Comparing Native Crystal Structures and AlphaFold2 Predicted Water-Soluble G Protein-Coupled Receptor QTY Variants

Michael A. Skuhersky [1,2,3], Fei Tao [1,3], Rui Qing [1,3], Eva Smorodina [4], David Jin [5] and Shuguang Zhang [1,*]

1 Laboratory of Molecular Architecture, Media Lab E15-391, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; vex@mit.edu (M.A.S.); taofei@sjtu.edu.cn (F.T.); ruiqing@mit.edu (R.Q.)
2 Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
3 Laboratory of Food Microbial Technology, State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiaotong University, Shanghai 200240, China
4 Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119991 Moscow, Russia; ribes.ev@gmail.com
5 Avalon GloboCare Corp., Freehold, NJ 07728, USA; david@avalon-globocare.com
* Correspondence: Shuguang@MIT.EDU; Tel.: +1-617-258-7514

Abstract: Accurate predictions of 3-dimensional protein structures by AlphaFold2 is a game-changer for biology, especially for structural biology. Here we present the studies of several native chemokine receptors including CCR5, CCR9, CXCR2 and CXCR4 determined by X-ray crystallography, and their water-soluble QTY counter parts predicted by AlphaFold2. In the native structures, there are hydrophobic amino acids leucine (L), isoleucine (I), valine (V) and phenylalanine (F) in the transmembrane helices. These hydrophobic amino acids are systematically replaced by hydrophilic amino acids glutamine (Q), threonine (T), and tyrosine (Y). Thus, the QTY variants become water-soluble. We also present the superimposed structures of native CCR10, CXCR5, CXCR7 and an olfactory receptor OR1D2 and their water-soluble QTY variants. Since the CryoEM structural determinations for the QTY variants of CCR10^{QTY} and OR1D2^{QTY} are in progress, it will be of interest to compare them when the structures become available. The superimposed structures show remarkable similarity within RMSD 1 Å–2 Å despite significant sequence differences (~26%–~33%). We also show the differences of hydrophobicity patches between the native GPCR and their QTY variants. Our study provides insight into the subtle differences between the hydrophobic helices and hydrophilic helices, and may further stimulate designs of water-soluble membrane proteins and other aggregated proteins.

Keywords: chemokine receptors; membrane protein design; protein structural predictions; QTY code; water-soluble GPCRs

1. Introduction

The recent release of the highly accurate protein prediction software suite AlphaFold2 has revolutionized protein science, especially for structural biology [1–3]. Likewise, a similar prediction program RoseTTAFold also made similar accurate protein structural predictions [4]. Together, they not only provided a promising solution for the long-standing difficult protein folding problem, and limited predictions of protein–protein interactions, but they also further stimulate research activities for protein design including design of water-soluble membrane proteins.

We previously designed several detergent-free transmembrane (TM) protein chemokine receptors and cytokine receptors for various uses using the conventional computing programs [5–9]. AlphaFold2 significantly reduces the time-consuming and expensive molecular simulation process and broadens the applications to wider research and biotech communities [1–3].

These chemokine receptors were chosen because they play many critical roles in health and diseases, and have been well characterized [10–13]. Chemokine receptors CCR5, CCR9, CCR10, CXCR2, CXCR4, CXCR5 and CXCR7 are also involved in various cancers initiation and metastasis, autoimmune diseases, viral infections and cytokine release syndrome, also called "cytokine storm". Moreover, the crystal structures of CXCR4, CCR5, CCR9 and CXCR2 have been published [14–17]. allowing direct comparison with the AlphaFold2 predicted water-soluble QTY variants.

We also chose several chemokine receptors CCR10, CXCR5 and CXCR7 although their structures are not yet available, due to their importance in biology and medicine [10–13]. 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 [18,19]. CXCR5 and its ligand CXCL13 are involved in chronic inflammation, infectious and immune responses, and autoimmune disorders [19]. CXCR7 is an atypical chemokine receptor (ACKR3) that does not activate G-protein-mediated signal transduction but recruits β-arrestins [20]. Rather, CXCR7 can heterodimerize with CXCR4 in order 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 [13,21–23].

It is intriguing and still an enigma how humans are able to detect and discriminate enormous numbers of odorants using only ~400 hundred olfactory receptors [24–27]. Despite the importance of understanding the structure and function of olfactory receptors as well as significant efforts and research funding, not a single olfactory receptor structure is known. One of the GPCRs we selected for the comparison is a human olfactory receptor (OR) OR1D2 which is a class I olfactory receptor with a narrow odorant recognition repertoire [28].

We are interested in the determinations of the CryoEM structures for the QTY variants of CCR10^{QTY} and OR1D2^{QTY}. After these structures become available, it will be of great interest to directly compare their structures. For now, it is very interesting to compare the native structures predicted by AlphaFold2 and their water-soluble QTY counter parts.

2. Materials and Methods

2.1. Protein Sequence Alignments

The native GPCR structures and their QTY variants are predicted using AlphaFold2. The CCR5^{QTY} sequence is 67.05% identical to the native CCR5, CCR9^{QTY} sequence is 73.98% identical to the native CCR9, CXCR2^{QTY} is 74.44% identical to the native CXCR2, and the CXCR4^{QTY} sequence is 70.74% identical to the native CXCR4. The CCR10^{QTY} sequence is 77.64% identical to the native CCR10, CXCR5^{QTY} is 76.31% identical to the native CXCR5, and the CXCR7^{QTY} sequence is 73.67% identical to the native CXCR7, OR1D2^{QTY} sequence is 82.70% identical to the native OR1D2. The molecular weights (MW) and pI values of the proteins were calculate using the service provided by Expasy (https://web.expasy.org/compute_pi/, accessed on 21 November 2021).

2.2. AlphaFold2 Predictions

Structure predictions of the QTY variants were performed with the AlphaFold2 [1] software following the instructions at the website https://github.com/sokrypton/ColabFold (accessed on 21 November 2021) on 2 × 20 Intel Xeon Gold 6248 cores, 384 GB RAM, and a Nvidia Volta V100 GPU. Other AlphaFold2 predicted structures are obtained from the European Bioinformatics Institute (EBI, Hinxton, Cambridgeshire, UK, https://alphafold.ebi.ac.uk, accessed on 21 November 2021) and also available at Uniprot website https://www.uniprot.org (accessed on 21 November 2021). Please see Table 1 for details.

Table 1. AlphaFold 2 structure prediction parameters.

Parameter	Value
homooligomer	1
msa_method	mmseqs2
msa_format	fas
pair_mode	unpaired
pair_cov	50
pair_qid	20
rank_by	pLDDT
use_turbo	True
max_msa	512:1024
show_images	True
num_models	5
use_ptm	True
num_ensemble	1
max_recycles	3
tol	0
num_samples	1
subsample_msa	True
num_relax	None

Each UniProt ID from the dataset was extended with ID, entry name, description, and FASTA sequence. The data were taken from UniProt using custom Python code. The FASTA sequences were converted into their soluble versions using the QTY method (https://pss.sjtu.edu.cn/, accessed on 21 November 2021), followed by Protter (http://wlab.ethz.ch/protter/start/, accessed on 21 November 2021) 2D diagrams plotting. These steps were optimized via Python libraries for web applications such as requests and splinter.

2.3. Superimpose the Structures

The published X-ray crystal structures of native CCR5 (4MBS), CCR9 (5LWE), CXCR2 (6LFL) and CXCR4 (3ODU) are obtained from the Protein Data Bank (PDB), https://www.rcsb.org (accessed on 21 November 2021). AlphaFold2 predictions of three native chemokine receptors CCR10, CXCR5, CXCR7, and one olfactory receptor OR1D2, and their QTY variants are carried out using the AlphaFold2 program at https://github.com/sokrypton/ColabFold (accessed on 21 November 2021). All other native receptor protein sequences are obtained from Uniprot https://www.uniprot.org (accessed on 21 November 2021). Structures were aligned with MUSTANG [29].

2.4. Structure Visualization

We used two programs for structure visualization: PyMOL https://pymol.org/2/ (accessed on 21 November 2021) and UCSF Chimera https://www.rbvi.ucsf.edu/chimera/ (accessed on 21 November 2021). All superimposed models were produced via PyMOL, while Chimera was used for hydrophobicity representation.

3. Results and Discussions

3.1. Protein Sequence Alignments

Since the crystal structures of native receptors CCR5, CCR9, CXCR2, and CXCR4 are already available, the sequence alignments were carried out for: CCR5 vs. CCR5^{QTY}, CCR9 vs. CCR9^{QTY}, CXCR2 vs. CXCR2^{QTY}, CXCR4 vs. CXCR4^{QTY} (Figure 1). Thus, their crystal structures and the AlphaFold2 predicted structures can be directly compared.

A

Name	pI	MW (kDa)	Variation rate(%)	Variation rate(TM,%)
CCR5	9.21	40.52	/	/
CCR5QTY	9.04	41.23	32.95	55.76

B

Name	pI	MW (kDa)	Variation rate(%)	Variation rate(TM,%)
CCR9	8.54	42.02	/	/
CCR9QTY	8.49	42.55	26.02	46.38

C

Name	pI	MW (kDa)	Variation rate(%)	Variation rate(TM,%)
CXCR2	8.66	40.76	/	/
CXCR2QTY	8.60	41.51	25.56	57.69

D

Name	pI	MW (kDa)	Variation rate(%)	Variation rate(TM,%)
CXCR4	8.46	39.75	/	/
CXCR4QTY	8.40	40.22	29.26	58.11
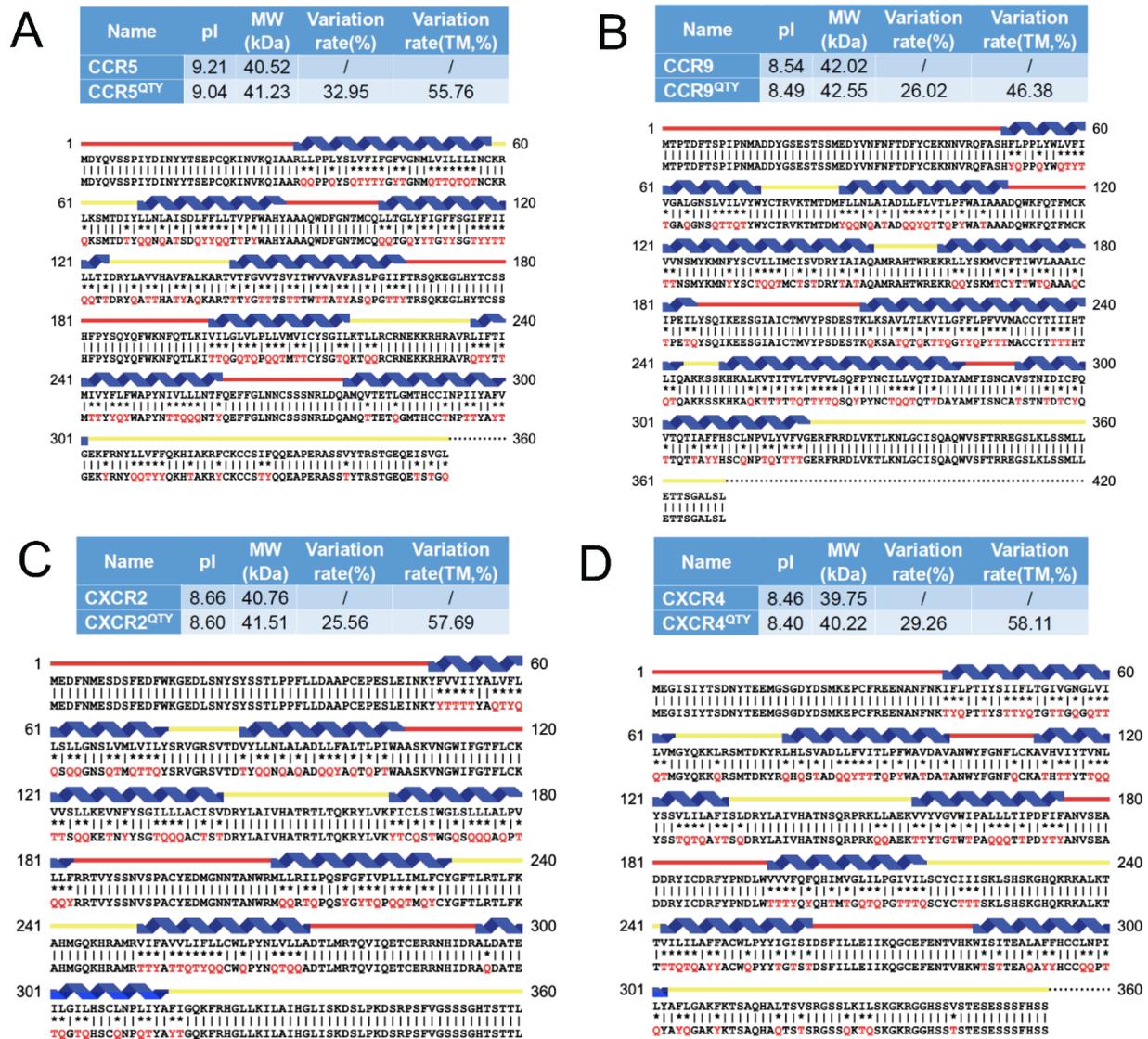
Figure 1. Alignment of known X-ray crystal structures of native chemokine receptors CCR5, CCR9, CXCR2 and CXCR4 with AlphaFold2 predicted native and their predicted water-soluble QTY variants. The Q, T, and Y amino acid substitutions are in red. The alpha-helical segments (blue) are shown above the protein sequences, the external (red) and internal (yellow) loops of the receptors are indicated. The symbols | and * indicate the similar and different amino acids, respectively. Characteristics of natural and QTY variants with pI, molecular weight, total variation rate and membrane variation rate, and the alignment: (**A**) CCR5 and CCR5QTY, (**B**) CCR9 and CCR9QTY, (**C**) CXCR2 and CXCR2QTY, and (**D**) CXCR4 and CXCR4QTY. Since the internal regions ICL1, ICL2, ICL3 and the C-terminus do not interact with the ligands, additional residues in these regions are sometimes QTY modified.

The QTY code selects 3 neutral and polar amino acids without any charges: glutamine (Q), threonine (T), and tyrosine (Y) to replace hydrophobic amino acids leucine (L), isoleucine (I), valine (V), and phenylalanine (F), since their electron density maps share remarkable structure similarities [5] Supporting information). After applying the QTY code, the hydrophobic amino acids in the transmembrane segments are replaced by Q, T, and Y; therefore, the transmembrane segments are no longer hydrophobic. Although there are greater than 26% overall replacements, the protein alignments of the transmembrane α-helical segments of the native receptors and their QTY variants show similar isoelectrical point (pI) and molecular weights (Figure 1).

Although there are substantial changes, ~26–~33% overall, in the amino acid composition between the native receptors and their QTY variants, and ~46–58% in the 7TM domains,

the pI changes are rather small. Respectively, being 0.06 units for CXCR2/CXCR2^{QTY}, 0.06 units for CXCR4/CXCR4^{QTY}, 0.05 units for CCR9/CCR9^{QTY} and 0.17 units for CCR5/ CCR5^{QTY}, these pI changes are insignificant regarding surface charges and unlikely to disrupt delicate structures. Likewise, the molecular weight differences are small through replacing saturated carbon side chains with $-OH$ and $-NH_2$ that can form 3–4 hydrogen bonds with water molecules. For example, CCR5/CCR5^{QTY} = 710 Daltons (1.75%), CCR9/CCR9^{QTY} = 530 Daltons (1.26%), CXCR2/CXCR2^{QTY} = 750 Daltons (1.84%), CXCR4/ CXCR4^{QTY} = 470 Daltons (1.18%), these changes may be tolerated by protein structures (Figure 1).

We also selected several receptors that do not yet have any experimental determined structures, including CCR10, CXCR5, CXCR7 and OR1D2 and their QTY variants for alignment: CCR10 vs. CCR10^{QTY}, CXCR5 vs. CXCR5^{QTY}, CXCR7 vs. CXCR7^{QTY} and OR1D2 vs. OR1D2^{QTY} (Figure 2). They also have very small pI changes, i.e., 0.24 units for CCR10/CCR10^{QTY}, 0.04 units for CXCR5/ CXCR5^{QTY}, 0.02 units for CXCR7/CXCR7^{QTY} and 0.05 units for OR1D2/OR1D2^{QTY}. Similarly, the molecular weight differences are small: CCR10/CCR10^{QTY} = 760 Daltons (1.98%), CXCR5/CXCR5^{QTY} = 590 Daltons (1.6%), CXCR7/CXCR7^{QTY} = 460 Daltons (1.26%), and OR1D2/OR1D2^{QTY} = 310 Daltons (0.82%).

3.2. AlphaFold2 Predictions

For over 6 decades, structural biologists and protein scientists have sought to predict how proteins fold naturally and rapidly. AlphaFold2's accurate predictions of 3-dimensional protein structures have made such predictions a reality, revolutionizing structural biology and protein science. Such an advancement not only allows us to study protein structure more in detail but also facilitate to obtain previously unattainable protein structures through AlphaFold2 predictions, at least in the framework.

It is estimated that ~25–30% genes code for membrane proteins in most organisms [30]. However, even the determination of a single membrane protein is a dauting task, and there are many hurdles along the way, from gene expression, protein production, detergent selection, purification, to maintaining the stability, integrity, and function to avoid aggregation. Thus, despite significant advances in tools and method developments, determinations of membrane protein structures still lag far behind of those of water-soluble proteins.

The advent of AlphaFold2 has completely changed this. Using AlphaFold2 predictions, DeepMind AlphaFold's team has already deposited over 100 million protein structures in the European Bioinformatics Institute (EBI, Hinxton, Cambridgeshire, UK, https://alphafold.ebi.ac.uk, accessed on 21 November 2021). The number will continuously increase over time.

We used the AlphaFold2 accurate prediction tool to predict water-soluble G protein-coupled receptors in this study and compare them to the known experimentally determined structures. The speed and accuracy of AlphaFold2 predictions of our designed receptors is unprecedented. Instead of taking weeks or days to predict one structure, AlphaFold2 can predict new structure in a few hours, or even minutes for smaller and simpler proteins. AlphaFold2 significantly accelerates studies of protein structures, stabilities, design new proteins, discovery of new protein interactions, and perhaps new functions that were previously unknown through experimental studies. It is a game changer for science, especially for protein science.
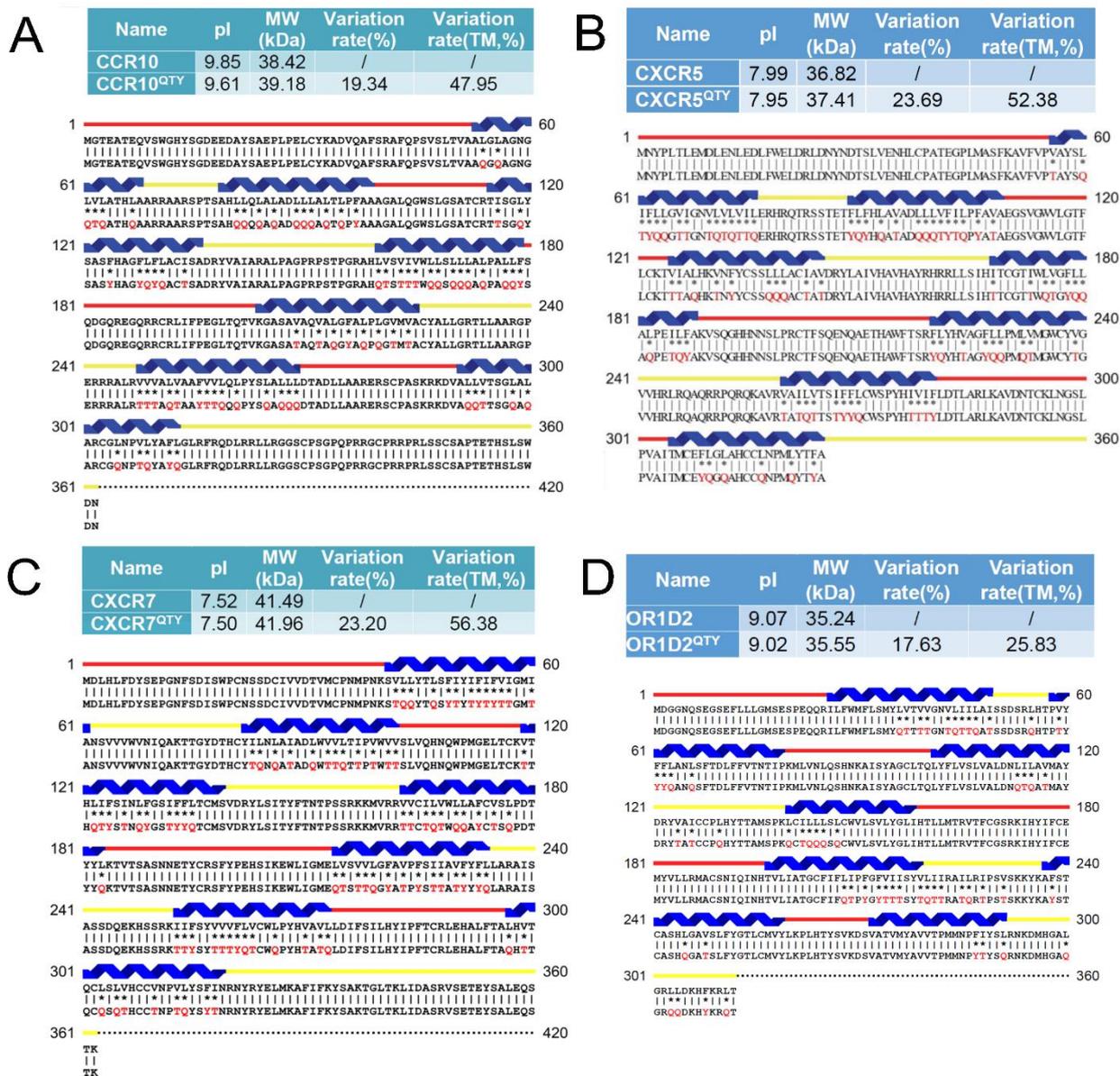
Figure 2. Alignment of AlphaFold2 predicted 3 native chemokine receptors and 1 olfactory receptor with their water-soluble QTY variants. The Q, T, and Y amino acid substitutions are in red. The alpha-helical segments (blue) are shown above the protein sequences, the external (red) and internal (yellow) loops of the receptors are indicated. The symbols | and * indicate the similar and different amino acids, respectively. Characteristics of natural and QTY variants with pI, molecular weight, total variation rate and membrane variation rate, and the alignment: (**A**) CCR10 and CCR10^{QTY}, (**B**) CXCR5 and CXCR5^{QTY} (**C**) CXCR7 and CXCR7^{QTY} and (**D**) OR1D2 and OR1D2^{QTY}. Since the internal regions ICL1, ICL2, ICL3 and the C-terminus mostly do not interact with the ligands, additional residues in these regions are sometimes also QTY modified.

3.3. Superimpositon of Native and Water-Soluble QTY Variants

The structures of native and water-soluble QTY variants superimposed very well except the loops and N- and C-termini. For simplicity and clarity, less-predictable N- and C-termini are not included in our predicted structures, only the 7TMs are included in the superimpositions (Figure 3). In the first sets of structures for CCR5, CCR9, CXCR2 and CXCR4, they are superimposed among: (i) the X-ray crystal determined structures CCR5 (Protein data bank code: 4MBS), CCR9 (5LWE), CXCR2 (6LFL) and CXCR4 (3ODU) (magenta color), (ii) the AlphaFold2 predicted native structures (green) and the (iii) AlphaFold2 predicted water-soluble QTY variants (cyan). As seen from the Figure 3, the structures are viewed from front, back and top, all 4 sets of structures superimposed well. The results

suggest that these structures share very similar folds despite of significant QTY amino acids replacement in the water soluble QTY receptor variants. These superimpositions also suggest that the AlphaFold2′s accurate prediction capabilities since the predicted structures are directly superimposable with the experimentally determined X-ray crystal structures.
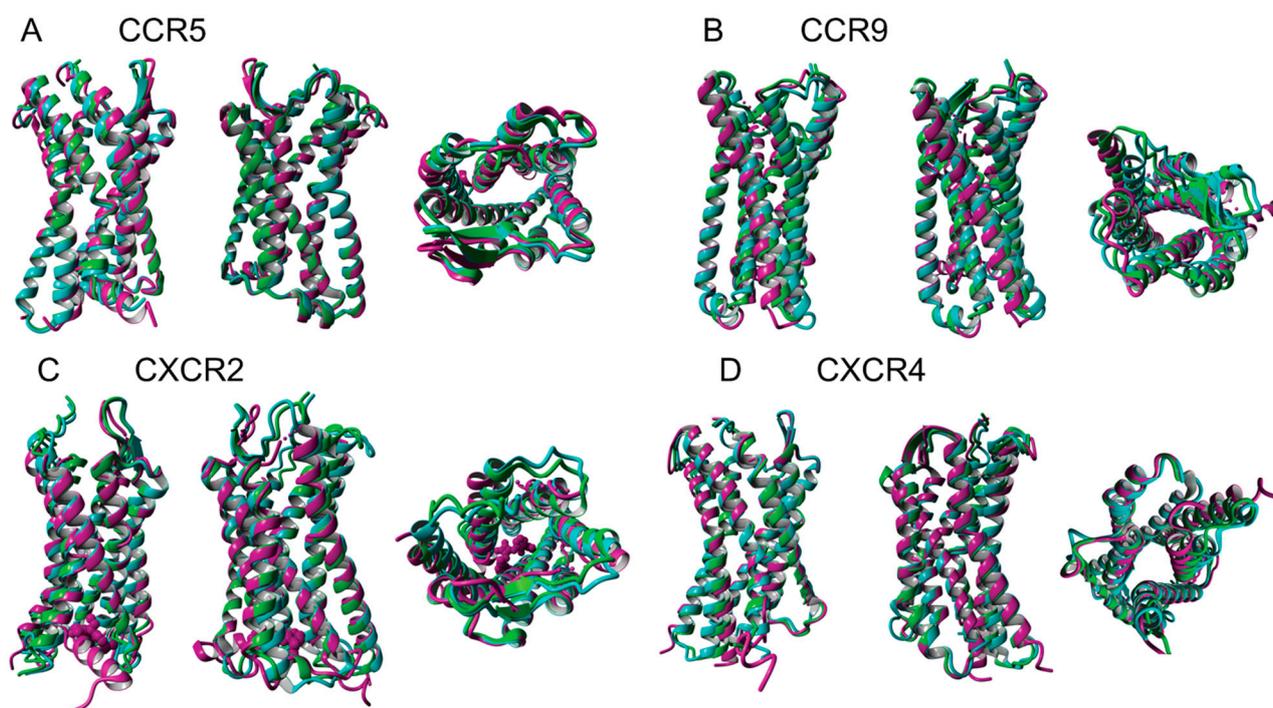
Figure 3. Superimposed determined 4 crystal structures, AlphaFold2 predicted 4 native chemokine receptors CCR5, CCR9, CXCR2, and CXCR4 and their QTY water-soluble variants. For each superimposition, 3 structures are show, front (left), back (middle) and view from top axis (right). The X-ray crystal structures of natural CCR5 (4MBS), CCR9 (5LWE), CXCR2 (6LFL), and CXCR4 (3ODU) are obtained from the Protein Data Bank (PDB). For clarity, the inserts for crystallization and N- and C-termini are removed. The crystal structure native (**A**) CCR5 (magenta) is superimposed with AlphaFold2 predicted native CCR5 (green) and water-soluble variant CCR5^{QTY} (cyan). Following the same order as CCR5, the superimposed determined and AlphaFold2 predicted structures of (**B**) CCR9, (**C**) CXCR2, (**D**) CXCR4.

Work is currently ongoing to confirm the QTY variant structures by CryoEM. We also directly superimposed the experimentally unknown but AlphaFold2-predicted structures of 4 native receptors and water-soluble QTY variants of chemokine receptors CCR10, CXCR5, CXCR7 and an olfactory receptor OR1D2 (Figure 4). Despite the significant differences of: (i) amino acid compositions, (ii) chemical characteristics, and (iii) sequence identities between ~67.5–83%, the structural similarity of the native and QTY variant receptors are presented by root mean square deviation (RSMD): (i) CCR10 vs. CCR10^{QTY} (RMSD 1.246 Å, 77% identity), (ii) CXCR5 vs. CXCR5^{QTY} (RMSD 1.110 Å, 71% identity), (iii) CXCR7 vs. CXCR7^{QTY} (RMSD 1.388 Å, 67.5% identity), and (iv) OR1D2 vs. OR1D2^{QTY} (RMSD 0.978 Å, 83% identity). These very closely superimposed structures again show the remarkable similarity of native and water-soluble QTY variants. This demonstrates that these alpha-helices have very similar molecular structures regardless of the hydrophobicity and hydrophilicity, namely: (i) 1.5 Å per amino acid rise, (ii) 100° per amino acid turn, (iii) 5.4 Å, 360° and 3.6 amino acids per helical turn [31–33].
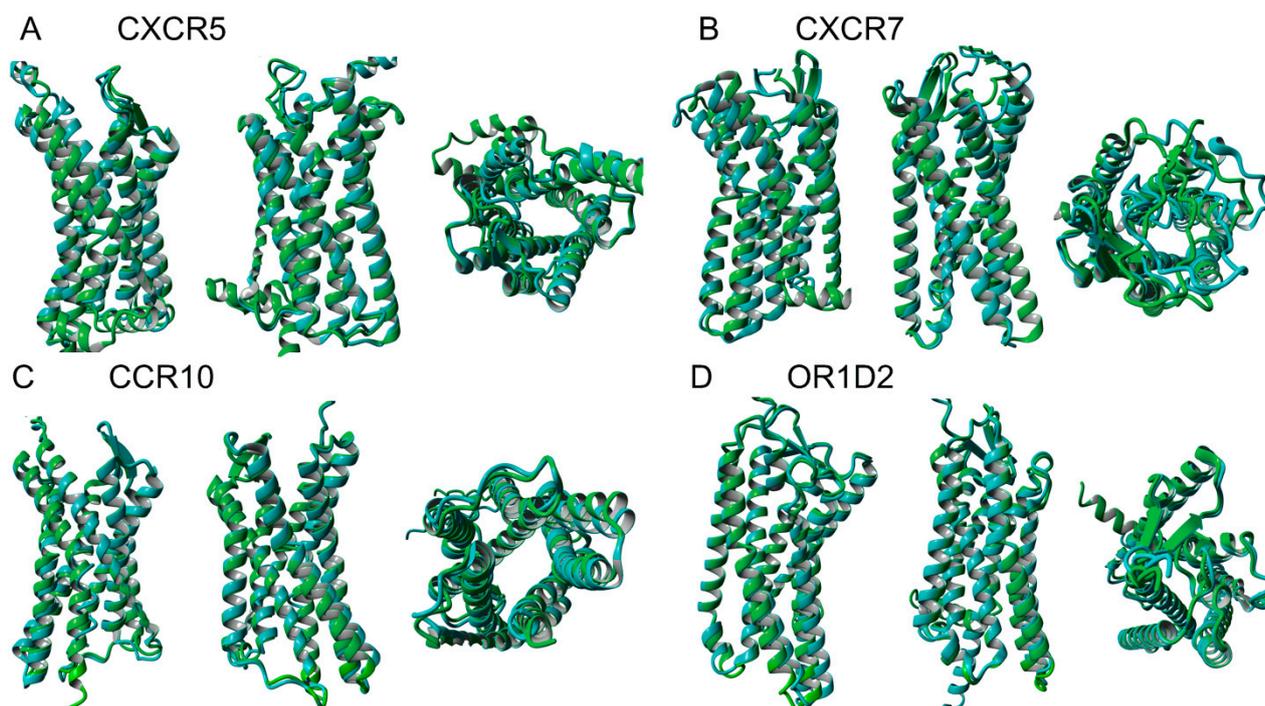
Figure 4. Superimposed AlphaFold2 predicted 3 native chemokine receptors and 1 olfactory receptor and their water-soluble QTY variants. For each superimposition, 3 structures are show, front (**left**), back (**middle**), and view from top axis (**right**). For clarity, N- and C-termini and large intracellular and extracellular loops are removed. (**A**) CCR10 vs. CCR10^{QTY} (RMSD 1.246 Å, 77% identity), (**B**) CXCR5 vs. CXCR5^{QTY} (RMSD 1.110 Å, 71% identity), (**C**) CXCR7 vs. CXCR7^{QTY} (RMSD 1.388 Å, 67.5% identity), and (**D**) OR1D2 vs. OR1D2^{QTY} (RMSD 0.978 Å, 83% identity).

We also carried out AlphaFold2 predictions of other variations. For example, we changed the QTY to alanine for CXCR4^{QTY}, arginine or glycine for CCR5^{QTY}. These results are shown in Supplementary Material Figure S1. Since alanine and arginine have propensity to form alpha-helix, the overall deviation is less than the structure change with glycine. This is not surprising, since glycine is an achiral amino acid that is very flexible and can influence the structures.

3.4. Analysis of the Hydrophobic Surface of Native and Water-Soluble QTY Variants

Native GPCRs are highly hydrophobic, especially in the 7TM helical domains, thus intrinsically water-insoluble, and require detergents to solubilize them after isolating them from the lipid bilayer membranes. Without the appropriate detergents, they immediately aggregate, precipitate, and lose their biological functions. The 7TM domains are directly embedded in the hydrophobic lipid bilayer so the hydrophobic side chains of amino acids leucine (L), isoleucine (I), valine (V), and phenylalanine (F) directly interact with the lipid molecules excluding water. Thus, they display highly hydrophobic patches on the 7TM domain (Figures 5 and 6).
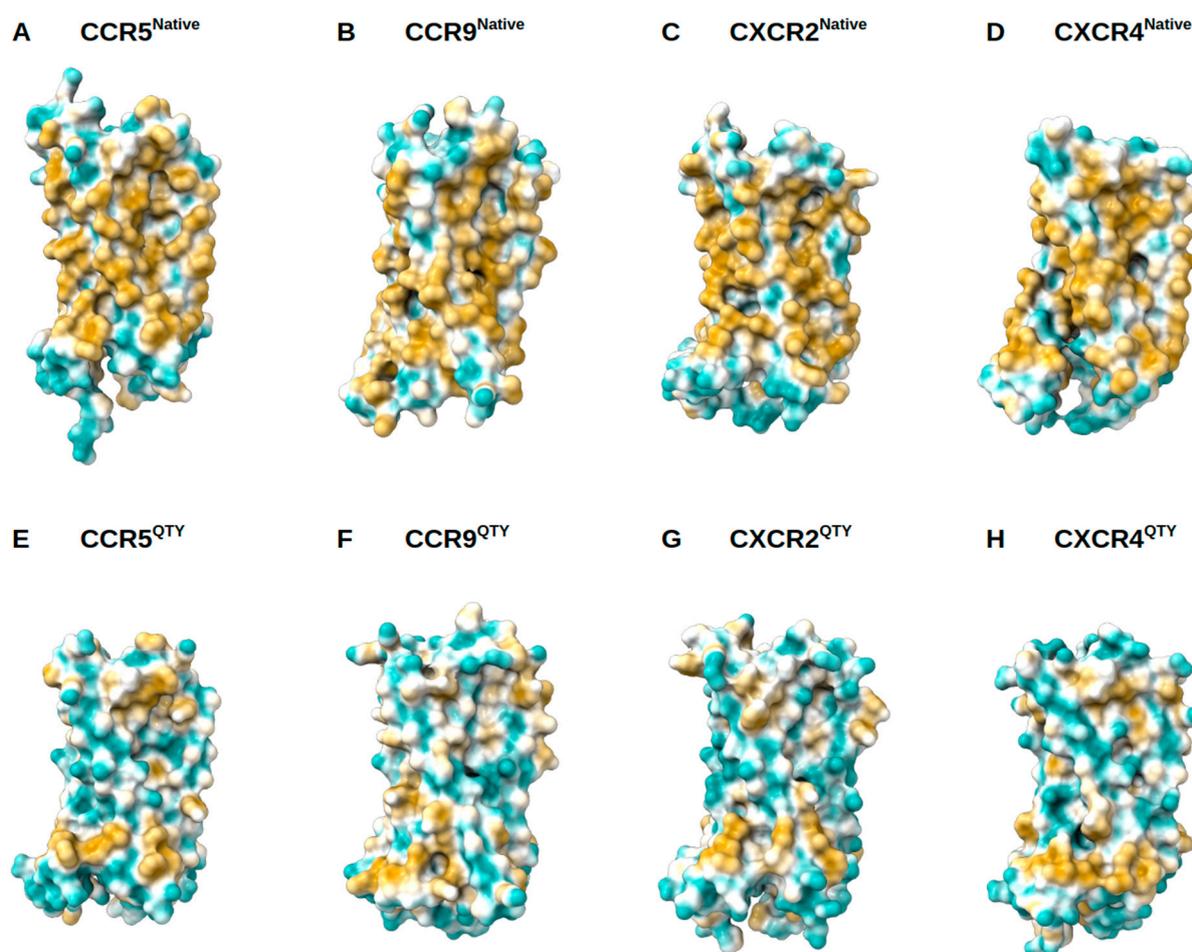
Figure 5. Surface hydrophobic patch of X-ray determined structures of native chemokine receptors and AlphaFold2 predicted water-soluble QTY variants. The native GPCR receptors mostly expose hydrophobic residues leucine (L), isoleucine (I), valine (V) and phenylalanine (F) facing outside to the hydrophobic lipid bilayer in cell membrane. After replacing the L, I, V, F with polar amino acids, glutamine (Q), threonine (T), and tyrosine (Y), the surfaces are much less hydrophobic. The large surface hydrophobic patch (yellow color) of the native receptors determined by X-ray crystal structures: (**A**) CCR5, (**B**) CCR9, (**C**) CXCR2 and (**D**) CXCR4. The hydrophobic patch is significantly reduced on the transmembrane domains for the AlphaFold2 predicted water-soluble QTY variants: (**E**) CCR5^{QTY}, (**F**) CCR9^{QTY}, (**G**) CXCR2^{QTY}, (**H**) CXCR4^{QTY}. These QTY variants become water-soluble without any detergent. The N- and C-termini are removed for clarity of directly comparisons.

On the other hand, after systematic QTY replacement of hydrophobic amino acids L, I, V, F, with hydrophilic amino acids glutamine (Q), threonine (T), and tyrosine (Y), these hydrophobic patches become largely reduced (Figures 5 and 6). Such transformation from hydrophobic 7TM to hydrophilic 7TM did not significantly alter the alpha-helix structures, stability or their ligand binding functions (5–9). This was rather unexpected before our previously systematic experiments were carried out. However, the experimental evidence conclusively demonstrated preservation of both structure and function (5–9).

We note that nature has already made such alpha-helices in both hydrophilic form, such as those found in hemoglobin and many other water-soluble enzymes, and hydrophobic helices, such as in GPCRs and other membrane proteins. These helices have very similar molecular structures irrespective of the hydrophobicity and hydrophilicity [31–33].
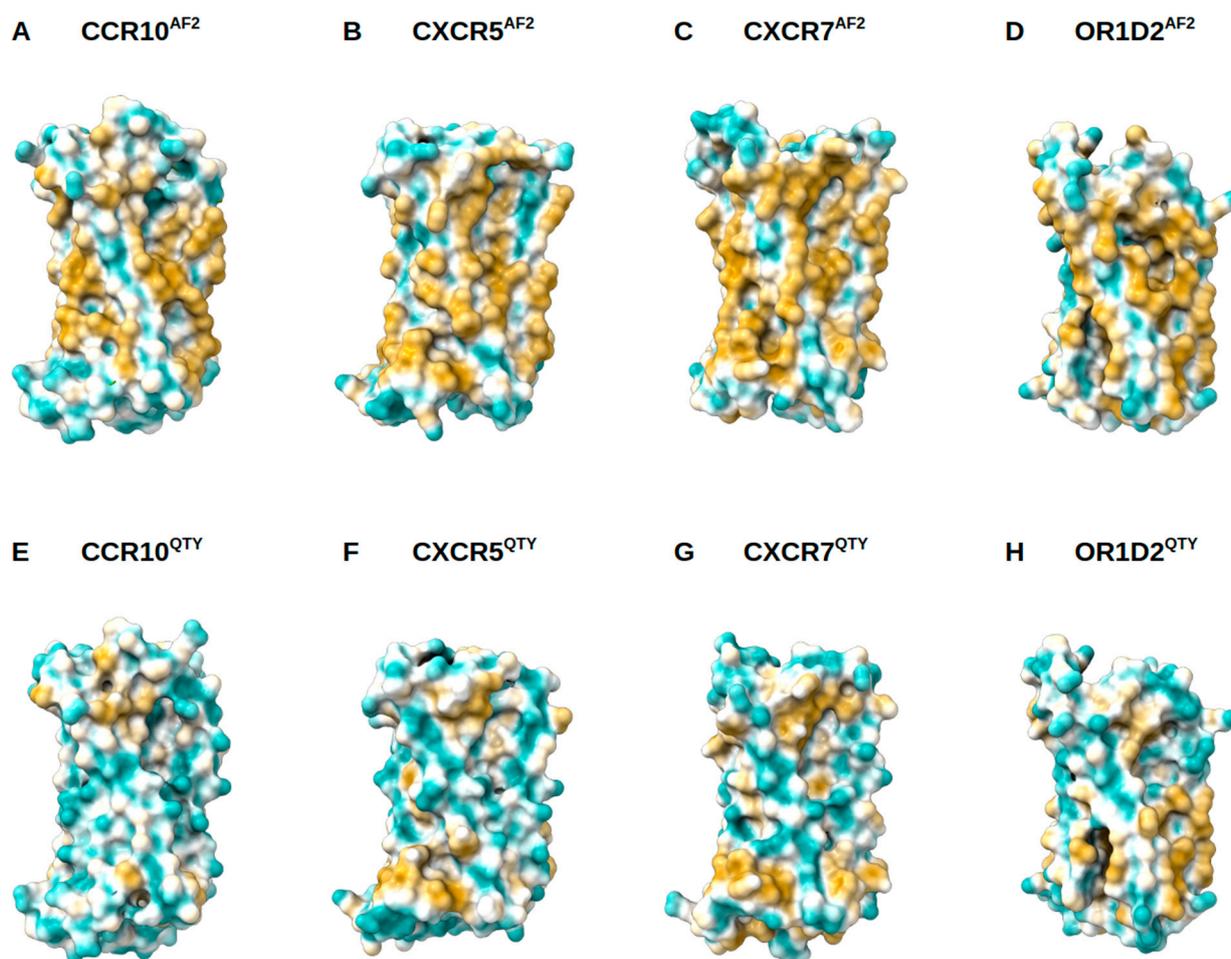
Figure 6. Surface hydrophobic patch of AlphaFold2 predicted structures of native chemokine receptors CCR10, CXCR5, CXCR7, and OR1D2 and their water-soluble QTY variants. The AlphaFold2 predicted structure receptors CCR10, CXCR5, CXCR7, and OR1D2 are presented. The large surface hydrophobic patch (yellow color) of the native receptors determined by AlphaFold2 predicted structures (cyan color): (**A**) CCR10, (**B**) CXCR5, (**C**) CXCR7, and (**D**) OR1D2. The hydrophobic patch is significantly reduced on the transmembrane domains for the AlphaFold2 predicted water-soluble QTY variants: (**E**) CCR10^{QTY}, (**F**) CCXC5^{QTY}, (**G**) CXCR7^{QTY}, and (**H**) OR1D2^{QTY}. These QTY variants become water-soluble without any detergent. The N- and C-termini are removed for clarity of directly comparisons.

4. Conclusions

Our study of systematically comparing experimentally determined structures with AlphaFold2 predicted water-soluble membrane receptors provides insight into the subtle differences between the hydrophobic helices and hydrophilic helices. They not only show the QTY code's simplicity to directly replace hydrophobic amino acid residues in the membrane proteins while maintaining their structures, but also demonstrate the validity of the QTY code. Thus, applying the QTY code may further stimulate designs of water-soluble membrane proteins and other aggregated proteins.

Author Contributions: Conceptualization, M.A.S. and S.Z.; methodology, M.A.S., F.T. and E.S.; software, AlphaFold2, M.A.S. and E.S.; validation, M.A.S.; formal analysis, M.A.S. and F.T.; investigation, M.A.S., F.T. and R.Q.; resources, D.J.; data curation, M.A.S. and E.S.; writing—original draft prepara-

tion, S.Z.; writing—review and editing, M.A.S., F.T., R.Q., D.J. and S.Z.; visualization, M.A.S. and E.S.; supervision, S.Z.; project administration, S.Z.; funding acquisition, D.J. All authors have read and agreed to the published version of the manuscript.

Competing Financial Interests

Massachusetts Institute of Technology (MIT) filed several patent applications for the QTY code and OH_2Laboratories licensed the technology from MIT. SZ is an inventor and has a minor equity of OH_2Laboratories shares. The other authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [CrossRef]
2. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nat. Cell Biol.* **2021**, *596*, 583–589. [CrossRef]
3. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nat. Cell Biol.* **2021**, *596*, 590–596. [CrossRef]
4. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [CrossRef] [PubMed]
5. Zhang, S.; Tao, F.; Qing, R.; Tang, H.; Skuhersky, M.; Corin, K.; Tegler, L.; Wassie, A.; Wassie, B.; Kwon, Y.; et al. QTY code enables design of detergent-free chemokine receptors that retain ligand-binding activities. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E8652–E8659. [CrossRef]
6. Qing, R.; Han, Q.; Fei, T.; Skuhersky, M.; Badr, M.; Chung, H.; Schubert, T.; Zhang, S. QTY code designed thermostable and water-soluble chimeric chemokine receptors with tunable ligand-binding activities. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 25668–25676. [CrossRef]
7. Tegler, L.; Corin, K.; Pick, H.; Brookes, J.; Skuhersky, M.; Vogel, H.; Zhang, S. The G protein coupled receptor CXCR4 designed by the QTY code becomes more hydrophilic and retains cell signaling activity. *Sci. Rep.* **2020**, *10*, 1–12. [CrossRef] [PubMed]
8. Hao, S.; Jin, D.; Zhang, S.; Qing, R. QTY Code-designed Water-soluble Fc-fusion Cytokine Receptors Bind to their Respective Ligands. *QRB Discov.* **2020**, *1*, 1–18. [CrossRef] [PubMed]
9. Qing, R.; Tao, F.; Chatterjee, P.; Yang, G.; Han, Q.; Chung, H.; Ni, J.; Suter, B.P.; Kubicek, J.; Maertens, B.; et al. Non-full-length Water-Soluble CXCR4QTY and CCR5QTY Chemokine Receptors: Implication for Overlooked Truncated but Functional Membrane Receptors. *iScience* **2020**, *23*, 101670. [CrossRef]
10. Proudfoot, A.E.I. Chemokine receptors: Multifaceted therapeutic targets. *Nat. Rev. Immunol.* **2002**, *2*, 106–115. [CrossRef]
11. Dorsam, R.T.; Gutkind, J.S. G-protein-coupled receptors and cancer. *Nat. Rev. Cancer* **2007**, *7*, 79–94. [CrossRef]
12. O'Hayre, M.; Degese, M.S.; Gutkind, J.S. Novel insights into G protein and G protein-coupled receptor signaling in cancer. *Curr. Opin. Cell Biol.* **2014**, *27*, 126–135. [CrossRef]
13. Poeta, V.M.; Massara, M.; Capucetti, A.; Bonecchi, R. Chemokines and Chemokine Receptors: New Targets for Cancer Immunotherapy. *Front. Immunol.* **2019**, *10*, 379. [CrossRef]
14. Wu, B.; Chien, E.Y.T.; Mol, C.D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F.C.; et al. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science* **2010**, *330*, 1066–1071. [CrossRef] [PubMed]
15. Tan, Q.; Zhu, Y.; Li, J.; Chen, Z.; Han, G.W.; Kufareva, I.; Li, T.; Ma, L.; Fenalti, G.; Li, J.; et al. Structure of the CCR5 Chemokine Receptor–HIV Entry Inhibitor Maraviroc Complex. *Science* **2013**, *341*, 1387–1390. [CrossRef] [PubMed]

16. Oswald, C.; Rappas, M.; Kean, J.; Doré, A.S.; Errey, J.C.; Bennett, K.; Deflorian, F.; Christopher, J.A.; Jazayeri, A.; Mason, J.S.; et al. Intracellular allosteric antagonism of the CCR9 receptor. *Nat. Cell Biol.* **2016**, *540*, 462–465. [CrossRef]

17. Liu, K.; Wu, L.; Yuan, S.; Wu, M.; Xu, Y.; Sun, Q.; Li, S.; Zhao, S.; Hua, T.; Liu, Z.-J. Structural basis of CXC chemokine receptor 2 activation and signalling. *Nat. Cell Biol.* **2020**, *585*, 1–9. [CrossRef] [PubMed]

18. Korbecki, J.; Grochans, S.; Gutowska, I.; Barczak, K.; Baranowska-Bosiacka, I. CC Chemokines in a Tumor: A Review of Pro-Cancer and Anti-Cancer Properties of Receptors CCR5, CCR6, CCR7, CCR8, CCR9, and CCR10 Ligands. *Int. J. Mol. Sci.* **2020**, *21*, 7619. [CrossRef]

19. Karnezis, T.; Farnsworth, R.H.; Harris, N.C.; Williams, S.P.; Caesar, C.; Byrne, D.J.; Herle, P.; Macheda, M.L.; Shayan, R.; Zhang, Y.F.; et al. CCL27/CCL28-CCR10 Chemokine Signaling Mediates Migration of Lymphatic Endothelial Cells. *Cancer Res.* **2019**, *79*, 1558–1572. [CrossRef]

20. Ehrlich, A.T.; Semache, M.; Couvineau, P.; Wojcik, S.; Kobayashi, H.; Thelen, M.; Gross, F.; Hogue, M.; Le Gouill, C.; Darcq, E.; et al. Ackr3-Venus knock-in mouse lights up brain vasculature. *Mol. Brain* **2021**, *14*, 151. [CrossRef]

21. Sun, X.; Cheng, G.; Hao, M.; Zheng, J.; Zhou, X.; Zhang, J.; Taichman, R.S.; Pienta, K.J.; Wang, J. CXCL12/CXCR4/CXCR7 chemokine axis and cancer progression. *Cancer Metastasis Rev.* **2010**, *29*, 709–722. [CrossRef]

22. Gustavsson, M.; Wang, L.; Van Gils, N.; Stephens, B.; Zhang, P.; Schall, T.J.; Yang, S.; Abagyan, R.; Chance, M.R.; Kufareva, I.; et al. Structural basis of ligand interaction with atypical chemokine receptor 3. *Nat. Commun.* **2017**, *8*, 14135. [CrossRef] [PubMed]

23. Shi, Y.; Riese, D.J.I.; Shen, J. The Role of the CXCL12/CXCR4/CXCR7 Chemokine Axis in Cancer. *Front. Pharmacol.* **2020**, *11*, 574667. [CrossRef]

24. Buck, L.; Axel, R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **1991**, *65*, 175–187. [CrossRef]

25. Mombaerts, P.; Wang, F.; Dulac, C.; Vassar, R.; Chao, S.K.; Nemes, A.; Mendelsohn, M.; Edmondson, J.; Axel, R. The molecular biology of olfactory perception. *Cold Spring Harb. Symp. Quant. Biol.* **1996**, *61*, 135–145. [PubMed]

26. Buck, L.B. Unraveling smell. *Harvey Lect.* **2007**, *101*, 117–134.

27. Touhara, K. Odor discrimination by G protein-coupled olfactory receptors. *Microsc. Res. Tech.* **2002**, *58*, 135–141. [CrossRef] [PubMed]

28. Sanz, G.; Schlegel, C.; Pernollet, J.-C.; Briand, L. Comparison of Odorant Specificity of Two Human Olfactory Receptors from Different Phylogenetic Classes and Evidence for Antagonism. *Chem. Senses* **2005**, *30*, 69–80. [CrossRef]

29. Konagurthu, A.; Whisstock, J.; Stuckey, P.; Lesk, A. MUSTANG: A multiple structural alignment algorithm. *Proteins Struct. Funct. Bioinform.* **2006**, *64*, 559–574. [CrossRef] [PubMed]

30. Almén, M.S.; Nordström, K.J.; Fredriksson, R.; Schiöth, H.B. Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **2009**, *7*, 50. [CrossRef]

31. Pauling, L.; Corey, R.B. Atomic Coordinates and Structure Factors for Two Helical Configurations of Polypeptide Chains. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 235–240. [CrossRef] [PubMed]

32. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W.H. Freeman: New York, NY, USA, 1998; pp. 523–532.

33. Brändén, C.-I.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Publishing: London, UK; New York, NY, USA, 1999; pp. 15–17.