Selection of human-zebrafish GPCR orthologs

Human GPCR gene annotations with HGNC-approved nomenclature were obtained from the HGNC gene group database (1), and receptors associated for olfaction or taste were filtered out. Human-zebrafish orthology data were retrieved from the Zebrafish Information Network (ZFIN) (2). These datasets were cross-referenced to identify overlapping human genes and their corresponding zebrafish orthologs. Structural models for the identified human and zebrafish GPCRs were then downloaded from AlphaFold Protein Structure Database (3,4).

Human-zebrafish GPCR ortholog alignment and binding pocket extraction

For structural alignment, the following representatives of GPCR Classes A, B1, C, and F were used as references:

[Class A] 2RH1 (ligand CAU, chain A)

[Class B1] 4K5Y (ligand 1Q5, chain A)

[Class C] 7M3G (ligand H43, chain A)

[Class F] 4JKV (ligand 1KS, chain A)

The specified chains were isolated using the “select” command in PyMOL (5). Each selected human and zebrafish gene structure was aligned to all four reference structures using TM-align (6), resulting in four alignments per gene. The aligned structures, along with their respective reference, were then loaded with the PyMOL 2.0 package (7); binding pocket residues were defined by selecting all residues from the aligned gene structure within 5.0 Å of the reference ligand and saved as individual .pdb files.

Human-zebrafish GPCR ortholog binding pocket conservation analysis

For all binding pockets, the centers of mass (COMs) of the included residues were computed. For each human gene and a corresponding zebrafish ortholog, the nearest zebrafish residue within a defined threshold (2.0 or 2.5 Å) of each human residue was identified and marked as a match (same residue identity) or mismatch (different residue identity). Human residues with no zebrafish residue within the threshold were noted as a residue that “failed” to find an alignment. Percent similarity between each human–zebrafish pair was calculated as the number of matched residues divided by the total number of matched and mismatched residues.

For each pair, the alignment yielding the highest percent similarity across GPCR classes was retained. Pairs with fewer than six matched residues were excluded from further analysis. Alignments with percent similarity between 60–80% were manually inspected using PyMOL, and visibly poor alignments were flagged. Pairings with a relatively high ratio of “failed” residues to matched and mismatched residues were also flagged. All flagged zebrafish pockets, as well as those from pairs with similarity ≤60%, were compared against a database of the successfully aligned zebrafish pockets using the NIH Basic Local Alignment Search Tool (BLAST) (8).

Each poorly aligned zebrafish pocket was realigned using TM-align to the most sequentially similar, successfully aligned zebrafish pocket identified via BLAST. If the best-matching pocket originated from a different GPCR class than the problematic alignment, both class contexts were evaluated. Specifically, the realigned pocket was compared against the human pocket from both the original class (e.g., Class B1) and the class associated with the BLAST-identified pocket (e.g., Class A). Any alignment showing improved quality—reflected by a higher match-to-mismatch ratio or fewer unaligned residues—was incorporated into the final dataset. For each human gene, the zebrafish ortholog with the highest percent similarity was retained for visualization.

**Supplementary References**

1. Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, Bruford EA. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res. PMID: 36243972 DOI: 10.1093/nar/gkac888
2. Bradford, Y.M., Van Slyke, C.E., Ruzicka, L., Singer, A., Eagle, A., Fashena, D., Howe, D.G., Frazer, K., Martin, R., Paddock, H., Pich, C., Ramachandran, S., Westerfield, M. (2022) Zebrafish Information Network, the knowledgebase for Danio rerio research. Genetics. 220(4).
3. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9. doi:10.1038/s41586-021-03819-2.
4. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022 Jan 7;50(D1):D439–44. doi:10.1093/nar/gkab1061.
5. The PyMOL Molecular Graphics System, Version 3.1.0a0, Schrödinger, LLC.
6. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005 Apr 1;33(7):2302–9. doi:10.1093/nar/gki524.
7. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.0. New York, NY: Schrödinger, LLC; 2017.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.