**"Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications", Cvicek et al.**

## Supporting Text 1

Here we compare the GRoSS alignment to two other available GPCR alignments:

1. Structure based alignment with gaps developed by Isberg et al. [24], downloaded from GPCRDB [78]
2. HMM-HMM based alignment created by hhalign [77]

### Comparison of GRoSS sequence alignment to HMM-HMM and GPCRDB

*GRoSS*

Within each GPCR class, the GRoSS alignment is an alignment that preserves the most conserved (BW) residues and has been curated not to contain gaps in the TM regions. First, we used Clustal Omega [37] to align small (~40) groups of proteins together so that no gaps are created in the TM regions. These individual alignments were then aligned together again using Clustal Omega, which uses hidden Markov model for profile-profile alignments. Between classes, the sequences were aligned to maximize the number of conserved contacts.

*HMM-HMM*

We compare our alignment to a general HMM-HMM alignment computed for each target—template pair taken from the available crystal structures. For each crystal structure, we searched for related sequences using hhblitz (http://toolkit.lmb.uni-muenchen.de/hhblits with database: uniprot20_2013_03). The representative alignment was stored and used as a HMM model for HMM-HMM alignment performed with hhalign [77].

*GPCRDB*

Isberg et al. [24] performed a detailed structural comparison of the GPCR crystal structures and concluded that the optimal alignments of TMs between some pairs of GPCR proteins have single residue gaps, which correspond to bulges or constrictions on the helices. We downloaded the Isberg alignment from GPCRDB [78], which contains several differences from the published version in [24] (panels A, I, and N in Figure 3 of ref. [24]). We assume that the differences are caused by a need to reconcile the pairwise alignments into a global alignment.

To the best of our knowledge, the gaps in the GPCRDB alignment cannot be predicted without prior knowledge of the protein structures. For example, sequence CIGWG in CRF1 aligns to IG-WG in GLR (Fig. 3G in [24]). Also HMM-HMM has

difficulty identifying the GPCRDB's gaps. Table A highlights the differences between HMM-HMM and GPCRDB alignments in terms of number of misaligned residues in all TM regions for the crystallized GPCRs considered. HMM-HMM manages to predict all GPCRDB TM gaps for only a small number of target—template pairs (value 0 in Table A). However, for most pairs even within class A HMM-HMM cannot correctly determine the gaps (small positive values in the table). Large number of misaligned residues between different classes means that HMM-HMM is not suitable for comparisons across the different GPCR classes.

Table A. Number of misaligned residues in the TM regions between GPCRDB and HMM-HMM.

| | | Template | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | | | | | | | | | | | | | | | | | | B | | C | | F |
| Target | | RHO | Beta1AR | Beta2AR | D3 | H1 | M2 | M3 | 5HT1B | 5HT2B | A2A | S1P1 | CXCR4 | CCR5 | KappaOR | MuOR | NOP | DeltaOR | PAR1 | P2Y12 | CRF1 | GLR | MGLU1 | MGLU5 | SMO |
| A | RHO | | 1 | 1 | 3 | 1 | 0 | 0 | 0 | 12 | 4 | 33 | 4 | 4 | 0 | 2 | 2 | 2 | 41 | 23 | 30 | 18 | 96 | 94 | 93 |
| | Beta1AR | 1 | | 0 | 5 | 2 | 1 | 1 | 1 | 15 | 5 | 34 | 5 | 7 | 1 | 3 | 3 | 3 | 43 | 27 | 34 | 19 | 92 | 88 | 77 |
| | Beta2AR | 1 | 0 | | 6 | 3 | 1 | 1 | 1 | 15 | 5 | 34 | 5 | 7 | 1 | 3 | 3 | 3 | 43 | 26 | 33 | 19 | 97 | 91 | 77 |
| | D3 | 3 | 5 | 6 | | 5 | 5 | 3 | 1 | 6 | 3 | 33 | 3 | 3 | 7 | 8 | 8 | 8 | 35 | 21 | 48 | 26 | 100 | 96 | 93 |
| | H1 | 1 | 2 | 1 | 5 | | 1 | 1 | 1 | 5 | 5 | 34 | 3 | 3 | 7 | 8 | 9 | 9 | 39 | 23 | 40 | 24 | 101 | 97 | 83 |
| | M2 | 0 | 1 | 1 | 4 | 1 | | 0 | 0 | 12 | 4 | 32 | 5 | 5 | 0 | 2 | 3 | 3 | 37 | 23 | 44 | 18 | 97 | 94 | 71 |
| | M3 | 0 | 1 | 1 | 3 | 1 | 0 | | 0 | 12 | 4 | 32 | 4 | 4 | 0 | 2 | 2 | 2 | 36 | 22 | 43 | 18 | 101 | 95 | 72 |
| | 5HT1B | 0 | 1 | 1 | 1 | 1 | 0 | 0 | | 9 | 4 | 33 | 2 | 2 | 1 | 2 | 2 | 2 | 34 | 21 | 45 | 16 | 96 | 92 | 84 |
| | 5HT2B | 12 | 15 | 15 | 6 | 5 | 12 | 12 | 9 | | 13 | 41 | 12 | 11 | 15 | 15 | 16 | 16 | 42 | 30 | 51 | 32 | 108 | 100 | 89 |
| | A2A | 4 | 5 | 5 | 3 | 5 | 4 | 4 | 4 | 13 | | 23 | 6 | 5 | 4 | 6 | 6 | 6 | 35 | 19 | 27 | 19 | 94 | 92 | 89 |
| | S1P1 | 33 | 34 | 34 | 33 | 34 | 32 | 32 | 33 | 41 | 23 | | 33 | 33 | 33 | 33 | 33 | 33 | 40 | 35 | 45 | 18 | 114 | 108 | 101 |
| | CXCR4 | 4 | 5 | 4 | 3 | 3 | 5 | 4 | 2 | 12 | 6 | 33 | | 3 | 7 | 6 | 5 | 6 | 32 | 25 | 21 | 17 | 98 | 91 | 98 |
| | CCR5 | 4 | 7 | 7 | 3 | 3 | 5 | 4 | 2 | 11 | 5 | 33 | 3 | | 4 | 4 | 4 | 5 | 31 | 26 | 21 | 20 | 103 | 99 | 103 |
| | KappaOR | 0 | 1 | 1 | 7 | 7 | 1 | 0 | 1 | 15 | 4 | 33 | 7 | 4 | | 1 | 0 | 1 | 35 | 29 | 24 | 9 | 98 | 93 | 103 |
| | MuOR | 2 | 3 | 3 | 8 | 8 | 2 | 2 | 2 | 15 | 6 | 33 | 6 | 4 | 1 | | 0 | 0 | 36 | 26 | 24 | 15 | 97 | 93 | 97 |
| | NOP | 2 | 3 | 3 | 8 | 9 | 3 | 2 | 3 | 16 | 6 | 33 | 5 | 4 | 0 | 0 | | 0 | 36 | 25 | 24 | 20 | 94 | 92 | 99 |
| | DeltaOR | 2 | 3 | 3 | 8 | 9 | 3 | 2 | 2 | 16 | 6 | 33 | 6 | 5 | 1 | 0 | 0 | | 36 | 26 | 24 | 10 | 94 | 91 | 100 |
| | PAR1 | 41 | 43 | 43 | 35 | 39 | 37 | 36 | 34 | 42 | 35 | 40 | 32 | 31 | 35 | 36 | 35 | 36 | | 3 | 31 | 24 | 100 | 96 | 106 |
| | P2Y12 | 23 | 27 | 26 | 21 | 23 | 23 | 22 | 21 | 30 | 19 | 35 | 25 | 26 | 29 | 26 | 25 | 26 | 3 | | 25 | 18 | 99 | 95 | 79 |
| B | CRF1 | 29 | 34 | 33 | 48 | 42 | 49 | 44 | 45 | 53 | 27 | 44 | 21 | 21 | 24 | 24 | 24 | 24 | 32 | 25 | | 23 | 85 | 61 | 81 |
| | GLR | 16 | 19 | 18 | 26 | 24 | 18 | 18 | 16 | 32 | 19 | 18 | 17 | 20 | 9 | 14 | 20 | 17 | 24 | 18 | 23 | | 90 | 91 | 66 |
| C | MGLU1 | 98 | 94 | 98 | 100 | 100 | 99 | 102 | 97 | 108 | 76 | 118 | 98 | 103 | 98 | 97 | 94 | 97 | 102 | 99 | 79 | 73 | | 0 | 105 |
| | MGLU5 | 95 | 90 | 92 | 96 | 98 | 95 | 97 | 92 | 102 | 94 | 113 | 93 | 99 | 93 | 93 | 93 | 94 | 98 | 95 | 56 | 74 | 0 | | 99 |
| F | SMO | 93 | 77 | 77 | 100 | 83 | 71 | 72 | 85 | 89 | 81 | 100 | 98 | 103 | 103 | 97 | 99 | 100 | 106 | 70 | 81 | 66 | 87 | 99 | |

*Comparing GRoSS to the reference alignments*

The GRoSS alignment refines the BW TM.50 residues extended from class A to classes B, C, and F. In terms of the notation used in [24], this would be referred to as TM.50a, denoting the use of class A as a reference. For each TM and each target—template pair we compute the relative offset of the BW residues in the HMM-HMM alignment. If the offset is 0, we use "_" as a label; if the offset is 9 or more residues, we use "9"; and if template BW residue maps to a loop or a wrong TM, we use "X". Table B shows the relative alignment of the BW residues between the GRoSS and the HMM-HMM alignments with the labels for the 7 TMs concatenated into one string.

When both the target and the template are from the same class, HMM-HMM aligns correctly all 7 BW .50 residues. When using class A templates for class B targets, HMM-HMM often gives the correct BW correspondence, and in some instances is off

by one turn (4 residues) on TM5. Class C gets TM7 off by 6 residues, which already constitutes too large of an error for homology models. The alignment between class A and SMO (class F) varies, but most often it disagrees by 4 residues on TM5 and 1 residue on TM7.

Table C shows the same comparison but between the GRoSS and the GPCRDB alignments. Here, classes A and B are aligned identically, class C differs by 1 residue on TM7 and class F differs on 3 TMs by 2 or 3 residues.

Table B. Relative alignment of BW residues between GRoSS and HMM-HMM.

| | | Mis-alignment of BW residues | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Template | | | | | | | | | | | | | | | | | | | | | |
| | | A | | | | | | | | | | | | | | | | | | B | | C | F |
| | Target | RHO | Beta1AR | Beta2AR | D3 | H1 | M2 | M3 | 5HT1B | 5HT2B | A2A | S1P1 | CXCR4 | CCR5 | KappaOR | MuOR | NOP | DeltaOR | PAR1 | P2Y12 | CRF1 | GIR | MGLU1 | MGLU5 | SMO |
| A | RHO | | | | | | | | | | | | | | | | | | | | | | 11_336 | 11_336 | 411 |
| | Beta1AR | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 1 |
| | Beta2AR | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 1 |
| | D3 | | | | | | | | | | | | | | | | | | | | 4 | | 11_X9X | 11_X9X | 4XX |
| | H1 | | | | | | | | | | | | | | | | | | | | 4 | | 11_136 | 11_136 | 1 |
| | M2 | | | | | | | | | | | | | | | | | | | | 4 | | 11_136 | 11_136 | 1 |
| | M3 | | | | | | | | | | | | | | | | | | | | 4 | | 11_136 | 11_136 | 1 |
| | 5HT1B | | | | | | | | | | | | | | | | | | | | 4 | | 11_136 | 11_136 | 451 |
| | 5HT2B | | | | | | | | | | | | | | | | | | | | 4 | | 11_136 | 11_136 | 1 |
| | A2A | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 4_1 |
| | S1P1 | | | | | | | | | | | | | | | | | | | | 4 | | 11_336 | 11_336 | 4_1 |
| | CXCR4 | | | | | | | | | | | | | | | | | | | | | | 11_336 | 11_336 | 4_1 |
| | CCR5 | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 431 |
| | KappaOR | | | | | | | | | | | | | | | | | | | | | | 11_336 | 11_336 | 4_1 |
| | MuOR | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 4_1 |
| | NOP | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 4_1 |
| | DeltaOR | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 4_1 |
| | PAR1 | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 4_1 |
| | P2Y12 | | | | | | | | | | | | | | | | | | | | | | 11_136 | 11_136 | 1 |
| B | CRF1 | | | | 4 | 4 | 4 | 4 | 4 | 4 | | 4 | | | | | | | | | | | 336 | 36 | 52 |
| | GIR | | | | | | | | | | | | | | | | | | | | | | 3_736 | 3_736 | 52 |
| C | MGLU1 | 11_336 | 11_136 | 11_136 | 11_999 | 11_136 | 11_136 | 11_136 | 11_136 | 11_136 | 11_1_6 | 11_336 | 11_X36 | 11_136 | 11_X36 | 11_136 | 11_136 | 11_136 | 11_136 | 11_136 | 336 | 3_7_6 | | | 1_3717 |
| | MGLU5 | 11_336 | 11_136 | 11_136 | 11_999 | 11_136 | 11_136 | 11_136 | 11_136 | 11_136 | 11_136 | 11_X36 | 11_136 | 11_136 | 11_X36 | 11_136 | 11_136 | 11_136 | 11_136 | 11_136 | 36 | 3_7_6 | | | 1_3717 |
| F | SMO | 4X1 | 1 | 1 | 499 | 1 | 1 | 1 | 431 | 1 | 1 | 4_1 | 431 | 431 | 4_1 | 431 | 4_1 | 4_1 | 4_1 | 1 | 9X | 4X | 1_719 | 1_3719 | |

Table C. Relative alignment of BW residues between GRoSS and GPCRDB.

| | | Mis-alignment of BW residues | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Template | | | | | | | | | | | | | | | | | | | | | |
| | | A | | | | | | | | | | | | | | | | | | B | | C | F |
| | Target | RHO | Beta1AR | Beta2AR | D3 | H1 | M2 | M3 | 5HT1B | 5HT2B | A2A | S1P1 | CXCR4 | CCR5 | KappaOR | MuOR | NOP | DeltaOR | PAR1 | P2Y12 | CRF1 | GIR | MGLU1 | MGLU5 | SMO |
| A | RHO | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | Beta1AR | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | Beta2AR | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | D3 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | H1 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | M2 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | M3 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | 5HT1B | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | 5HT2B | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | A2A | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | S1P1 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | CXCR4 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | CCR5 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | KappaOR | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | MuOR | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | NOP | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | DeltaOR | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | PAR1 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | P2Y12 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| B | CRF1 | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| | GIR | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | 2_33 |
| C | MGLU1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 2_32 |
| | MGLU5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | 2_32 |
| F | SMO | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_33 | 2_32 | 2_32 | |

GRoSS, HMM-HMM and GPCRDB agree on alignment of the BW .50 residues for all TMs within class A. In this case, the only differences between these alignments are the gaps in HMM-HMM. Some gaps can be present in both target and template at matching locations, which would simplify homology modeling. Table D shows the number of residues aligned to gaps for each target—template pair when HMM-HMM is used. Similarly to the notation in the previous tables, we label "_" when there are no gaps; "9" for 9 or more gaps; "Y" for misaligned BW residues; "X" for any template residues aligned to loop regions of the target.

Table E shows the number of residues aligned to gaps if the GPCRDB alignment is used. Most target—template pairs have at least one gap. However, the number of

gaps predicted by HMM-HMM (Table D) is larger and often falls at wrong positions, which disagree with GPCRDB (Table A).

Table D. Gaps in HMM-HMM: Number of residues in template TMs aligned to gaps in target sequence.

| | | Number of TM gaps | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Template | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | A | | | | | | | | | | | | | | | | | | | | B | | C | | F | |
| | Target | RHO | Beta1AR | Beta2AR | D3 | H1 | M2 | M3 | 5HT1B | 5HT2B | A2A | S1P1 | CXCR4 | CCR5 | KappaOR | MuOR | MOP | DeltaOR | PAR1 | P2Y12 | GIR | CRF1 | | MGLU1 | MGLU5 | SMO | |
| A | RHO | | | | | 1 | | | | 1 | | 32 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | 1_1 | 2 | YV4YYY | YV4YYY | 11_YYY | |
| | Beta1AR | | | | | 1 | | | | | | 32 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | 1 | | YV_YYY | YV_YYY | 2Y | |
| | Beta2AR | | | | 1 | 1 | | | | | | 32 | 1 | 1 | 1 | 1 | 1_1 | 1_1 | | 1 | | 1 | 1 | YV5YYY | YV4YYY | 1Y | |
| | D3 | | | 1 | | 11 | 2 | 1 | 1 | 12 | 1 | 32 | 1_21 | 1_2 | 1_2 | 1_2 | 1_2 | 1_2 | 1_2 | 1_2 | 1_2 | 1_1Y | 11_1 | YV4XYX | YV5XYX | 1_YXX | |
| | H1 | 1 | 1 | 1 | 11 | | 1 | 1 | | | 1 | 321 | 1_1 | 1_1 | 1_1 | 1 | 1_1 | 1_1 | 1 | 1 | 1_1 | 1_1Y | 2 | YV_YYY | YV3YYY | 1_11Y | |
| | M2 | | | | 1 | 1 | | | | | | 32 | 1_1 | 1_1 | 1 | 1 | 1_1 | 1_1 | 1 | 1 | 1_1Y | 22 | YV4YYY | YV4YYY | 1_11Y | |
| | M3 | | | | 1 | 1 | | | | | | 32 | 1_1 | 1_1 | 1_1 | 1_1 | 1_1 | 1_1 | 1 | 1 | 1_1Y | 22 | YV4YYY | YV4YYY | 1_11Y | |
| | 5HT1B | | | | 1 | | | | | | | 32 | 1_1 | 1_1 | 1 | 1 | 1_1 | 1 | 1 | 1 | 1_1Y | 1 | YV3YYY | YV3YYY | 2_YYY | |
| | 5HT2B | 1 | | | 12 | | | | | | 1 | 321 | 1_31 | 1_5 | 1_6 | 1_4 | 1_5 | 1_5 | 1_5 | 1_4 | 2Y | 2 | YV_YYY | YV1YYY | 1_11Y | |
| | A2A | | | | 1 | 1 | | | | 1 | | 32_1 | 2_1 | 1 | 1 | 1 | 1_1 | 1 | | 1 | | 3 | YV4YYY | YV4YYY | Y1Y | |
| | S1P1 | 32 | 32 | 32 | 32 | 321 | 32 | 32 | 32 | 321 | 32_1 | | 22_1 | 22 | 22_1 | 22 | 22 | 22_1 | 32_1 | 22_1 | 22_Y | 32_2 | YV_YYY | YV_YYY | 2_Y1Y | |
| | CXCR4 | 1 | 1 | 1 | 1_21 | 1_1 | 1_1 | 1_1 | 1_1 | 1_31 | 2_1 | 22_1 | | 2 | 1_2 | 2 | 2 | 2 | 2 | 1 | 1_1 | 1_3 | YV5YYY | YV5YYY | 1_Y3Y | |
| | CCR5 | 1 | 1 | 1 | 1_2 | 1_1 | 1_1 | 1_1 | 1_1 | 1_5 | 1 | 22 | 2 | | | 1 | 1 | 1 | | 2 | 1_1 | 11_3 | YV_YYY | YV_YYY | YYY | |
| | KappaOR | 1 | 1 | 1 | 1_2_1 | 1_1 | 1_1 | 1_1 | 1 | 1_6 | 1 | 22_1 | 1_2 | | | 1 | 1 | 1 | | 2 | | 3 | YV_YYY | YV_YYY | 1_Y1Y | |
| | MuOR | 1 | 1 | 1 | 1_2 | 1 | 1 | 1_1 | 1 | 1_4 | 1 | 22 | 2 | 1 | 1 | | | | | 1 | 1 | 1_2 | YV_YYY | YV_YYY | Y1Y | |
| | MOP | 1 | 1_1 | 1_1 | 1_2 | 1_1 | 1_1 | 1_1 | 1_1 | 1_5 | 1_1 | 22 | 2 | 1 | 1 | | | | 1 | 1 | 1_2 | YV_YYY | YV_YYY | 1_Y1Y | |
| | DeltaOR | 1 | 1 | 1_1 | 1_2 | 1_1 | 1_1 | 1_1 | 1 | 1_5 | 1 | 22_1 | 2 | 1 | 1 | | | | | 1 | 1 | 2 | YV_YYY | YV_YYY | Y1Y | |
| | PAR1 | | | | 1_2 | 1 | 1 | 1 | 1 | 1_5 | | 32_1 | 1 | | | | | | | 2 | 1 | 3 | YV4YYY | YV4YYY | 11_Y3Y | |
| | P2Y12 | 1 | 1 | 1 | 1_2 | 1_1 | 1 | 1_1 | 1 | 1_4 | 1 | 22_1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | | 1 | 1_13 | YV_YYY | YV_YYY | 1_G1Y | |
| B | CRF1 | 1_1 | 1 | 1 | 1_1Y | 1_1Y | 1_1Y | 1_Y | 1_1Y | 1_2V | | 12_Y | 1_1 | 1_1 | | 1 | | 1 | 1_1 | 1 | | | | 1_5YW | 1_53W | YY | |
| | GIR | 1 | | 1 | 11_1 | 2 | 22 | 22 | 1 | 2 | 3 | 12_2 | 1_3 | 11_2 | 3 | 1_2 | 1_2 | 11_2 | 2 | 1_13 | | | | Y4_5YW | YG_5YW | 21_YY | |
| C | MGLU1 | YV4YYY | YV_YYY | YY5YYY | YV4YYY | YV3YYY | YV4YYY | YV4YYY | YV3YYY | YV_YYY | YV4Y4Y | YV_YYY | YV_XYY | YV_YYY | YY_XYY | YV_YYY | YV_YYY | YV_YYY | YV4YYY | YV_YYY | 5YYY | Y3_5Y4Y | | | Y_YYW | | |
| | MGLU5 | YV4YYY | YV_YYY | YY4YYY | YV5YYY | YV3YYY | YV4YYY | YV4YYY | YV3YYY | YV_YYY | YV4YYY | YV_XYY | YV_YYY | YV_YYY | YY_XYY | YY_YYY | YV_YYY | YV_YYY | YV4YYY | YV_YYY | 5_3YY | Y5_5Y3Y | | | 1Y_YYYY | | |
| F | SMO | 12__YXY | ____1Y | ____1Y | ___YYY | 1__11Y | ____11Y | ____11Y | 2___YYY | 1__11Y | ____2_Y | 12_Y1Y | 1__YYY | ___YYY | 1__Y_Y | ___YYY | 1__Y_Y | ___Y1Y | 11__Y3Y | ___33Y | ___YX | 21___YX | | Y_1YYY | Y_YYYY | | |

Table E. Gaps in GPCRDB: Number of residues in template TMs aligned to gaps in target sequence.

| | | Number of TM gaps | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Template | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | A | | | | | | | | | | | | | | | | | | | | B | | C | | F | |
| | Target | RHO | Beta1AR | Beta2AR | D3 | H1 | M2 | M3 | 5HT1B | 5HT2B | A2A | S1P1 | CXCR4 | CCR5 | KappaOR | MuOR | MOP | DeltaOR | PAR1 | P2Y12 | GIR | CRF1 | | MGLU1 | MGLU5 | SMO | |
| A | RHO | | | | 1 | 1 | | | | 1_1 | 3 | 1_1 | 1_1 | 1_1 | 1 | 1 | 1 | 1 | 1_111 | 1_111 | 1_21 | 1_1 | 1_1Y | 1_1Y | Y_1YY | |
| | Beta1AR | | | | 1 | 1 | | | | 1_1 | 3 | 1_1 | 1_1 | 1_1 | 1 | 1 | 1 | 1 | 1_111 | 1_111 | 1_21 | 1_1 | 1_1Y | 1_1Y | Y_1YY | |
| | Beta2AR | | | | 1 | 1 | | | | 1_1 | 3 | 1_1 | 1_1 | 1_1 | 1 | 1 | 1 | 1 | 1_111 | 1_111 | 1_21 | 1_1 | 1_1Y | 1_1Y | Y_1YY | |
| | D3 | 1 | 1 | 1 | | 2 | 1 | 1 | | 1 | 3 | 1_11 | 1 | 1 | 1_1 | 1_1 | 1_1 | 1_1 | 1_11 | 1_211 | 1_31 | 1_11 | 1_11Y | 1_11Y | Y_11YY | |
| | H1 | 1 | 1 | 1 | 2 | | 1 | 1 | 1 | 1_2 | 13 | 1_31 | 1_2 | 1_2 | 1_1 | 1_1 | 1_1 | 1_1 | 1_211 | 1_211 | 1_31 | 1_11 | 1_11Y | 1_11Y | Y_11YY | |
| | M2 | | | | 1 | 1 | | | | 1_1 | 3 | 1_1 | 1_1 | 1_1 | 1 | 1 | 1 | 1 | 1_111 | 1_111 | 1_21 | 1_1 | 1_1Y | 1_1Y | Y_1YY | |
| | M3 | | | | 1 | 1 | | | | 1_1 | 3 | 1_1 | 1_1 | 1_1 | 1_1 | 1 | 1 | 1 | 1_111 | 1_111 | 1_21 | 1_1 | 1_1Y | 1_1Y | Y_1YY | |
| | 5HT1B | | | | | 1 | | | | 1 | 3 | 1_1 | 1 | 1 | 1 | 1 | 1 | 1 | 1_11 | 1_11 | 1_21 | 1_1 | 1_1Y | 1_1Y | Y_1YY | |
| | 5HT2B | 1_1 | 1_1 | 1_1 | 1 | 1_2 | 1_1 | 1_1 | 1 | | 1_3 | 11_11 | 11 | 11 | 11_1 | 11_1 | 11_1 | 11_1 | 11_11 | 11_211 | 11_31 | 11_11 | 11_11Y | 11_11Y | 1Y_11YY | |
| | A2A | 3 | 3 | 3 | 3 | 13 | 3 | 3 | 3 | 1_3 | | 1_2 | 1_3 | 1_3 | 1_3 | 1_3 | 1_3 | 1_3 | 1_21 | 1_21 | 1_21 | 1_2 | 1_2Y | 1_2Y | Y_2YY | |
| | S1P1 | 1_1 | 1_1 | 1_1 | 1_11 | 1_11 | 1_1 | 1_1 | 1_1 | 11_11 | 1_2 | | 11 | 11 | 1 | 1 | 1 | 1 | 1_1 | 1 | 2 | | Y | | Y_YY | |
| | CXCR4 | 1_1 | 1_1 | 1_1 | 1 | 1_2 | 1_1 | 1_1 | 1 | 11 | 1_3 | 11 | | | 1 | 1 | 1 | 1 | 11 | 211 | 31 | 11 | 11_Y | 11_Y | Y_11YY | |
| | CCR5 | 1_1 | 1_1 | 1_1 | 1 | 1_2 | 1_1 | 1_1 | 1 | 11 | 1_3 | 11 | | | 1 | 1 | 1 | 1 | 11 | 211 | 31 | 11 | 11_Y | 11_Y | Y_11YY | |
| | KappaOR | 1 | 1 | 1 | 1_1 | 1_1 | 1 | 1 | 1 | 11_1 | 1_3 | 1 | 1 | 1 | | | | | 111 | 111 | 21 | 1 | 1_Y | 1_Y | Y_1YY | |
| | MuOR | 1 | 1 | 1 | 1_1 | 1_1 | 1 | 1 | 1 | 11_1 | 1_3 | 1 | 1 | 1 | | | | | 111 | 111 | 21 | 1 | 1_Y | 1_Y | Y_1YY | |
| | MOP | 1 | 1 | 1 | 1_1 | 1_1 | 1 | 1 | 1 | 11_1 | 1_3 | 1 | 1 | 1 | | | | | 111 | 111 | 21 | 1 | 1_Y | 1_Y | Y_1YY | |
| | DeltaOR | 1 | 1 | 1 | 1_1 | 1_1 | 1 | 1 | 1 | 11_1 | 1_3 | 1 | 1 | 1 | | | | | 111 | 111 | 21 | 1 | 1_Y | 1_Y | Y_1YY | |
| | PAR1 | 1_111 | 1_111 | 1_111 | 1_11 | 1_211 | 1_111 | 1_111 | 1_11 | 11_11 | 1_21 | 1_1 | 11 | 11 | 111 | 111 | 111 | 111 | | 2 | 3_1 | 1_1 | 1_1W | 1_1W | Y_1_YY | |
| | P2Y12 | 1_111 | 1_111 | 1_111 | 1_211 | 1_211 | 1_111 | 1_111 | 1_11 | 11_211 | 1_21 | 1 | 211 | 211 | 111 | 111 | 111 | 111 | 2 | | 1_1 | 1_1 | 1_1W | 1_1W | Y_1_YY | |
| B | CRF1 | 1_21 | 1_21 | 1_21 | 1_31 | 1_31 | 1_21 | 1_21 | 1_21 | 11_31 | 1_21 | 2 | 31 | 31 | 21 | 21 | 21 | 21 | 3_1 | 3_1 | | 2 | 2_Y | 2_Y | Y_2_YY | | |
| | GIR | 1_1 | 1_1 | 1_1 | 1_11 | 1_11 | 1_1 | 1_1 | 1_1 | 11_11 | 1_2 | | 11 | 11 | 1 | 1 | 1 | 1 | 1_1 | 1_1 | 2 | | Y | Y | Y_YY | |
| C | MGLU1 | 1_1_Y | 1_1_Y | 1_1_Y | 1_11_Y | 1_11_Y | 1_1_Y | 1_1_Y | 1_1_Y | 11_11_Y | 1_2_Y | | 11_Y | 11_Y | 1_Y | 1_Y | 1_Y | 1_Y | 1_1W | 1_1W | 2_Y | Y | | | Y_YY | | |
| | MGLU5 | 1_1_Y | 1_1_Y | 1_1_Y | 1_11_Y | 1_11_Y | 1_1_Y | 1_1_Y | 1_1_Y | 11_11_Y | 1_2_Y | | 11_Y | 11_Y | 1_Y | 1_Y | 1_Y | 1_Y | 1_1W | 1_1W | 2_Y | Y | | | Y_YY | | |
| F | SMO | V__1W | V__1W | V__WY | V_11W | V_11W | V__1W | V__1W | V__WY | 1V_11W | V__2W | V__YY | Y_11WY | V_11W | V__1W | V__1W | V__1W | V__WY | V_1_VY | V_1_W | V_2_YY | V__YY | Y__YY | V__YY | | |

## Geometrical quality of homology models

Each alignment can be used to produce a homology model for a given target—template pair. For the following analysis, we constructed simple homology models for the backbone atoms only. We ignored any missing residues, which were gaps in the target—template alignment. For all pairs considered in the previous tables, we evaluated RMSD, TM-score, and the number of common inter-helical contacts. The results of these three measures comparing HMM-HMM with GRoSS are shown in Figure A panels 1, 2, and 3. The same measures comparing GPCRDB with GRoSS are shown in Figure A panels 4, 5, and 6.

The RMSD comparisons (lower number is better) show that GRoSS alignment outperforms HMM-HMM for essentially all cases (Fig. A1). The RMSD comparisons of GRoSS to GPCRDB (Fig. A4), show that GRoSS outperforms GPCRDB in cross-class cases and only slightly underperforms in intra-class cases. This is expected as

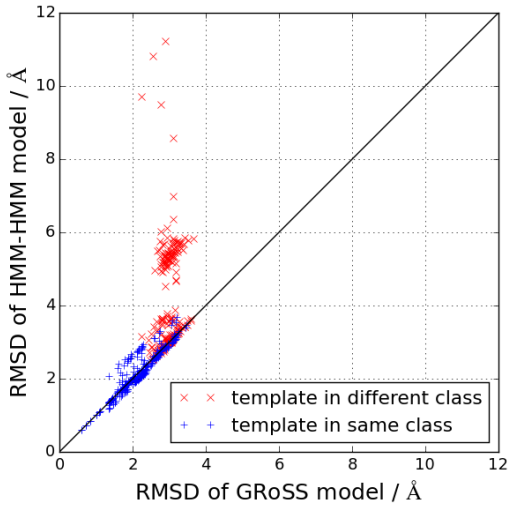
GPCRDB alignments include gaps/bulges based on pairwise structural comparison input, whereas GRoSS alignment ignores these gaps and bulges. TM-Score comparisons (higher number is better) show similar results as RMSD comparisons for GRoSS versus HMM-HMM (Fig. A2) and GRoSS versus GPCRDB (Fig. A5).

The comparison of the number of conserved contacts (higher number is better) shows that HMM-HMM performs slightly better than GRoSS (Fig. A3) for intra-class cases, but fails for inter-class cases. Same is true for GPCRDB comparison with GRoSS (Fig. A6).

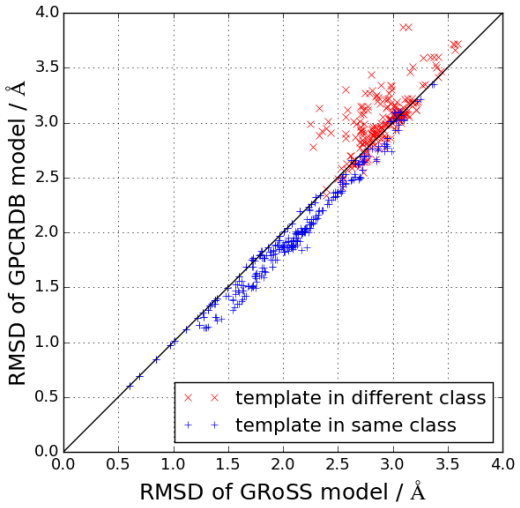
Overall, these comparisons show that GRoSS alignments perform better than both HMM-HMM and GPCRDB. The cross-class sequence alignments for GRoSS are significantly better, whereas the intra-class sequence alignments are of similar quality.

Figure A. Comparing geometrical quality of the homology models.

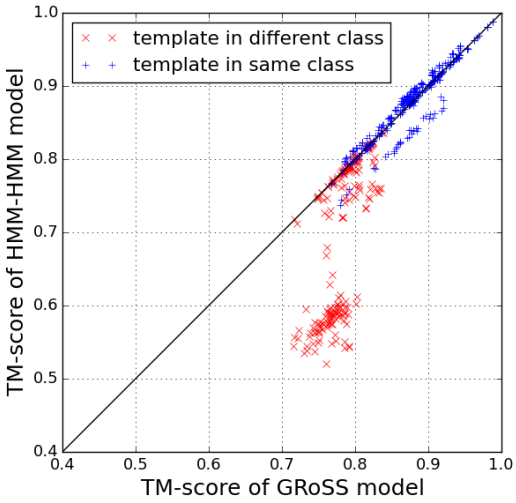
1



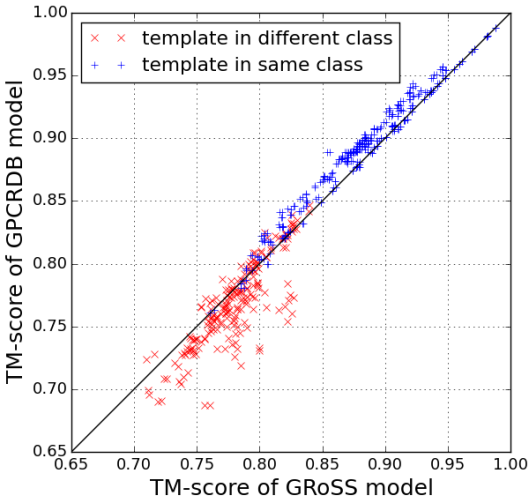

4



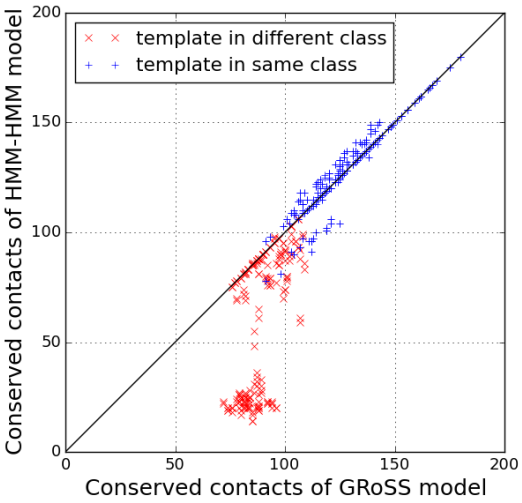

2




5




3




6