Chapter 2

Prediction of the 3D Structure for FMRF-amide Peptides Bound to Mouse MrgC11 Receptor with Subsequent Experimental Verification[1]

2.1 Introduction

G protein-coupled receptors (GPCRs) play an essential role in cell communications and sensory functions as mentioned in chapter 1. Consequently they are involved in wide variety of diseases and are targets for many drug therapies. Particularly important is the large number of orphan GPCRs (for which the native ligands remain unknown), which may play important, albeit unknown, functions in various cells. To understand their respective physiological roles, it is important to identify their endogenous ligands, and to find small molecule ligands that would serve as selective agonists or antagonists. One example here is the family of GPCRs called the Mas-related gene (Mrg) receptor for mouse or the sensory neuron specific receptor (SNSR) in mouse and human[1, 2]. A subset of these receptors including mMrgC11 and mMrgA1 is localized mainly to isolectin $B4^{+}$, the small diameter nociceptors in the dorsal root ganglia (DRG). Dong et al. showed that some of these receptors were activated by RFamide neuropeptides such as NPFF and NPAF and suggested them to be involved in pain sensation or modulation[1]. These Mrg receptors have been paired with structurally diverse transmitter peptides[3].

[1] Portions of this chapter have been submitted to the *Journal of Medical Chemistry* for publication.

Clearly deorphanization would be greatly aided by having three-dimensional (3D) structures of the orphan receptors to help select the most promising new ligands for experimental assays, but it is not yet possible to obtain experimental 3D structures for human GPCRs. Consequently our group developed the MembStruk computational method[4, 5] to predict such structures and we demonstrate in this study that the predicted structures are sufficiently accurate to predict binding sites and relative binding energies. Previously MembStruk was applied to several GPCRs, obtaining ligand binding sites in excellent agreement with experiments. However in these studies the structural data were known prior to our calculations. Although any experimental data were not utilized in making our predictions, such validations are not completely convincing. We undertook this study on mMrgC11 and mMrgA1 receptors for the specific purpose of validating the MembStruk method. Thus prior to our calculations there were no data on how mutations affect binding. In addition the experiments had shown that the F-M-R-F-NH_2, (D)F-M-R-F-NH_2 and F-(D)M-R-F-NH_2 tetrapeptides activate mMrgC11 receptor at ~100 nM concentration, while F-M-(D)R-F-NH_2 and F-M-R-(D)F-NH_2 are inactive (>10 μM). We assumed that explaining such an effect of chirality on binding should provide a strong test of the predicted structures.

2.2 Computational methods

All energy and force calculations were done using DREIDING force field (FF)[6] with the charges from CHARMM22[7] FF and were executed in the molecular dynamics program, MPSIM[8]. The cell multipole method[9] was used for the calculation of nonbond interaction. Unless otherwise specified all simulations were performed in gas phase with the dielectric constant of 2.5.

2.2.1 Structure predictions of the Mrg receptor

The 3D structure of the mMrgC11 and mMrgA1 receptors were predicted independently using MembStruk (version 4.05)[10]. The details of the MembStruk (version 3.5) were described

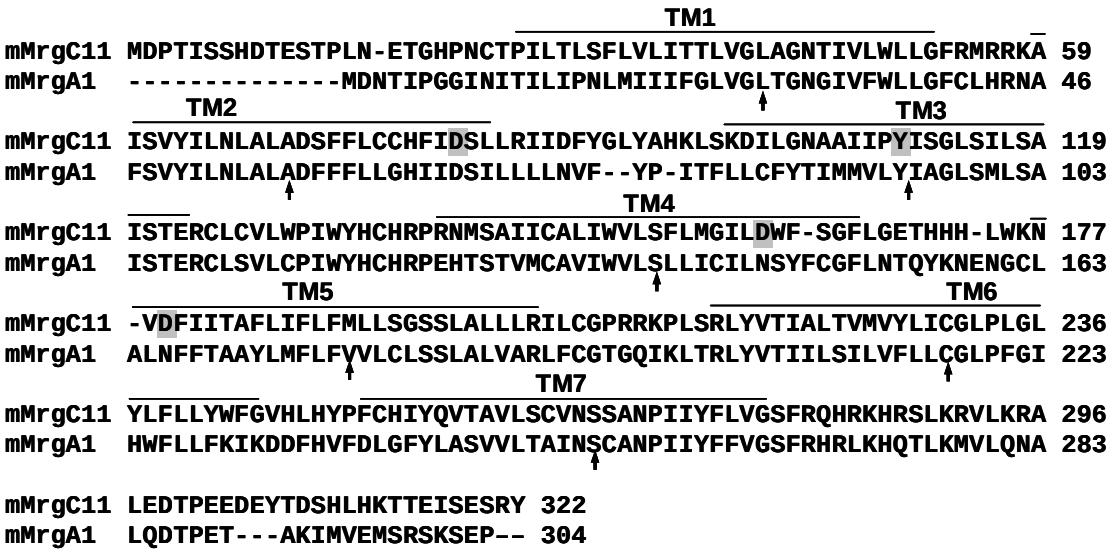

Figure 2.1 Predicted transmembrane (TM) regions. The sequence alignment of mMrgC11 and mMrgA1 is based on the alignment with the entire set of sequences obtained by BLAST search with the mMrgC11 sequence (see Fig. S2.1). The hydrophobic center of each TM is indicated with an arrow. The residues involved in the mutagenesis experiment are highlighted.

in reference 10. Here we outline the procedure, highlighting aspects also relevant to Mrg receptors or that were improved in version 4.05.

Prediction of transmembrane regions

The transmembrane (TM) regions and the hydrophobic maximum for each TM helix were predicted using the TM2ndS method[5]. We used NCBI BLAST[11] to search the non redundant protein database to find sequences homologous to the mMrgC11 receptor with bit scores greater than 200. These 27 sequence hits had sequence identities to mMrgC11 ranging from 41% to 88%. This set of sequences included the mMrgA1 receptor whose sequence identity to mMrgC11 is 44%. Twenty-two of these 27 sequences belong to Mrg receptor family, with remaining 5 corresponding to unnamed GPCRs. We then carried out a multiple sequence alignment with these 27 sequences using ClustalW[12]. These results (Fig. S2.1) were used as input to TM2ndS. The hydrophobicity profile (Fig. S2.2) resulting from TM2ndS had no clear separation between TM2 and TM3, leading to uncertainty in the boundaries between TM2 and TM3. A similar ambiguity

was observed between TM6 and TM7. To eliminate such problems, the MembStruk 4.05 procedure calculates the hydrophobicity profile from a second round of seven TM predictions in which each sequence of the core of the seven TM (15 amino acids around the hydrophobic center) was used as a template. This second set of independent BLAST searches was executed under high gap penalty with each TM core. Here we selected GPCR sequences with sequence identities of >50% (the identity with the entire sequence of mMrgC11 was as low as 23%), see table S1. Then a second round of TM predictions was performed using the multiple sequence alignment of these 7 sets of sequences, see Table S1. The final refined TM region and its hydrophobic center for each of the 7 TM domains were determined from this second round of prediction. For mMrgA1 we used the same TM regions as assigned from alignment with mMrgC11.

Assembly of TM helical bundle

For each TM domain we built canonical α-helices with fully extended conformation of side chains. These were assembled such that the 7 predicted hydrophobic centers are all in the xy plane with the x and y coordinates adapted from the 7.5 Å electron density map of frog rhodopsin[13]. Each helix oriented about its axis so that its hydrophobic moment pointed away from the center of the seven helices (toward the membrane). The tilt of each helix with respect to the z axis and its azimuthal angle were adapted from the 7.5 Å electron density map of frog rhodopsin[13].

Then we carried out 200 ps of molecular dynamics (MD) at 300 K without solvent or lipid, but with charged side chains neutralized by adding Na^+ or Cl^- ions. This allows the conformation of each individual helix to bend or kink as appropriate. We then selected the snapshot with the lowest potential energy from the last 100 ps of the MD trajectory and the net hydrophobic moment was calculated for the middle 15 residues around the hydrophobic center for each helix using this conformation. Each helix was rotated again so that its hydrophobic moment faces toward the membrane. This hydrophobicity-based rotation works well for the six TM helices with

extensive contacts to lipid bilayers. Moreover, since the optimal orientation of a helix depends on the relative orientations of the neighboring helices, we often carry out a combinatorial rotation of the 7 helices. However we found that the structure predicted by the above process placed the highly conserved Asn44-Asp71 pair between TM1 and TM2 and the Asn66-Trp151 pair between TM2 and TM4 close enough to form hydrogen bonds (based only on the coarse hydrophobicity-based rotation step). Therefore we carried out extensive 360º rotational orientation optimization only for TM3, TM5 and TM6. Here the rotational angle of TM3 was scanned for 360 º (in 30 º increments) because TM3 has the least surface area exposed to lipid, but TM5 and 6 were rotated only over the range of -60º to 60º since the orientation had already been optimized roughly using the hydrophobic moment. For every rotation we reassigned the side chain conformation using SCWRL3.0[14] before energy-minimization. The orientation with the best energy was then selected. The results of these scans are shown in Table S2. The rotational orientation of TM7 was scanned over 360º in 5º increments, where for each angle all atoms were optimized. In fact the initial orientation showed the best energy.

Rigid body dynamics in lipid bilayers and addition of loops

Next we added two layers of explicit lipid molecules (52 molecules of dilauroylphosphatidyl choline (DPC) lipid) surrounding the TM bundle. The initial structures for the lipid DPC these were based on the crystal structure in Cambridge Structural Database (ID: LAPETM10). To achieve proper packing of the TM helices, the 7-helix-lipid complex was optimized using rigid body MD for 50 ps where each helix and lipid molecule was treated as a rigid body, with just 6 degrees of freedom (translation and rotation).

The conformation of each TM helix was further optimized in the lipid environment with full atom Cartesian MD simulation for 50 ps while the coordinates of lipid molecules were kept fixed. Then we carried out an additional equilibration of the whole system for 40 ps and selected the structure with the lowest potential energy. For this structure each side chain conformation was

re-assigned using SCWRL and the bundle (helices plus lipid) was minimized to an RMS force of 0.5 (kcal/mol)/Å using conjugate gradients.

The loops were added to the helices using MODELLER6v2[15]. The side chains were re-assigned using SCWRL and subsequently a full atom conjugate gradient minimization of the receptor was performed.

In many GPCRs (including bovine rhodopsin and the catechol amine receptors, such as dopamine and adrenergic receptors) there are conserved cysteines near the top of TM3 and in the second extracellular loop (EC2) that are expected to form a disulfide bond leading to a closed loop. However the mMrgC11 and mMrgA1 receptors do *not* contain such cysteines so the loops were allowed to remain in an open conformation. From five loop structures generated with MODELLER6v2 we selected the one with the lowest internal strain and then optimized the coordinates using annealing MD while keeping the coordinates of TM helices fixed. In this process the system was heated from 50 K to 600 K and cooled down back to 50 K in 50 K steps, with 1 ps of equilibration between the temperature jumps. At the end of the annealing cycle the structure was fully optimized using the conjugate gradients. This final structure shown in Figure 2.3 (top and side views) was used for all docking studies.

2.2.2 Docking predictions with peptide ligands

Using the 3D structure of the mMrgC11 structure we used a refined version (MSCDock) of the docking procedure described in Cho et al.[16]. Since peptide ligands are highly flexible we modified the step in HierDock2.0 (described in Vaidehi et al.[4]), involving scan of the entire receptor with RFa to locate the binding site. This hierarchical docking protocol to predict the binding sites for various ligands as used in this study is described below.

Scanning of the binding sites

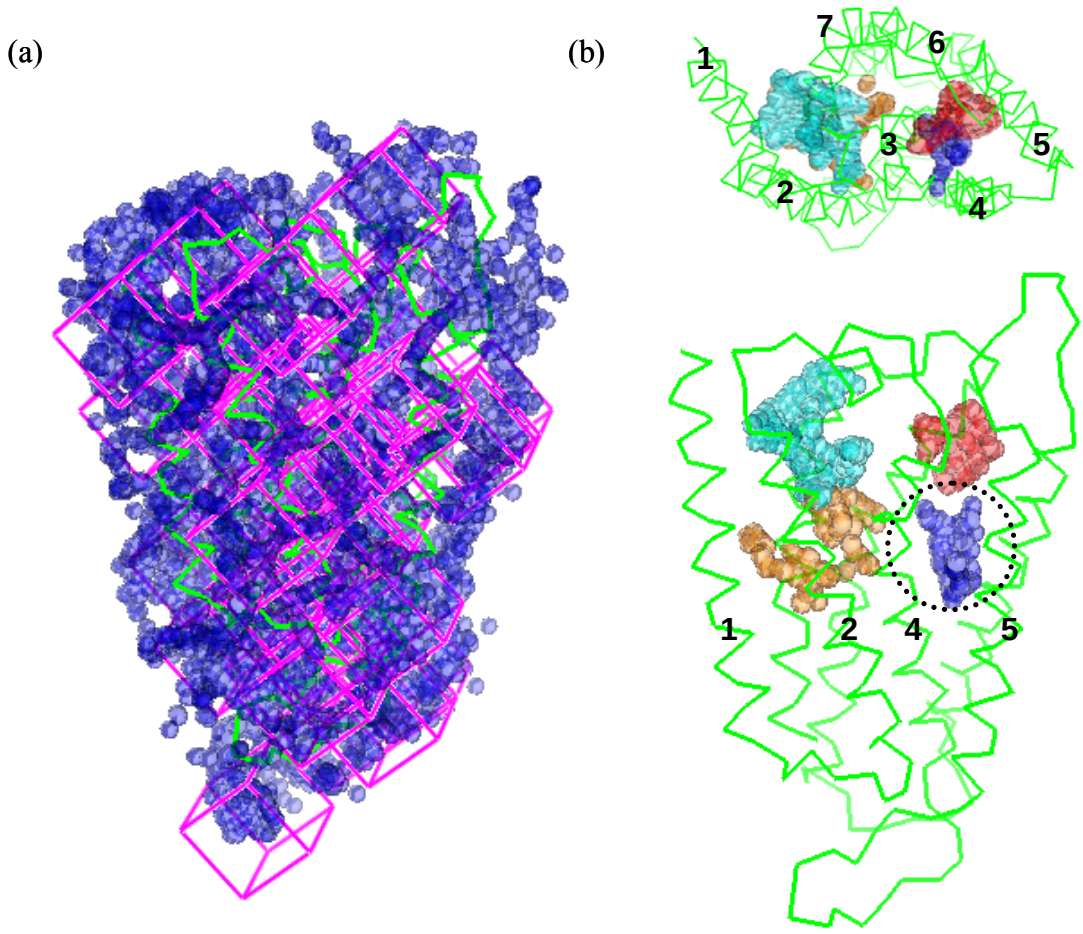

Figure 2.2 The scanning regions used to determine the binding sites for the mMrgC11 receptor. (a) The 9186 Spheres generated with SPHGEN to fill the void spaces of the receptor. The 40 cubic boxes used for docking are shown. (b) The four regions pre-selected for the docking studies. The region enclosed by the dotted circle was identified as the best site.

The entire receptor structure was scanned with the Arg-Phe-NH_2 (RFa) dipeptide known to agonize the receptor (EC50 = 460 nM) to locate the putative binding site. First, the molecular surface was created using the autoMS utility in DOCK4.0[17] with the default values for surface density (3.0dots/$Å^2$) and probe radius (1.4Å). Then we generated spheres from each that filled the void space in the receptor. To do this we used SPHGEN in DOCK4.0. We then constructed a total of 40 cubic boxes (sides of 10Å) and spaced by 8Å that covered this set of spheres. Assuming that the ligand binds inside the TM bundle from the extracellular region, we analyzed these

spheres using a buried surface criterion to pre-select four non-peripheral regions, located on the upper half of the receptor as shown in Figure 2.2. The spheres inside each box were used to define the docking region as input to DOCK4.0.

The RFa ligand was docked independently into each of the four regions as follows. Since the peptide ligands have a significant number of independent dihedral angles (the smallest dipeptide, RFa, contains 10 torsions), we wanted to ensure that this extensive conformation space is sampled in the docking. Thus for each peptide ligand we used the Metropolis Monte Carlo (MC) Method in Cerius2[18] (with a MC temperature of 5000 K in vacuum) to generate a set of 1,000 low energy conformations having a diversity of CRMS = 1.0 Å.

Level0: Then for each of the 1,000 conformers we used DOCK4.0 to generate a set of 3,000 configurations within each of the four binding regions of the receptor. From these we selected the 100 best configurations for each of the 1,000 conformers based on the DOCK score. This led to a total of 100*1000 = 100,000 configurations which were combined together and saved for the next scoring step. In these configurational searches, the rigid ligand and torsion drive options in DOCK4.0 were used. The bump filter option was turned on (maximum bump = 10) and the reduced (to 75%) van der Waals radius was used.

Level1: The configurations from level0 with a ligand-buried surface area below 65% were discarded and the remaining configurations were ranked by the number of hydrogen bonds between receptor and ligand, then by the percentage of buried surface area, and then by DOCK4.0 energy score. This ordered list was trimmed using a diversity criterion of CRMS = 0.6 Å and the top 100 configurations selected. Each of these was minimized in MPSIM using 100 steps of conjugate gradient method, while the receptor coordinates were fixed.

Level2: The 10 best configurations by energy were selected from level1 and the full ligand-protein complex was minimized in MPSIM with 100 steps. The side chain rotamers for all the residues within 5 Å of the ligand were reassigned using the SCREAM side chain replacement

program [Kam, Vaidehi and Goddard unpublished], which uses a side chain rotamer library of 1,478 rotamers with a diversity of 1.0 Å in coordinates.

Level3: The binding energies were then calculated for these 10 optimized ligand-receptor complex configurations. The calculated binding energy (BE) is defined by

BE = E (ligand in fixed protein) – E (ligand in water)

where the E (ligand in fixed protein is the potential energy of the ligand calculated in the ligand-receptor complex with the coordinates of the receptor fixed. This potential energy includes the internal energy of the ligand and the interaction energy of the ligand with the receptor. E (ligand in water) is the potential energy of the free ligand in its docked conformation (snap bind energy) and its solvation energy calculated using the analytical volume generalized born (AVGB) continuum solvation method[19]. In these calculations the dielectric constant was set to 78.2 for the exterior region and to 1.3 for the interior region. The final best ligand-receptor structure was selected as the one with the most negative binding energy.

Among four regions we found that the RFa ligand had the best binding energy in the region involving TM3, 4, 5 and 6 (blue in Fig. 2.2(b)), which we call the putative binding site. The best structure of the RFa-receptor complex was further refined using one cycle of annealing MD heating from 50 K to 600 K and cooling down back to 50 K in 50 K steps, with 1 ps of equilibration between the temperature jumps. Here only the ligand and the residues within 10 Å of the binding pocket (including backbone atoms) were allowed to move during the annealing cycle. At the end of the annealing cycle, the system was minimized to an RMS force of 0.3 (kcal/mol)/Å and the side chains of the residues in the receptor within 4 Å from the ligand was reassigned again with SCREAM. The spheres for docking other peptide ligands were defined with this final optimized RFa-receptor complex.

Docking of other peptide ligands

For the five F-M-R-F tetra-peptide stereoisomers, we first docked the R-F amide part of the C-terminal. This motif is common to most peptide agonists of mMrgC11[20]. Indeed the efficacy results for the five chirally modified F-M-R-F-amides (see Fig. 2.11) show that the chirality of the R-F part dramatically affects activation. Therefore we first docked the three dipeptides: acetylated R-F-NH_2, (D)R-F-NH_2 and R-(D)F-NH_2. Then we used these as an anchor in building the remaining F-M amino acids to construct the docked tetrapeptide.

For docking the three dipeptides, we used the Dock-Diversity Completeness protocol (DDCP) described in Cho et al.[16] to generate a set of diverse configurations and improve completeness in searching the configurations in DOCK (Level0). Briefly, DDCP attempts to generate a complete set of ligand configurations families with a fixed coordinate diversity (1.0 Å). Completeness is defined as the point where the fraction of new configuration that belong to previously generated families to the fraction that leads to a new family is 2.2 (but restricted the list to 5000 families). Then we selected the 50 families with the best energies (by DOCK4.0 energy score) and continued generating configurations while keeping only those that belonged to one of these 50 families until there was an average of six members in each family. Then 50 family heads (best energy in each family) were conjugate gradient minimized (100 steps or 0.1 kcal/mol/Å of RMS force) with the ligand atoms movable and the receptor atoms fixed. Then the 10 best scoring ligands (one from each family by binding energy) were selected for further side chain optimization. Here the binding energy was calculated as the difference between the energy of the ligand in the fixed receptor and the energy of the ligand in solution. The energy of the free ligand was calculated for the docked conformation and its solvation energy was calculated using surface generalized Born model (SGB)[21]. The side chain rotamers of the residues in the receptor within 5 Å of the bound ligand were reassigned by using the SCREAM side chain replacement program. After side chain optimization, the final 10 complex structures were minimized (100 steps or 0.1 kcal/mol/Å of RMS force) with all atoms movable.

The above docking procedure was applied to each of the 1,000 conformers of each peptide ligand generated using the MC Method in Cerius2 (with a MC temperature of 5,000 K in vacuum) using diversity of CRMS = 2.0 Å. Prior to docking the structures of the 1,000 conformers were minimized in gas phase and ordered by energy. Then they were re-clustered with the diversity of 2.0 Å and the conformer of each family head (the best energy among the family) was chosen for docking. This led at least 10 family heads for the R-F dipeptides and over 20 family heads for acetylated R-F dipeptides.

For each such structure the docking process ends up with 10 structures for the ligand/protein complex. Thus we obtained ~100 structures for the dipeptide and ~200 for acetylated peptides. The number of hydrogen bonds (intermolecular between receptor and ligand and intramolecular for a ligand) was calculated for each structure of each ligand/protein complex. This was combined with the binding energy and the number of hydrogen bonds to select the final best structure.

The final structure of ligand-receptor complex obtained from the hierarchical docking procedure was further refined by annealing MD as described in section 2.1.3. Here only the ligand and the side chains of residues within 3.5 Å of the binding pocket were allowed to move. At the end of the annealing cycle, the system was minimized to an RMS force of 0.1 (kcal/mol)/Å.

Building the terminal F-M residues from the bound acetylated R-F-NH_2

The conformations of the terminal F-M residues were sampled using moleculeGL, a recursive, Metropolis Monte Carlo-based rotamer design technique [Kekenes-Huskey, Vaidehi and Goddard in preparation] from the R-F-NH_2 dipeptide docked in mMrgC11 receptor where the extracellular loops were removed. Either the psi angle of Met or the phi angle of Arg is defined as an anchor. We used moleculeGL to generate 1000 structures for the terminal FM, using a diversity of 1.0. Then we selected the lowest energy conformation and minimized the ligand structure (0.3 kcal/mol/Å of RMS force) with the coordinates of receptor fixed. Then the side

chain rotamers of residues within 5 Å of the ligand were assigned using SCREAM and the structure of the whole complex was minimized. The final best structure was refined by annealing as described in previous section.

2.3 Experimental procedures

2.3.1 *In vitro* mutagenesis

The point mutation was incorporated into mMrgC11-GFP coding sequence in pcDNA3.1/Zeo (+) plasmid (Invitrogen) using the QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). The mutagenic oligonucleotide primers were synthesized and purified in the oligonucleotide synthesis center of Caltech. All mutant constructs were verified by DNA sequencing. Later the wild type and mutant gene in pcDNA3.1/Zeo (+) were sub-cloned into pcDNA5/FRT expression vector (Invitrogen) for stably expressing cell lines.

2.3.2 Cell culture and transfection

Flp-In™-293 cells (Invitrogen) were co-transfected with mMrgC11-GFP gene in pcDNA5/FRT vector and pOG44 plasmid (Invitrogen) using FuGENE-6 reagent (Roche Applied Science) according to the manufacturer's instructions. The cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), penicillin/streptomycin and L-glutamine. The cells were split into fresh medium 48 h after transfection and then selected with 400 μg/ml of hygromycin. After two weeks of selection period the hygromycin-resistant clones were picked and then maintained in the selective medium with 200 μg/ml of hygromycin.

2.3.3 Biotinylation and immunoprecipitation

Flp-In™-293 cells stably expressing wild type and mutant receptors were placed into 10 cm culture dish coated with poly-L-lysine and cultured for 24h. The cells were washed twice with ice-cold PBS and incubated with 3 mL of 0.5 mg/mL Sulfo-NHS-LC-Biotin (Pierce

Biotechnology) in PBS supplemented with 0.1 mM HEPES, pH 7.5 at room temperature for 30 min. The biotinylation reaction was quenched by washing cells three times with Tris-buffered saline (10 mM Tris pH 7.5, 154 mM NaCl). The washed cells were incubated with 5 mL of cold lysis buffer (10 mM HEPES, pH 7.4, 1 mM EGTA) supplemented with 100 μM 4-(2-aminoethyl)-benzene sulfonyl fluoride hydrochloride at 4 °C for 15 min. Cells were scraped from the dish and homogenized with a Dounce homogenizer (20-25 strokes with a tight pestle). The cell lysate was centrifuged at 750x g for 10min at 4 °C to remove the nuclei and cell debris. The resulting supernatant was centrifuged at 75,000x g for 30 min at 4 °C. The membrane pellet was solubilized in 500 μL of ice-cold TX/G buffer (300 mM NaCl, 1% TX-100, 10% Glycerol, 1.5 mM $MgCl_2$, 1 mM $CaCl_2$, 50 nM Tris pH 7.4, 0.5 mM PMSF, protease inhibitor cocktail) and incubated with gentle mixing at 4 °C for 1h. Insoluble material was removed by centrifugation at 10,000x g for 15 min at 4 °C. The protein concentration was estimated using the *DC* Protein Assay Kit (Bio-Rad).

The solubilized protein was incubated with 50 μL of streptavidin-agarose (Pierce Biotechnology) overnight at 4 °C on an inversion wheel. The streptavidin-agarose was washed four times with ice-cold TX/G buffer in absence of protease inhibitor and then twice with ice-cold PBS. The precipitates were resuspended with protein sample buffer and then boiled for 15 min. The protein sample was analyzed by SDS-PAGE and transferred to nitrocellulose membrane. The membrane was blocked in Tris-buffered saline with 0.1% Tween 20 containing 5% non-fat milk for 1h. GFP-tagged mMrgC11 receptors were detected by blotting with anti-GFP polyclonal primary antibody (Molecular Probes) in blocking solution followed by anti-rabbit horseradish peroxidase-conjugated secondary antibody and an ECL detection kit (Amersham Biosciences).

2.3.4 Intracellular calcium assay

The cells were placed into 96-well cell culture plate coated with MATRIGEL matrix (BD Biosciences). After 16-24 h, the cells were washed twice with Hank's balanced salt solution

supplemented with 10 mM D-glucose, 20 mM HEPES and 1.6 mM NaOH (assay buffer) and loaded with 2 μM fura-2/AM (Molecular Probes) in assay buffer at room temperature for 20min. Then the cells were washed four times with assay buffer to get rid of the residual fura-2/AM present outsides cell membranes. The fluorometric imaging plate reader (FLIPR) assay was carried out at various concentrations of peptide ligands (1 nM to 10 μM) with the FlexStation II system (Molecular Devices). The fluorescence emitted from the excitation at 340 nm and 380 nm was measured respectively along the time and the ratio of emission at two excitation wavelengths was evaluated together. The difference between maximum and minimum value of the ratio was plotted along with the logarithm of the ligand concentration. The curve was fitted with ORIGIN6.0 software to compute EC_{50} value.

2.4 Results and discussion

2.4.1 Characteristics of the predicted mMrgC11 receptor structure

The predicted TM regions for mMrgC11 are given in Figure 2.1 and the predicted 3D structure of the mMrgC11 receptor is shown in Figure 2.3. TM6 is bent by 28º at Pro233 and TM7 is bent by 15º at Pro271. These two prolines are highly conserved over all family A GPCRs including rhosopsin (in rhodopsin TM 6 and TM7 are bent by 24º and 33º, respectively). Moreover, Pro109 in the middle of TM3 leads to bending of 23º (in rhodopsin TM3 is bent by 13º). We find that these distortions lead to a cavity lined by TM3, TM5, and TM6 that provides the space required for binding our tetrapeptides. The remaining four TMs have relatively straight α-helical conformations.

The predicted 3-D structure of mMrgC11 receptor is superimposed with the 2.2 Å X-ray crystal structure of bovine rhodopsin[22] in Figure 2.3. Here each TM between mMrgC11 and rhodopsin was aligned separately with Clustal-W, imposing a high gap penalty and only the TM

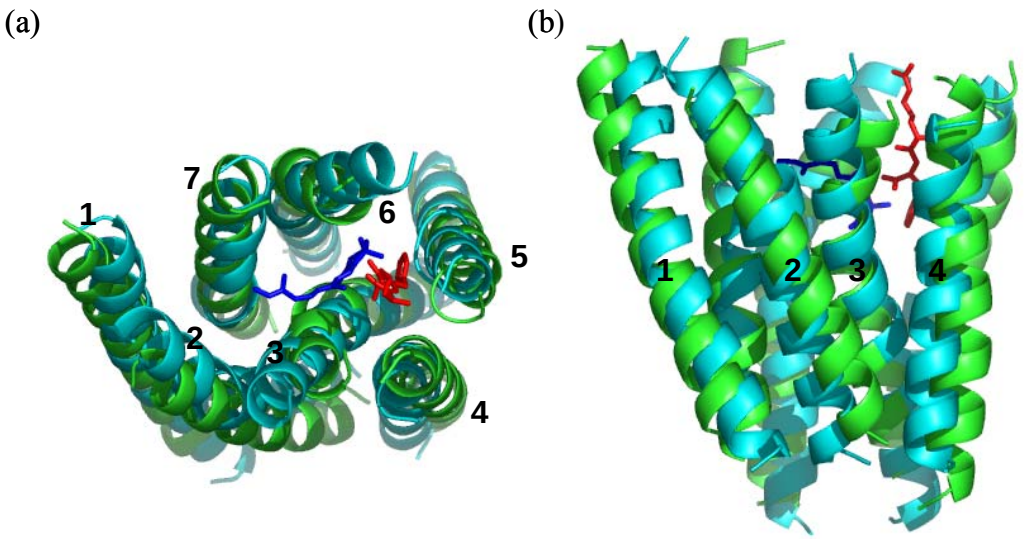

Figure 2.3 Comparison of the predicted 3D structure for the RFa/mMrgC11 complex (green) with the X-ray crystal structure of retinal/rhodopsin (PDB code: 1U19, 2.2 Å resolution). The RFa dipeptide is colored red while the retinal is blue. (a) top view (from extracellular region); (b) side view (with EC at the top). As expected by the low sequence identity (22% for the TM regions) there are significant differences. The CRMS difference in the Cα atoms is 3.75 Å.

regions were fitted with each other for superposition. The sequence identity between the TM regions is ~22%, averaged over the seven TM region sequences. The RMSD in coordinates (CRMSD) of the Cα atoms in the TM regions between bovine rhodopsin and mMrgC11 is 3.75 Å. As expected from the low sequence identity the structures are rather different, but they share such structural features as the kink in the TM6 and TM7 helices. Indeed TM3 of rhodopsin has a slight kink at the two consecutive glycines present at the same position as the proline in mMrgC11.

Several conserved residues participate in the inter-helical hydrogen bonds that maintain the stability of the mMrgC11 receptor structure just as in the rhodopsin crystal structure. Thus Asn44 (TM1) (highly conserved in the family A GPCRs) forms a hydrogen bond with the Ser268 carbonyl group of the backbone in TM7 as shown in Figure 2.4. Asp71 (TM2) forms an interhelical hydrogen bond with this Asn in rhodopsin is in the proximity, but is not in hydrogen bond contact in the mMrgC11 receptor. Such differences are plausible since Miura and Karnik reported TM2 movement from activation in angiotensin II type 1 receptor (using substituted

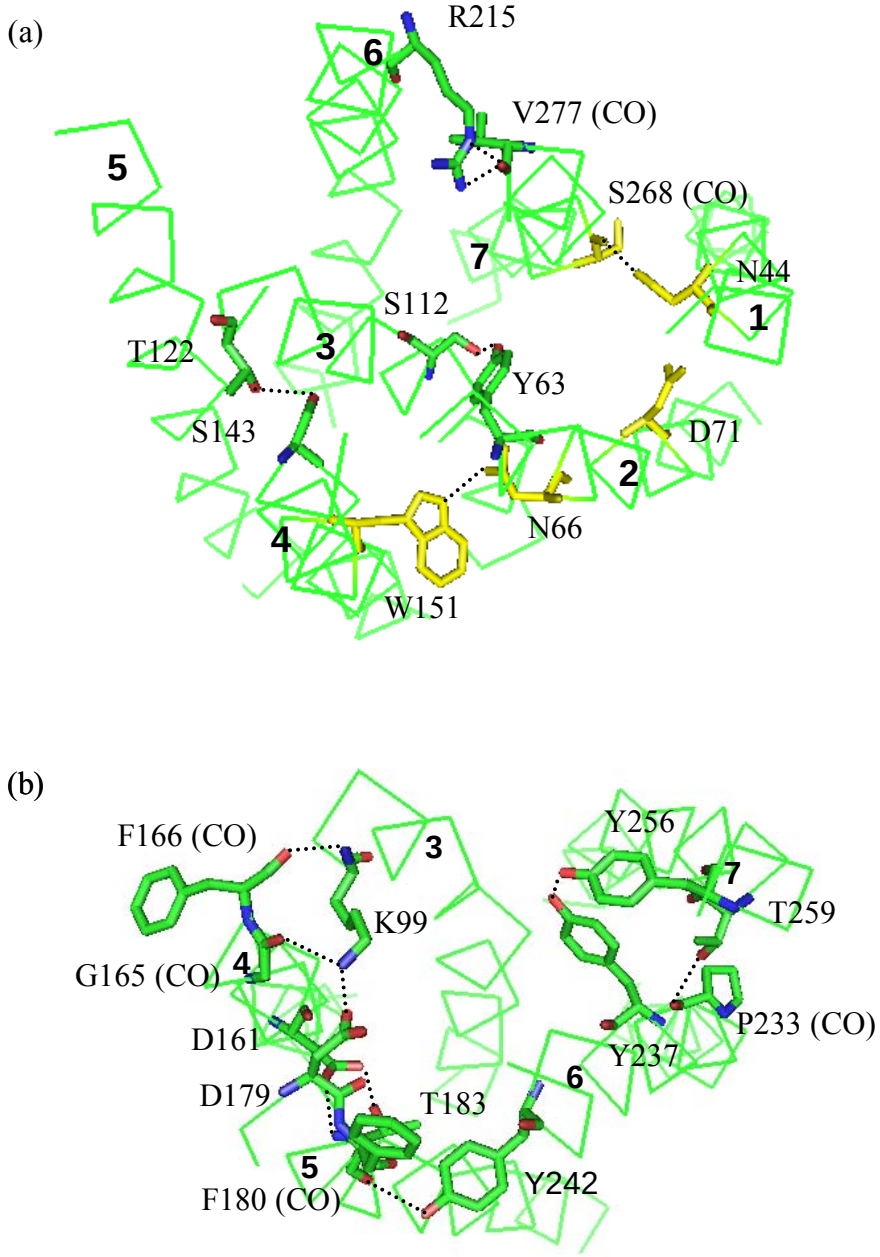

Figure 2.4 Interhelical hydrogen bond networks in the mMrgC11 receptor. The interhelical hydrogen bonds (dashed lines) are specified with residues participating in hydrogen bonds. The highly conserved residues in the family A of GPCRs that form interhelical hydrogen bonds in rhodopsin are colored by yellow. (a) Viewed from the intracellular region. (b) Viewed from the extracellular region. The HBPLUS[23] program was used to calculate hydrogen bonds (maximum D-A distance = 3.9 Å, minimum D-H-A angle = 90.0°).

cystein accessibility mapping)[24]. Thus Asp in TM2 might interact differently, compared to one in the inactive rhodopsin structure. The Asn66 (TM2)–Trp151 (TM4) pair does form a hydrogen bond just like the analogous pair in rhodopsin.

Important points to note in the structure are:

Tyr63 (TM2) (one of residues conserved in the Mrg receptor family (with 39 sequences available on Swiss-Prot and TrEMBL)) participates in hydrogen bonding with Ser112 (TM3) as shown in Figure 2.4.

Another conserved residue, Ser143 (TM4) forms a hydrogen bond with the hydroxyl group in Thr122 (TM3) as shown in Figure 2.4.

Arg215 (TM6) contacts with the backbone carbonyl group of Val277 in TM7, as shown in Figure 2.4.

Asp179 (TM5), which is identified as a key residue for the ligand binding in this study is in contact with Lys99 (TM3) in the apo protein. Asp161 (TM4) also interacts with Thr183 (TM5) in the absence of a ligand.

Several other inter-helical hydrogen bonds are formed with non-conserved hydrophilic residues. Most of these are found in the regions of the TM regions near the intracellular loop. These regions pack more compactly than the near-extracellular regions as appropriate for ligand binding.

No direct contact between TM3 and TM6 or between TM3 and TM7 is found in the TM regions. However, these TM helices interact with each other through well-stacked aromatic rings as shown in Figure 2.5. Tyr110 (TM3), one of the aromatic residues participating in these interactions is conserved through the Mrg receptors (5 of 39 have Phe at this position instead of Tyr). Also Trp265 in TM6 known to be responsible in activating rhodopsin is replaced with Gly

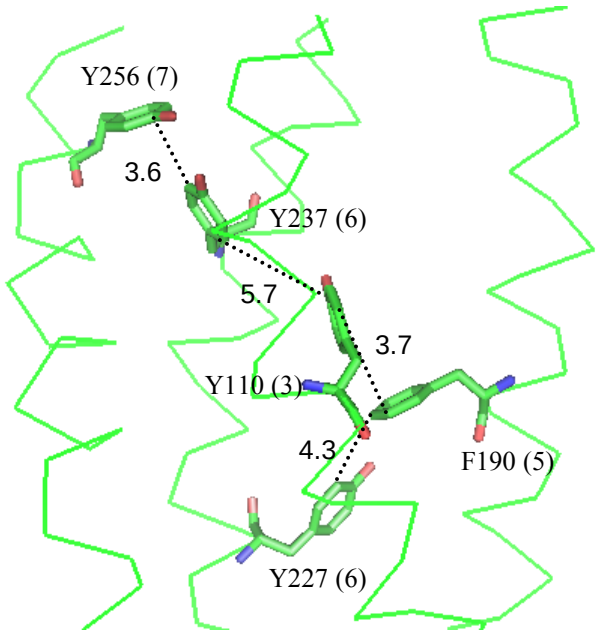

Figure 2.5 Aromatic interactions in TM regions of mMrgC11 receptor. Aromatic residues involved in the π-stacking through TM3, 5, 6 and 7 are shown with the closest C–C distance between two benzyl rings (Å).

in the mMrgC11 receptor. Thus activation in mMrgC11 might involve a different mechanism. As discussed in section 4.2 we find that the agonists to MrgC11 bind in the pocket located between TM3, 4, 5 and 6, which might affect the aromatic–aromatic interactions to help induce activation.

2.4.2 Description of the peptide binding sites

The predicted RFa binding site is located between TM3, TM4, TM5 and TM6 as shown in Figure 2.3. In contrast to 11-cis retinal in rhodopsin, we find that RFa orients vertically in the binding pocket. As seen in Figure 2.3, the aromatic rings stacked between TM3 and TM6 confines the ligand to the region between TM3, TM4, TM5 and TM6. A similar binding orientation has been suggested for the formylated peptide, *f*MLF[25], which binds parallel to the helix in the formyl peptide receptor (FPR). Since RFa is a small peptide ligand (like *f*MLF) it can be placed parallel in the pocket but for longer peptides, the additional amino acids might be kinked towards TM2 and TM7, having contact with these TMs mainly in the loop regions.

Predicted binding site of the dipeptides

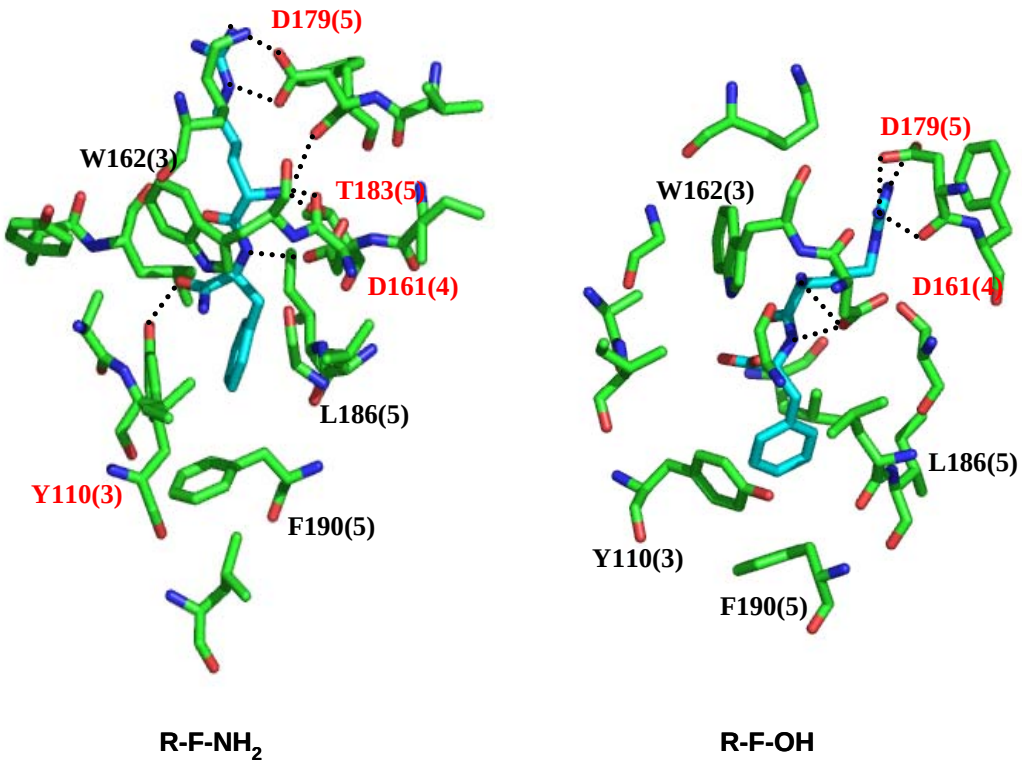

Figure 2.6 Predicted 5 Å binding pocket of the RFa and RF dipeptide agonists. The intermolecular hydrogen bonds calculated with explicit hydrogens using the same criteria as in Figure 2.4 are indicated by the dotted lines. A residue whose side chain participates in the hydrogen bond is specified in red, while one whose backbone is involved is in blue. The residues showing good hydrophobic interactions are specified in black. The top of the picture corresponds to the extracellular regions.

The detailed interactions of bound dipeptides with mMrgC11 receptors are described in Figure 2.6. The binding mode of R-F-OH (RF) is similar to R-F-NH_2 (RFa) although the side chain rotamers of certain residues are different. The common features are that the positively charged moieties are stabilized through the salt bridges and other hydrophilic interactions. Thus the Arg has a good electrostatic interaction with Asp179 (TM5) and the N-terminus has good electrostatic interaction with Asp161 (TM4). The N-terminus of RFa also forms a hydrogen bond with the hydroxyl group of Thr183 (TM5). In addition the C-terminus of RFa makes a hydrogen bond with the hydroxyl group of Tyr110 (TM3).

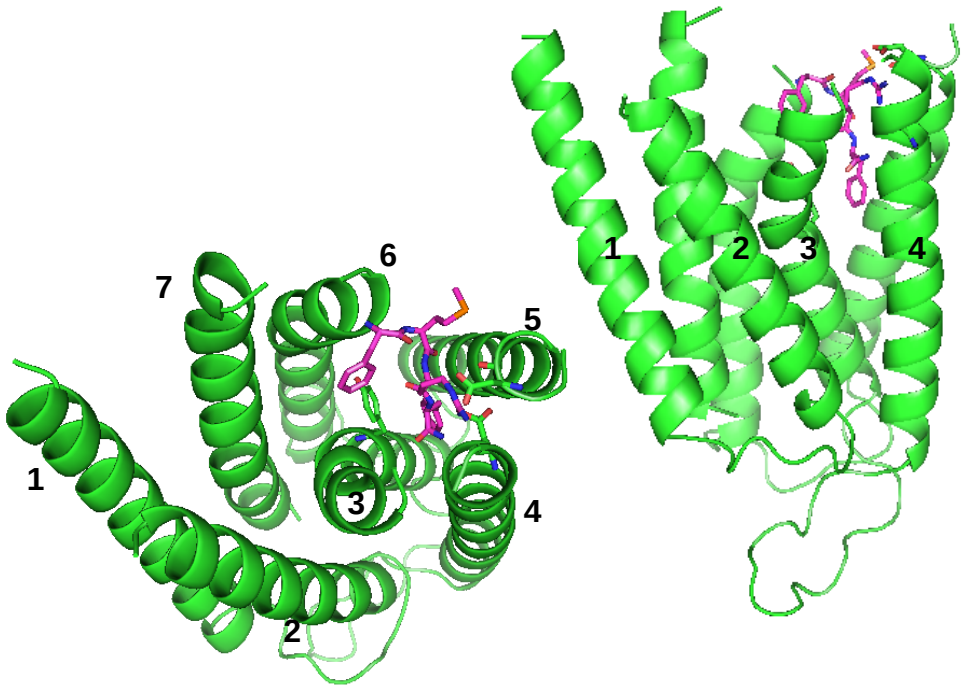

Figure 2.7 Predicted 3D structure for the FMRFa/mMrgC11 complex. The Cα atoms in the TM regions are traced in cartoon while the three key residues (Y110, D161, and D179) are shown in stick. The top view is from extracellular (EC) region and in the side view the EC region is at the top.

The phenyl group of the Phe is stabilized by several aromatic residues present in the binding pocket. Tyr110 interacts most closely with Phe of both dipeptides. For RFa the phenyl ring is in a sandwiched geometry with Tyr110 while for RF these two rings have the displaced T-shape. Phe190 (TM5) also has a good π-π interaction with Phe of the ligand, while Leu186 (TM5) also contributes a good hydrophobic environment for Phe.

Predicted binding sites of the tetrapeptide agonists, F-M-R-F-NH_2, (D)F-M-R-F-NH_2 and F-(D)M-R-F-NH_2

Three tetra-peptides known to be good agonists for mMrgC11[20] were docked into the binding region identified for RFa. The common C-terminal dipeptide part, which is parallel to the average helical axis with the C-terminus of the peptide toward the intracellular region, is bound similarly to RFa (or RF). The extra F-M peptide stretches out horizontally toward TM6 as shown

in Figure 2.7 for the F-(D)M-R-F-NH_2 (FdMRFa) case, where the chirality of Met is modified to be left-handed. In FdMRFa, the amide group of the C-terminus forms hydrogen bonds with the side chain of Asp161 (TM4) and the backbone carbonyl group of Gly158 (TM4). Phe at the C-terminus resides in good aromatic and hydrophobic environment formed by Tyr110 (TM3), Phe190 (TM5) and Leu186 (TM5). Arg is stabilized through the electrostatic interactions with Asp161 (TM4) and Asp179 (TM5). Thr183 (TM5) also interacts with the side chain of Arg. Asp161 (TM4) forms a hydrogen bond with a nitrogen atom of the backbone. Met located in the peripheral region between TM5 and TM6 is nearby such hydrophobic residues as Leu238 (TM6), Phe239 (TM6) and Ile187 (TM5), but has no specific interaction. The N-terminal Phe is sandwiched between Trp162 (TM4) and Tyr237 (TM6), leading to good aromatic interactions. The N-terminus is exposed to the extracellular region. Thus for longer peptide agonists the extra residues might be added starting from this N-terminal position. This might account for the binding of Met-Enk-RF-amide. This is all shown in Figure 2.8.

In F-M-R-F-NH_2 (FMRFa), the overall binding mode is similar to FdMRFa. Some differences are that Thr183 (TM5) no longer participates in the hydrogen bonding with the peptide and the side chain of the right-handed Met is closer to TM5 and interacts at the edge of aromatic ring of Phe180 (S–C distance = 4.0 Å). The preference of S atoms at the edge of aromatic ring has been observed in the study of the non-bond interaction involving sulfur atom of Met by analyzing the protein crystal structures[26].

(D)F-M-R-F-NH_2 (dFMRFa) shows similar interactions. Although the N-terminal Phe has a different chirality from the previous two ligands, it has a similar conformation of the side chain and fits in between Trp162 (TM4) and Tyr237 (TM6). In this case the Met leads to an intra-residue S…O interaction and an inter-residue interaction with Leu240, where the sulfur atom behaves as an electrophile [26].

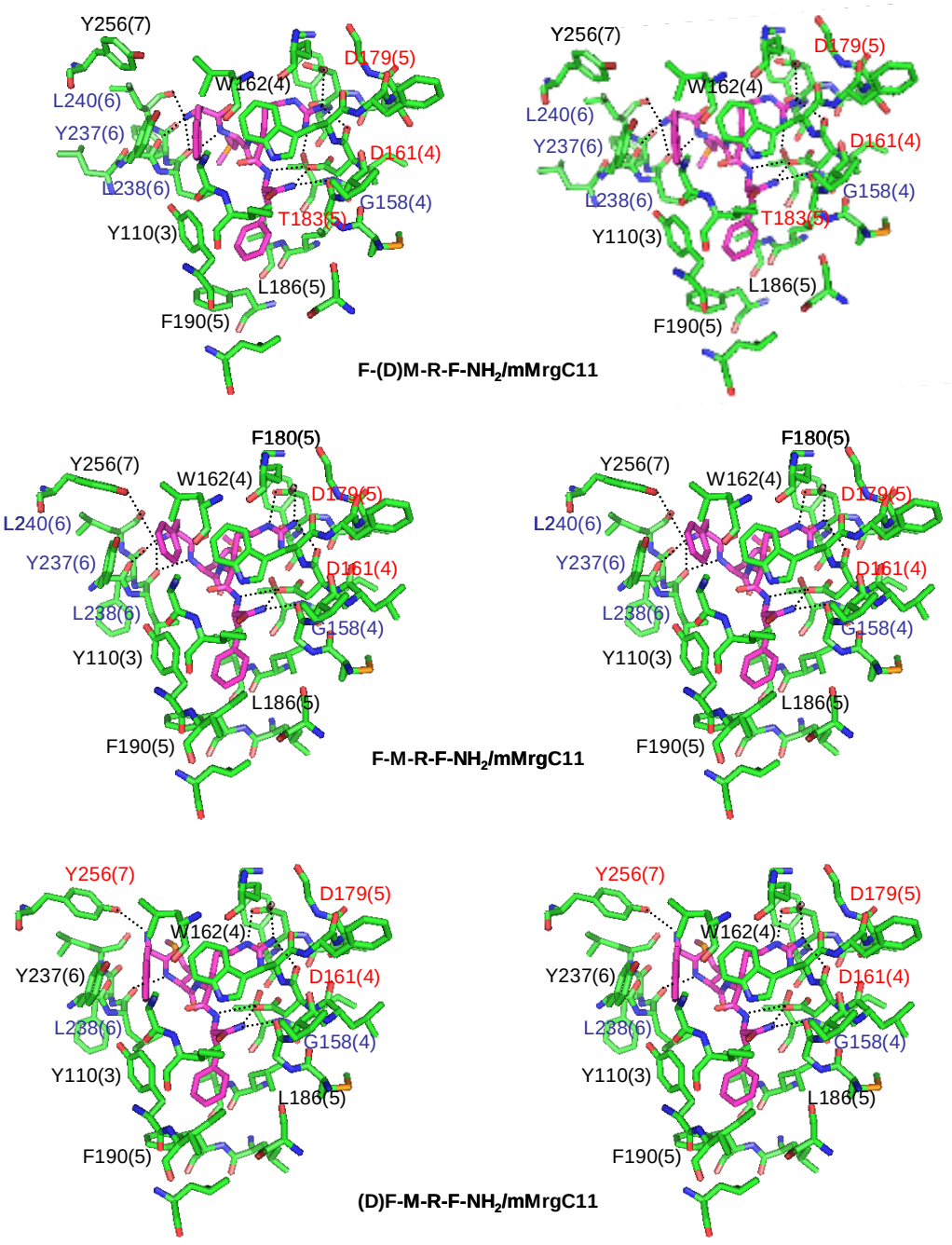

Figure 2.8 Predicted 5 Å binding site to mMrgC11 of the agonist tetra-peptides, F-(D)M-R-F-NH_2, F-M-R-F-NH_2 and (D)F-M-R-F-NH_2.

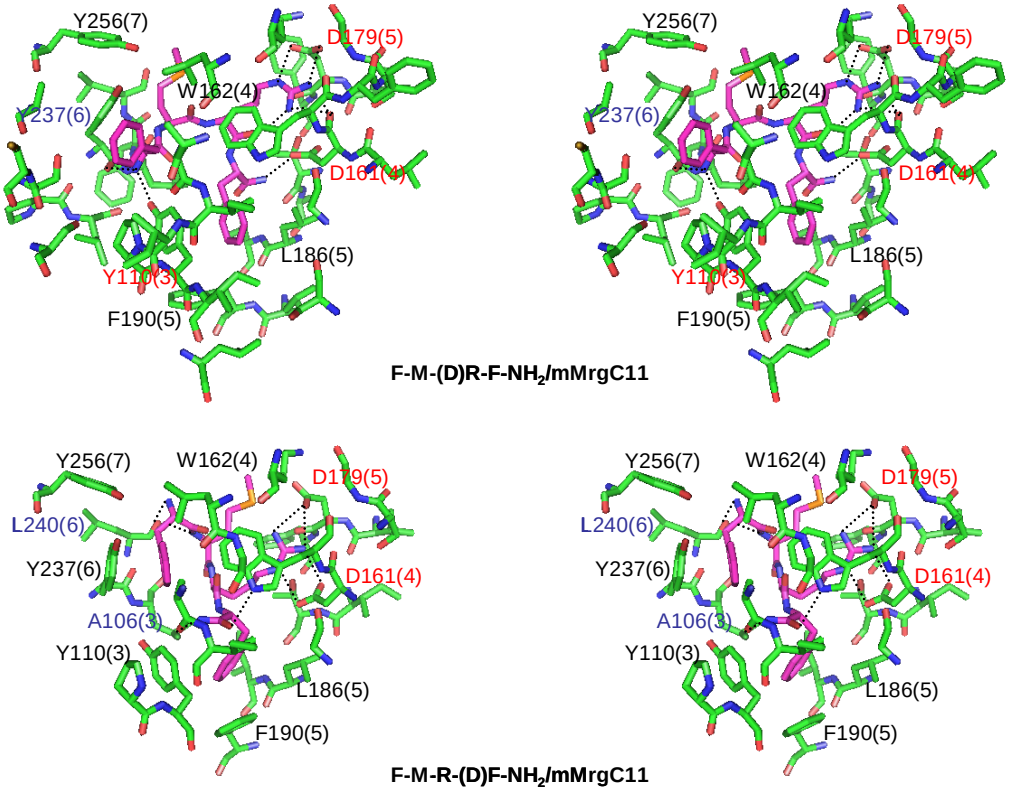

Figure 2.9 Predicted 5 Å binding pocket of the non-agonist tetra-peptides, F-M-(D)R-F-NH_2 and F-M-R-(D)F-NH_2 (neither case was observed experimental to bind even at 30 μM).

We calculate FdMRFa to bind strongest, with FMRFa and dFMRFa having binding energies just 7% and 11% weaker.

Predicted binding sites of the non-agonists, F-M-(D)R-F-NH_2 and F-M-R-(D)F-NH_2

The two other chirally modified FMRFa peptides, FMdRFa and FMRdFa, do not agonize mMrgC11. Our predicted 5 Å binding sites for them are shown in Figure 2.9. In both cases, the C-terminal Phe interacts with Tyr110 (TM3) and Phe190 (TM5) as seen for other agonists.

In F-M-(D)R-F-NH_2 (FMdRFa) the side chain of Arg is located near Asp161 (TM4) and Asp179 (TM5), with good electrostatic interactions. However the contact is less tight and the non-bond interaction energies with Asp161 and Asp179 decrease by 41% and 12% respectively,

compared with FdMRFa. We see the intra-residue S…O interaction for Met in this case. The N-terminal Phe loses π-π interaction with Trp162 (TM4).

In the other non-agonist, F-M-R-(D)F-NH_2 (FMRdFa), the side chain of Arg is between Asp161 (TM4) and Asp179 (TM5) and the interaction is weaker than in FdMRFa. The sulfur of Met shows the interaction with the backbone carbonyl group of Asp179. The N-terminal Phe is sandwiched with Trp162 (TM4) and Tyr237 (TM6). Overall these two non-agonist peptides show the similar binding characteristics to the agonist peptides, but the interaction energy is much weaker by 34% for FMdRFa and by 32% for FMRdFa, compared with FdMRFa.

Summary of binding sites

This study identified several residues critical for peptide binding in mMrgC11. The two aspartic acids, Asp161 (TM4) and Asp179 (TM5) contribute to good electrostatic interactions for the electropositive groups of the ligands; Arg for tetrapeptides and Arg and N-terminus for dipeptides. Several aromatic residues contribute to good π-π interactions. Tyr110 (TM3) and Phe190 (TM5) contact with the common C-terminal Phe of all five agonists. Tyr110 is highly conserved across MRG family of receptors. In the tetrapeptide agonists, the additional phenyl group interacts with Trp162 (TM4) and Tyr237 (TM6). As mentioned previously, these aromatic residues are well stacked in the receptor in the absence of a ligand and provide the interhelical interactions among TM3, TM5, TM6 and TM7. This coupling with two phenyl groups of the tetrapeptide ligand along with the strong electrostatic interaction of Arg with Asp161 (TM4) and Asp179 (TM5) is likely to induce the conformational change responsible for the activation.

2.4.3 Mutagenesis experimental results

Based on the predictions described above, we expect that Tyr110 (TM3) (highly conserved aromatic residue among Mrg family), Asp161 (TM4), and Asp179 (TM5) are all critical to binding. Thus we embarked on a series of mutation experiments to validate theses predictions.

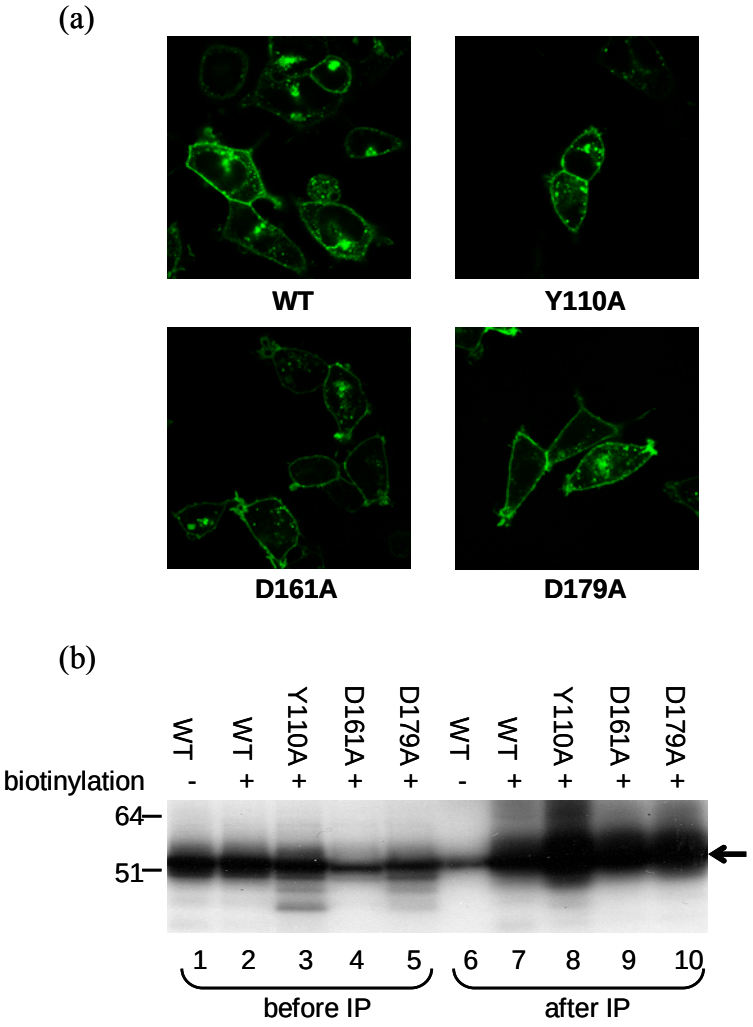

Figure 2.10 The expression of mMrgC11 wild type and mutant receptors in the Flp-In293 cells. (a) GFP images of wild type and mutant cells. The GFP was fused into the C-terminus of the receptor. (b) Biotinylation of the cell surface where receptors are localized and folded. The biotinylated cell extract is blotted with anti-GFP after immunoprecipitation (IP) with streptavidin. Lanes 1-5 are before IP and Lanes 6-10 are after IP. The molecular weight markers are shown on the left in kDa.

Table 2.1 The EC50 values of various peptide ligands determined by the intracellular calcium release assay with Flp-In293 cells expressing mMrgC11 receptor

Peptide	Sequence	EC50, nM	Han et. al[20]
RF	RF	1255 ± 239	632 ± 124
RFa	RF-NH_2	682 ± 186	460 ± 35
FMRF	FMRF	666 ± 228	544 ± 117
FMRFa	FMRF-NH_2	168 ± 26	114 ± 32
dFMRFa	(D)F-M-R-F-NH_2	276 ± 56	108 ± 1
FdMRFa	F-(D)M-R-F-NH_2	113 ± 18	11 ± 4
FMdRFa	F-M-(D)R-F-NH_2	inactive	inactive
FMRdFa	F-M-R-(D)F-NH_2	inactive	inactive
Bam15	VGRPEWWMDYQKRYG	292 ± 19	53 ± 2
γ1-MSH	YVMGHFRWDRF-NH_2	398 ± 189	17 ± 3
γ2-MSH	YVMGHFRWDRFG	340 ± 66	11 ± 5
NPFF	FLFQPQRF-NH_2	358 ± 25	54 ± 5

Inactive means that no activation was detected up to the highest concentration tested, 10μM.
Data represent the mean (±SEM) of four independent experiments.

Expression and localization of mMrgC11 wild type and mutant receptors

Based on the predictions, we carried out three sets of experiments in which key residues were mutated to alanine – Tyr110Ala, Asp161Ala and Asp179Ala. Figure 2.10(a) shows the GFP images for mMrgC11 wild type and for the three mutant receptors. All mutant cells show fluorescence signals as intense as the wild type and the cell boundaries are clearly identified. These images indicate that the mutant receptors are expressed at level similar to the wild type and are well localized at the cell membranes.

To determine whether the mutants properly fold across the cell membrane, we combined immunoprecipitation (IP) experiments with biotinylation. Lanes 1-5 in Figure 2.10(b) show total mMrgC11 receptor proteins including ones that are not biotinylated but present in cytosol and those that have not crossed properly through the membranes. These blots indicate again that all

three mutants are well expressed in the cells, although the expression levels of D161A and D179A mutants seem slightly lower. The results of blots after IP with streptavidin (lanes 6-10) show that the mutant receptors localized on the cell membranes take apical positions at similar amounts to the wild type. Since the band corresponding to the non-specific binding of streptavidin (lane 6) is much weaker, we conclude that the major portions of blots in lane 7-10 come from the biotin-specific binding. This suggests the mutant proteins folds properly on the membranes as well as the wild type protein.

Dose-dependent intracellular calcium release assay with stably expressed MrgC11 receptors

Table 2.1 shows the EC50 values of various peptide ligands determined by intracellular calcium assay experiment with Flp-In293 cells expressing the mMrgC11 receptor. The di- and tetra-peptides and some longer peptide agonists were selected from the ligands previously identified by Han *et al*.[20]. We obtained slightly higher EC50 values in our cellular system, compared to the previous measurements. This difference might result from a variety of sources such as different coupling efficiencies, different expression levels of receptor, and different cellular environment[27]. Nonetheless, the selectivity observed in this study is consistent with the previous results– for example; FMdRFa and FMRdFa still show no activity.

Out of the twelve ligands tested for the wild type receptor, we selected the six most potent ligands to measure the potencies for Y110A, D161A and D179A mutant receptors, as shown in Table 2.2.

We find that the Y110A mutant is not activated by any of the six tested ligands up to a concentration of 33 μM, indicating that Y110 is critical for binding and activation.

The D179A mutants show no potency for the three tetrapeptide ligands, while the other three are activated only under 10 times higher concentration of the ligand.

Table 2.2 Binding constants (EC50 values in nM) of mutant mMrgC11 receptors from intracellular calcium assays

	Binding[a]	Wild Type	Y110A[b]	D161A[b]	D179A[b]	Y110F[c]	Y110W[c]
FdMRFa	100%	113 ± 18	>33000	>33000	>33000	714 (6.3)	334 (3.0)
FMRFa	93%	168 ± 26	>33000	+ (18000)	>33000	1795 (10.7)	1531 (9.1)
dFMRFa	88%	276 ± 56	>33000	>33000	>33000	1500 (5.4)	1513 (5.5)
Bam15		292 ± 19	>33000	>33000	+ (3000)	749 (2.6)	1713 (5.9)
. γ1-MSH		398 ± 189	>33000	>33000	+ (3000)	331 (0.8)	302 (0.8)
. γ2-MSH		340 ± 66	>33000	+ (33000)	+ (2000)	340 (1.0)	349 (1.0)
FMdRFa	67%	>10000					
FMRdFa	68%	>10000					

[a] Calculated binding energy relative to FdMRFa (absolute value = 117kcal/mol). [b] + means that activation starts at a given concentration. [c] Numbers in parentheses are the ratio with respect to the EC50 values of WT.

For mutants D161A we find that 4 of the 6 ligands no longer activate while that other two only activate for 100 times the concentration.

These results that mutation of Tyr110, Asp161 and Asp179 very strongly reduce or eliminate the activity of mMrgC11 receptor validate the predictions that these residues are involved in the ligand binding.

For a positive control experiment, the mutant of Asp81 in TM2 to Ala was transiently expressed in HEK293 cells along with the Y110A, D161A, and D179A mutant receptors also transiently expressed under the same condition. Then the intracellular calcium assay experiment was carried out with 0.33 μM of FMRFa. Except for the D81A mutant, the other three showed no activity.

We investigated the implication of the hydroxyl group on the Tyr110 in ligand recognition by replacing this tyrosine with phenylalanine or tryptophan. The potencies of γ1-MSH and γ2-MSH ligands are not affected by the absence of the hydroxyl group, indicating that the hydroxyl group does not contribute to ligand activation for these ligands. For the three tetrapeptide agonists

the Y110F or Y110W mutations leads to a factor of 5 to 10 reduction in the potency. This is consistent with our predicted structure (Figures 2.7 and 2.8) which does not have the hydroxyl group of Tyr110 interacting with the ligand, but instead forms a hydrogen bond with the carbonyl group of the backbone. The missing hydroxyl group should results in a dangling hydrogen bond donor which might induce an overall conformational change in the binding pocket to explain the loss in activity. Since mutation of Tyr110 to Ala totally extinguishes the activity for all six ligands, we conclude that the aromatic ring must be significant for all six cases.

To investigate whether the two non-agonist tetrapeptides, FMdRFa and FMRdFa are antagonists or weak binders (or non-binders), we saturated the receptors either with FMdRFa or with FMRdFa in three concentrations, 3.3, 16 and 33 μM and then measured the EC50 value for FdMRFa. The intensity of calcium signal remained on the same level as in the absence of FMdRFa or FMRdFa and the EC50 values did not change much (within standard deviation). This result shows that FMdRFa and FMRdFa do not block the efficacy of FdMRFa and at best bind only weakly to the receptor.

Summarizing, the experimental results show that Tyr110 (TM3), Asp161 (TM4) and Asp179 (TM5) are possibly in the binding site in agreement with the predictions. These predicted mutations focused on the dipeptide binding region. Using the binding region for the tetrapeptide, we now suggest that mutations of Trp162 (TM4), Phe190 (TM5), and Tyr237 (TM6) to Ala would also dramatically decrease binding. Additional validations could be to mutate either the receptor or the peptide ligand and to carry out other cell assay experiments such as radiolabelled ligand binding assays. Such studies should further improve our understanding of the structure and ligand binding site.

2.4.4 Prediction of the structure of the mMrgA1 receptor and the binding site for ligands

The 3D structure of mMrgA1 was predicted using MembStruk procedure described in this chapter. The CRMSD of Cα atoms between mMrgC11 and mMrgA1 is 2.49Å in the TM

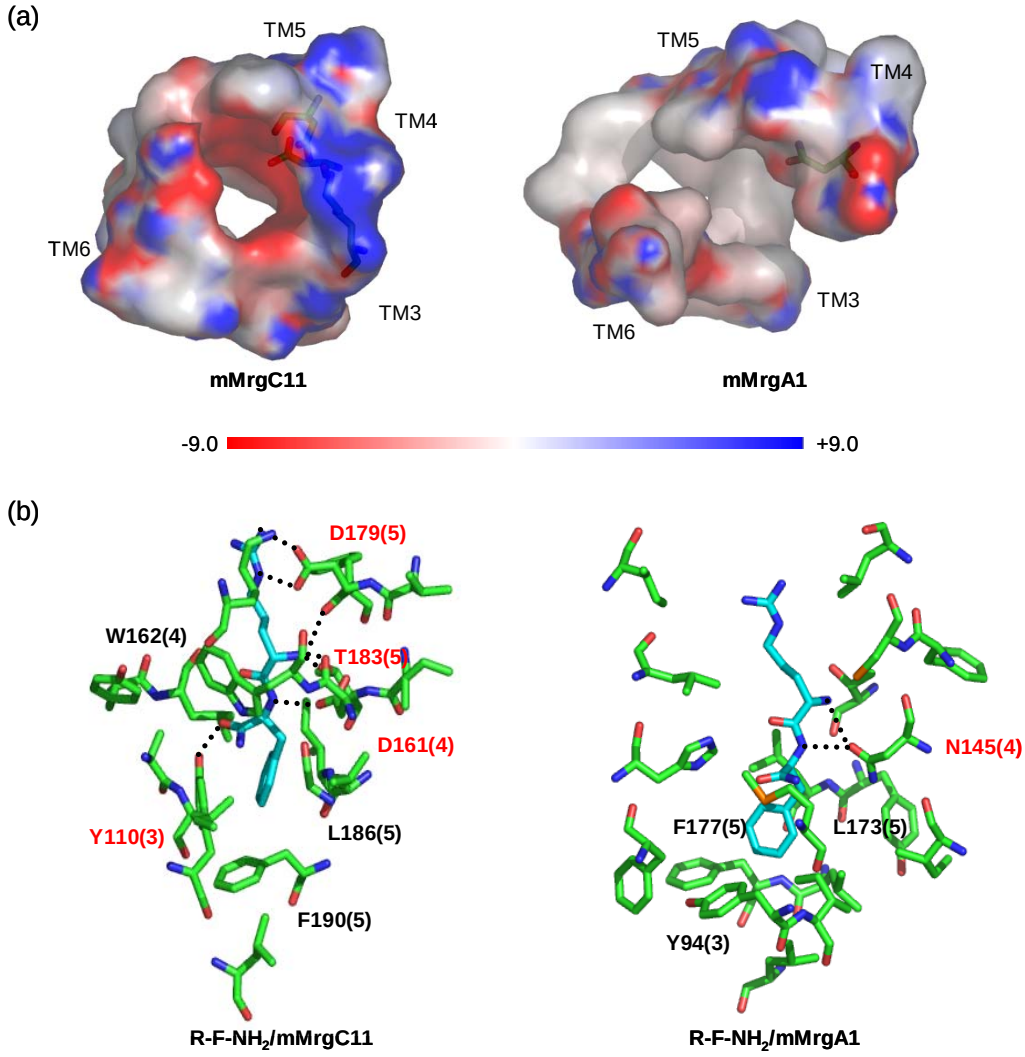

Figure 2.11 Comparison between mMrgC11 and mMrgA binding sites. (a) The electrostatic potential map of the binding pocket in mMrgC11 and mMrgA1. The residues within 5 Å from RFa ligand were selected for visualization. Asp161, Asp179 and Lys99 of mMrgC11 are specified in stick and Asn145 of mMrgA1 in stick. The electrostatic potential was computed using APBS and visualized on PyMOL. The van der Waals radii of DREIDING forcefield were used for APBS calculation. (b) The predicted 5 Å binding pocket of RFa in mMrgC11 and mMrgA1 receptor.

regions and 4.94 Å if the loops are included. The sequence identity between them is 53% for the TM regions and 46% for the entire sequence. It was observed experimentally that mMrgA1 receptor is activated much less potently by tetra-peptide ligands containing the RF-amide motif as compared to the mMrgC11 receptor, and that neither the amide nor acidic form of RF di-peptide activates mMrgA1[20].

We docked the RFa ligand into the mMrgA1 receptor by superimposing it with RFa-bound mMrgC11 receptor. The side chains of residues within 5 Å were reassigned using SCREAM and then the potential energy of the ligand-receptor complex structure was minimized. We found that Tyr94 (TM3), F177 (TM5) and L173 (TM5) (homologous residues of Tyr110, Phe190 and Leu186 in mMrgC11) form a hydrophobic pocket for Phe as in mMrgC11, but they are located slightly farther (the closet C–C distance between aromatic rings is 4 to 5 Å). Asn145 (TM4), the homologous residue of Asp161 in mMrgC11, is involved in the hydrogen bonding with the N-terminus. The Arg side chain of the peptide is surrounded with hydrophobic residues and does not have any favorable interaction with receptor. The calculated binding energy (positive value) predicts that it does not bind to mMrgA1 receptor.

Figure 2.11 shows the electrostatic potential maps of the binding pocket in mMrgC11 and of the corresponding region in mMrgA1. The electrostatic potential was calculated for the entire receptor using adaptive Poisson-Boltzmann solver (APBS)[28]. The binding pocket within 5 Å from RFa docked in mMrgC11 receptor is selectively presented here. We can see that the pocket of mMrgA1 is more hydrophobic than that of mMrgC11. In mMrgC11, two aspartic acids (Asp161 and Asp179) are located in the spot showing the fairly negative potential. We observed that in mMrgC11 the positively charged side chain of Arg and the N-terminus are favored in this region. In the absence of the ligand, Lys in TM3 (Lys99) compensates for this highly negative potential. For mMrgA1 these Asp residues are replaced by Asn. We expect that highly polar ligands such as a peptide containing an Arg residue might be unfavorable for the hydrophobic

character of the pocket in mMrgA1. This might explain why this ligand fails to bind strongly to the mMrgA1 receptor, explaining the low potency for RFa ligand to mMrgA1. This provides additional confirmation of our predicted binding site and protein structure for mMrgC11. We expect that the potency for these ligands to mMrgA1 might increase if these Asn residues are mutated to Asp, an experiment we intend to do soon.

2.4.5 Comparison of Mrg sequences

The 39 verified Mrg sequences were aligned using Clustal-W (v. 1.83) with the default parameters (protein gap open penalty = 10.0, protein gap extension penalty = 0.2, protein matrix = Gonnet) as shown in Figure S2.4. It includes 19 mouse, 13 rat, 1 monkey and 6 human Mrg receptors. The sequence identities range from 21% to 97 %. The mouse MrgF and rat MrgF have the highest sequence identity. The human MrgF also shows the relatively high sequence identity with rat and mouse orthologs (85% and 86% respectively). Across the 39 sequences we examined the sequence variations in the six key residues (Tyr110, Phe190, Asp161, Asp179, Trp162 and Tyr256) that we identified in this study. As mentioned before, Tyr110 is conserved throughout the Mrg sequences except for the 5 Mrgs that have the homologous Phe at the same position. Other five residues show various range of alteration;

- D161: 14 D or E, 7 N, 5 L, 2 A, 3 H, 2 P, 2 T, 1 Q, 1 K, 1 V and 1 Y.
- W162: 6 W, 14 G, 8 S, 4 N, 3 R, 2 A, 1 M and 1 E.
- D179: 14 D, 10 N, 5 I, 3 M, 2 A, 2 H, 2 W and 1 S.
- F190: 20 F, 8 C, 3 T, 2 M, 2 S, 1 A, 1 I, 1 L and 1 V.
- Y256: 20 Y or F, 3 C, 3 D, 3 Q, 2 H, 2 L, 2 N, 2 I, 1 S, and 1 T.

We observed that only rat MrgC has all six residues conserved and mouse MrgB1, mouse MrgB2 and rat MrgB2 have Tyr110, Asp161, Asp179, Phe190 and Y256 at their homologous positions. Trp162 are replaced with Gly for these MrgB receptors. It has been shown that γ2-MSH is the

most potent at activating rat MrgC and the active moiety recognized by rat MrgC is the C-terminal of γ2-MSH, F-R-W-D-R-F-G[29]. The rat MrgC also binds Met-Enkephalin RF-amide with F-M-R-F-NH_2 at the C-terminus. These experimental results suggest the similar characteristics in the binding site of rat MrgC receptor to that of mMrgC11, further supporting our predictions.

2.5 Summary and conclusions

We predicted the 3D structure of the mMrgC11 receptor and used it to predict the binding sites for a number of di- and tetra-peptide ligands. We find that in each case the peptide ligand binds in a pocket among TM3, 4, 5 and 6 oriented parallel to the helical axis. These predictions suggested that three residues (Tyr110 (TM3), Asp161 (TM4) and Asp179 (TM5) in the binding pocket) play a key role in the binding.

To test these predictions, we carried out several mutagenesis experiments. For 6 ligands exhibiting EC50 of 100 to 400 nM in wild type, we find that the EC50 for the Y110A, D161A and D179A mutant receptors are higher than 33 μM for 14 of 18 combinations and 50 to 100 times higher for the other 4 combinations. This validates the implication of these residues for the activation or binding of the ligand.

Since the peptide forms a zwitterion at pH 7 giving it relatively polar character and since the ligands that bind to MrgC11 contain an Arg whose side chain is positively charged at pH 7, it is plausible that the two aspartic acids in the binding pocket participate. On the other hand, mMrgA1 has increased hydrophobic character in the corresponding region (these Asp are replaced by Asn). This might be responsible for the low efficacy of the ligand.

Our predicted binding site also suggests additional mutation candidates to be tested, especially residues involving hydrophobic interaction such as Trp162, Leu186, Phe190, Tyr237, and Leu238.

This study indicates how collaboration between theory and experiment can provide insight into the structural characterization of these Mrg receptors to determine how they are related with function. This could lead to the design of small molecule antagonists to selectively inhibit these receptors as candidate drugs for treating pain. Such studies would be equally valuable for many other GPCR receptors, indicating that a systematic combination of computational tools along with biochemical experiments can provide an increased understanding membrane protein receptors and their activation.

References

1. Dong, X.Z., et al., *A diverse family of GPCRs expressed in specific subsets of nociceptive sensory neurons.* Cell, 2001. **106**(5): p. 619-632.
2. Lembo, P.M.C., et al., *Proenkephalin A gene products activate a new family of sensory neuron-specific GPCRs.* Nat. Neurosci., 2002. **5**(3): p. 201-209.
3. Civelli, O., *GPCR deorphanizations: the novel, the known and the unexpected transmitters.* Trends Pharmacol. Sci., 2005. **26**(1): p. 15-19.
4. Vaidehi, N., et al., *Prediction of structure and function of G protein-coupled receptors.* Proc. Natl. Acad. Sci. U. S. A., 2002. **99**(20): p. 12622-12627.
5. Trabanino, R.J., et al., *First principles predictions of the structure and function of G-protein-coupled receptors: Validation for bovine rhodopsin.* Biophys. J., 2004. **86**(4): p. 1904-1921.
6. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - a Generic Force-Field for Molecular Simulations.* J. Phys. Chem., 1990. **94**(26): p. 8897-8909.
7. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* J. Phys. Chem. B, 1998. **102**(18): p. 3586-3616.
8. Lim, K.T., et al., *Molecular dynamics for very large systems on massively parallel computers: The MPSim program.* J. Comput. Chem., 1997. **18**(4): p. 501-521.
9. Ding, H.Q., N. Karasawa, and W.A. Goddard, *Atomic Level Simulations on a Million Particles - the Cell Multipole Method for Coulomb and London Nonbond Interactions.* J. Chem. Phys., 1992. **97**(6): p. 4309-4315.
10. Hall, S.E., *Development of a structure prediction method for G-protein coupled receptors*, in *Division of Chemistry and Chemical Engineering*. 2005, California Institute of Technology: Pasadena.

11. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res., 1997. **25**(17): p. 3389-3402.

12. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice.* Nucleic Acids Res., 1994. **22**(22): p. 4673-4680.

13. Unger, V.M., et al., *Arrangement of rhodopsin transmembrane alpha-helices.* Nature, 1997. **389**(6647): p. 203-206.

14. Canutescu, A.A., A.A. Shelenkov, and R.L. Dunbrack, *A graph-theory algorithm for rapid protein side-chain prediction.* Protein Sci., 2003. **12**(9): p. 2001-2014.

15. *MODELLER6v2*, University of California San Francisco: San Francisco.

16. Cho, A.E., et al., *The MPSim-Dock hierarchical docking algorithm: Application to the eight trypsin inhibitor cocrystals.* J. Comput. Chem., 2005. **26**(1): p. 48-71.

17. Ewing, T.J.A. and I.D. Kuntz, *Critical evaluation of search algorithms for automated molecular docking and database screening.* J. Comput. Chem., 1997. **18**(9): p. 1175-1189.

18. *Cerius2 Modeling Environment, Release 4.0*, Accelrys Inc.: San Diego.

19. Zamanakos, G., *A fast and accurate analytical method for the computation of solvent effects in molecular simulations*, in *Division of Physics, Mathematics and Astronomy*. 2002, California Institute of Technology: Pasadena.

20. Han, S.K., et al., *Orphan G protein-coupled receptors MrgA1 and MrgC11 are distinctively activated by RF-amide-related peptides through the G alpha(q/11) pathway.* Proc. Natl. Acad. Sci. U. S. A., 2002. **99**(23): p. 14740-14745.

21. Ghosh, A., C.S. Rapp, and R.A. Friesner, *Generalized born model based on a surface integral formulation.* J. Phys. Chem. B, 1998. **102**(52): p. 10983-10990.

22. Okada, T., et al., *The retinal conformation and its environment in rhodopsin in light of a new 2.2 angstrom crystal structure.* J. Mol. Biol., 2004. **342**(2): p. 571-583.

23. McDonald, I.K. and J.M. Thornton, *Satisfying Hydrogen-Bonding Potential in Proteins.* J. Mol. Biol., 1994. **238**(5): p. 777-793.

24. Miura, S. and S.S. Karnik, *Constitutive activation of angiotensin II type 1 receptor alters the orientation of transmembrane helix-2.* J. Biol. Chem., 2002. **277**(27): p. 24299-24305.

25. Mills, J.S., et al., *Characterization of the binding site on the formyl peptide receptor using three receptor mutants and analogs of Met-Leu-Phe and Met-Met-Trp-Leu-Leu.* J. Biol. Chem., 2000. **275**(50): p. 39012-39017.

26. Pal, D. and P. Chakrabarti, *Non-hydrogen bond interactions involving the methionine sulfur atom.* J. Biomol. Struct. Dyn., 2001. **19**(1): p. 115-128.

27. Heringdorf, D.M.Z., et al., *Discrimination between plasma membrane and intracellular target sites of sphingosylphosphorylcholine.* Eur. J. Pharmacol., 1998. **354**(1): p. 113-122.

28. Baker, N.A., et al., *Electrostatics of nanosystems: Application to microtubules and the ribosome.* Proc. Natl. Acad. Sci. U. S. A., 2001. **98**(18): p. 10037-10041.

29. Grazzini, E., et al., *Sensory central neuron-specific receptor activation elicits and peripheral nociceptive effects in rats.* Proc. Natl. Acad. Sci. U. S. A., 2004. **101**(18): p. 7175-7180.

Supporting figures and tables

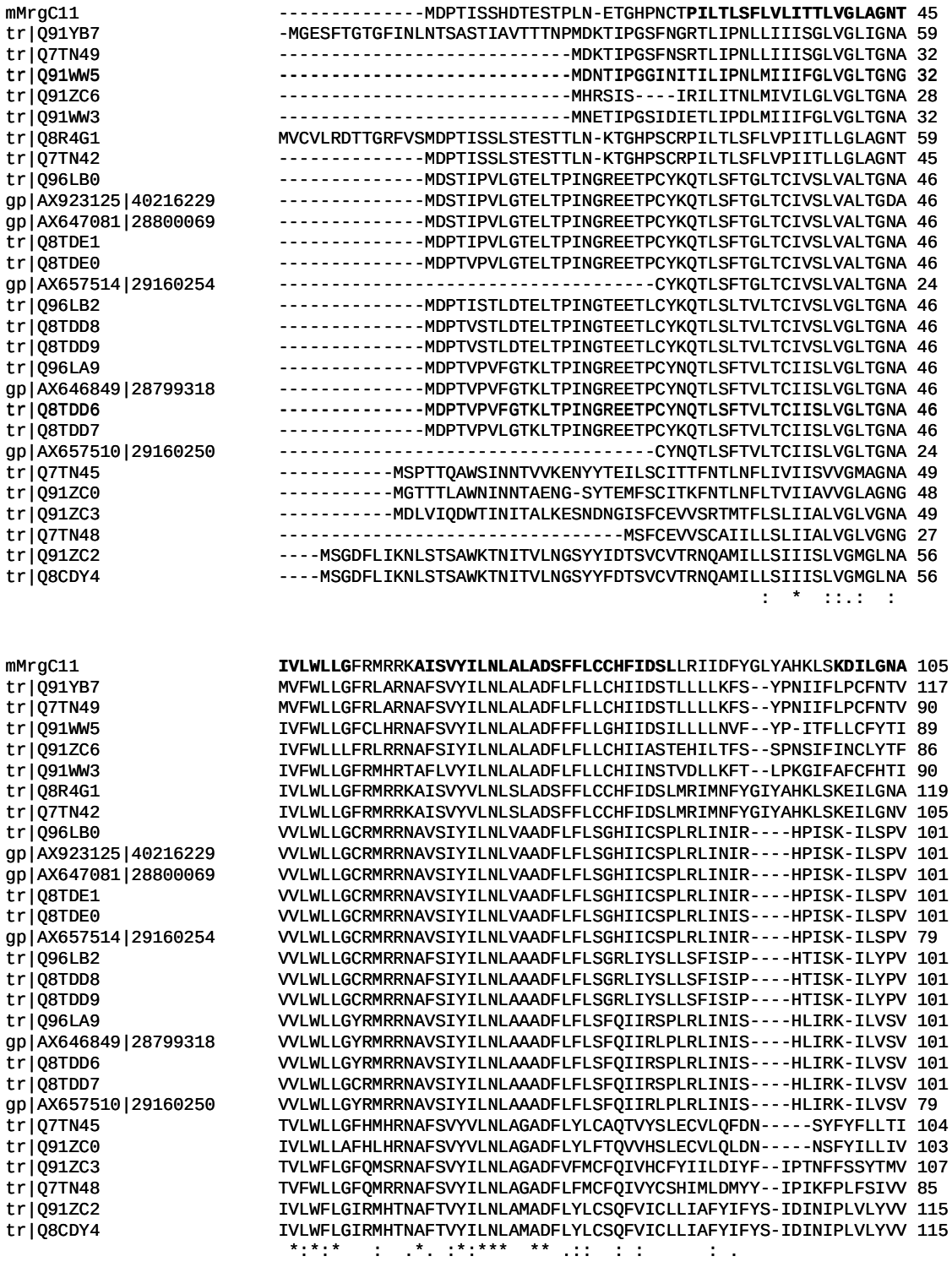

Figure S2.1 Multiple sequence alignment for mMrgC11 with 27 homologous sequences.

```
mMrgC11                 AIIPYISGLSILSAISTERCLCVLWPIWYHCHRPRNMSAIICALIWVLSFLMGILDWF-S 164
tr|Q91YB7               MMVPYIAGLSMLSAISTERCLSVVCPIWYRCRRPKHTSTVMCSAIWVLSLLICILNRYFC 177
tr|Q7TN49               MMVPYIAGLSMLSAISTERCLSVVCPIWYRCRRPKHTSTVMCSAIWVLSLLICILNRYFC 150
tr|Q91WW5               MMVLYIAGLSMLSAISTERCLSVLCPIWYHCHRPEHTSTVMCAVIWVLSLLICILNSYFC 149
tr|Q91ZC6               RVLLYIAGLSMLSAISIERCLSVMCPIWYRCHSPEHTSTVMCAMIWVLSLLLCILYRYFC 146
tr|Q91WW3               KRVLYITGLSMLSAISTERCLSVLCPIWYHCRRPEHTSTVMCAVIWVLSLLICILDGYFC 150
tr|Q8R4G1               AIIPYISGLSILSAISTERCLSVLWPIWYHCHRPRNMSAIICVLIWVLSFLMGILDWFFS 179
tr|Q7TN42               AFIPYISGLSILSAISTERCLSVLWPIWYHCHRPRNMSAIICVLIWVLSFLMGILDWFFS 165
tr|Q96LB0               MTFPYFIGLSMLSAISTERCLSILWPIWYHCRRPRYLSSVMCVLLWALSLLRSILEWMFC 161
gp|AX923125|40216229    MTFPYFIGLSMLSAISTERCLSILWPIWYHCRRPRYLSSVMCVLLWALSLLRSILEWMFC 161
gp|AX647081|28800069    MTFPYFIGLSMLSAISTERCLSILWPIWYHCRRPRYLSSVMCVLLWALSLLRSILEWMFC 161
tr|Q8TDE1               MTFPYFIGLSMLSAISTERCLSILWPIWYHCRRPRYLSSVMCVLLWALSLLRSILEWMFC 161
tr|Q8TDE0               MTFPYFIGLSMLNAISTERCLSILWPIWYHCRRPRYLSSVMCVLLWAPSLLRSILEWMFC 161
gp|AX657514|29160254    MTFPYFIGLSMLSAISTERCLSILWPIWYHCRRPRYLSSVMCVLLWALSLLRSILEWMFC 139
tr|Q96LB2               MMFSYFAGLSFLSAVSTERCLSVLWPIWYRCHRPTHLSAVVCVLLWALSLLRSILEWMLC 161
tr|Q8TDD8               MMFSYFAGLSFLSAVSTERCLSVLWPIWYRCHRPTHLSAVVCVLLWALSLLRSILEWMLC 161
tr|Q8TDD9               MMFSYFAGLSFLSAVSTERCLSVLWPIWYRCHRPTHLSAVVCVLLWALSLLRSILEWMLC 161
tr|Q96LA9               MTFPYFTGLSMLSAISTERCLSVLWPIWYRCRRPTHLSAVVCVLLWGLSLLFSMLEWRFC 161
gp|AX646849|28799318    MTFPYFTGLSMLSAISTERCLSVLWPIWYRCRRPTHLSAVVCVLLWGLSLLFSMLEWRFC 161
tr|Q8TDD6               MTFPYFTGLSMLSAISTERCLSVLWPIWYRCRRPTHLSAVVCVLLWGLSLLFSMLEWRFC 161
tr|Q8TDD7               MTFPYFTGLSMLSAISTERCLSVLWPIWYRCRRPTHLSAVVCVLLWGLSLLFSMLEWRFC 161
gp|AX657510|29160250    MTFPYFTGLSMLSAISTERCLSVLWPIWYRCRRPTHLSAVVCVLLWGLSLLFSMLEWRFC 139
tr|Q7TN45               LMFNYLAGFCMIAAISTERCLSVTWPIWYHCQRPRHTSATVCALFWAFSLLLSLLLGQGC 164
tr|Q91ZC0               TMFAYLAGLCMIAAISAERCLSVMWPIWYHCQRPRHTSAIMCALVWVSSLLLSLVVGLGC 163
tr|Q91ZC3               LNIAYLSGLSILTVISTERFLSVMWPIWYRCQRPRHTSAVICTVLWVLSLVLSLLEGKEC 167
tr|Q7TN48               LNIGYLCGMSILSAISIERCLSVMWPIWYRCQRPRHTSAVICTLLWVLALVWSLIEGKEC 145
tr|Q91ZC2               PIFAYLSGLSILSTISIERCLSVIWPIWYRCKRPRHTSAITCFVLWVMSLLLGLLEGKAC 175
tr|Q8CDY4               PIFAYLSGLSILSTISIERCLSVIWPIWYRCKRPRHTSAITCFVLWVMSLLLGLLEGKAC 175
                          . *: *:.:: .:* ** *.:  ****:*: *   *:  *  .*  :::  ::    .

mMrgC11                 GFLGETHHH-LWKN-VDFIITAFLIFLFMLLSGSSLALLLRILCGPRRKPLSRLYVTIAL 222
tr|Q91YB7               GFLDTKYEKDNRCLASNFFTAACLIFLFVVLCLSSLALLVRSFCGAGRMKLTRLYATIML 237
tr|Q7TN49               GFLDTKYEKDNRCLASNFFTAACLIFLFVVLCLSSLALLVRLFCGAGRMKLTRLYATIML 210
tr|Q91WW5               GFLNTQYKNENGCLALNFFTAAYLMFLFVVLCLSSLALVARLFCGTGQIKLTRLYVTIIL 209
tr|Q91ZC6               GFLDTKYEDDYGCLAMNFLTTAYLMFLFVVLCVSSLALLARLFCGAGRMKLTRLYVTITL 206
tr|Q91WW3               GYLDNHYFNYSVCQAWDIFIGAYLMFLFVVLCLSTLALLARLFCGARNMKFTRLFVTIML 210
tr|Q8R4G1               GFLGETHHH-LWKN-VDFIVTAFLIFLFMLLFGSSLALLVRILCGSRRKPLSRLYVTISL 237
tr|Q7TN42               GFLGETHHH-LWKN-VDFIVTAFLIFLFMLLFGSSLALLVRILCGSRRKPLSRLYVTISL 223
tr|Q96LB0               DFLFSGADS-VWCETSDFITIAWLVFLCVVLCGSSLVLLVRILCGSRKMPLTRLYVTILL 220
gp|AX923125|40216229    DFLFSGADS-VWCETSDFITIAWLVFLCVVLCGSSLVLLVRILCGSRKMPLTRLYVTILL 220
gp|AX647081|28800069    DFLFSGANS-VWCETSDFITIAWLVFLCVVLCGSSLVLLVRILCGSRKMPLTRLYVTILL 220
tr|Q8TDE1               DFLFSGANS-VWCETSDFITIAWLVFLCVVLCGSSLVLLVRILCGSRKMPLTRLYVTILL 220
tr|Q8TDE0               DFLFSGADS-VRCETSDFITIAWLVFLRVVLCGSSLVLLVRILCGSRKMPLTRLYVTILL 220
gp|AX657514|29160254    DFLFSGANS-VWCETSDFITIAWLVFLCVVLCGSSLVLLVRILCGSRKMPLTRLYVTILL 198
tr|Q96LB2               GFLFSGADS-AWCQTSDFITVAWLIFLCVVLCGSSLVLLIRILCGSRKIPLTRLYVTILL 220
tr|Q8TDD8               GFLFSGADS-AWCQTSDFITVAWLIFLCVVLCGSSLVLLIRILCGSRKIPLTRLYVTILL 220
tr|Q8TDD9               GFLFSGADS-AWCQTSDFITVAWLIFLCVVLCGSSLVLLIRILCGSRKIPLTRLYVTILL 220
tr|Q96LA9               DFLFSGADS-SWCETSDFIPVAWLIFLCVVLCVSSLVLLVRILCGSRKMPLTRLYVTILL 220
gp|AX646849|28799318    DFLFSGADS-SWCETSDFIPVAWLIFLCVVLCVSSLVLLVRILCGSRKMPLTRLYVTILL 220
tr|Q8TDD6               DFLFSGADS-SWCETSDFIPVVWLIFLCVVLCVSSLVLLVRILCGSRKMPLTRLYVTILL 220
tr|Q8TDD7               DFLFSGADS-SWCETSDFIPVAWLIFLCVVLCVSSLVLLVRILCGSRKMPLTRLYVTILL 220
gp|AX657510|29160250    DFLFSGADS-SWCETSDFIPVAWLIFLCVVLCVSSLVLLVRILCGSRKMPLTRLYVTILL 198
tr|Q7TN45               GFLFSKFDY-SFCRYCNFIATAFLIVIFMVLFVSSLALLAKIICGSHRIPVTRFYVTIAL 223
tr|Q91ZC0               GFLFSYYDY-YFCITLNFITAAFLIVLSVVLSVSSLALLVKIVWGSHRIPVTRFFVTIAL 222
tr|Q91ZC3               GFLYYTSGP-GLCKTFDLITTAWLIVLFVVLLGSSLALVLTIFCGLHKVPVTRLYVTIVF 226
tr|Q7TN48               GFLFDTNGP-GWCETFDLIATAWLIVLIVVLLGSSLALVINIFCGLYRIPVTRLYVTIVF 204
tr|Q91ZC2               GLLFNSFDS-YWCETFDVITNIWSVVFFGVLCGSSLTLLVRIFCGSQRIPMTRLYVTITL 234
tr|Q8CDY4               GLLFNSFDS-YWCETFDVITNIWSVVFFGVLCGSSLTLLVRIFCGSQRIPMTRLYVTITL 234
                        . *            :.:     :.:  :*  *:*.*:   . *  .  .:*::.** :
```

Figure S2.1 (*continued*)

```
mMrgC11                 TVMVYLICGLPLGLYLFLLYWFGVHLHYPFCHIYQVTAVLSCVNSSANPIIYFLVGSFRQ 282
tr|Q91YB7               TVLVFLLCGLPFGIHWFLLIWIKIDYGKFAYGLYLAALVLTAVNSCANPIIYFFVGSFRH 297
tr|Q7TN49               TVLVFLLCGLPFGIHWFLLIWIKIDYGKFAYGLYLAALVLTAVNSCANPIIYFFVGSFRH 270
tr|Q91WW5               SILVFLLCGLPFGIHWFLLFKIKDDFHVFDLGFYLASVVLTAINSCANPIIYFFVGSFRH 269
tr|Q91ZC6               TLLVFLLCGLPCGFYWFLLSKIKNVFTVFEFSLYLASVVLTAINSCANPIIYFFVGSFRH 266
tr|Q91WW3               TVLVFLLCGLPWGITWFLLFWIAPGVFVLDYS---PLLVLTAINSCANPIIYFFVGSFRQ 267
tr|Q8R4G1               TVMVYLICGLPLGLYLFLLYWFGIHLHYPFCHIYQVTVLLSCVNSSANPIIYFLVGSFRH 297
tr|Q7TN42               TVMVYLICGLPLGLYLFLLYWFGIHLHYPFCHIYQVTVLLSCVNSSANPIIYFLVGSFRH 283
tr|Q96LB0               TVLVFLLCGLPFGIQWALFSRIHLDWKVLFCHVHLVSIFLSALNSSANPIIYFFVGSFRQ 280
gp|AX923125|40216229    TVLVFLLCGLPFGIQWALFSRIHLDWKVLFCHVHLVSIFLSALNSSANPIIYFFVGSFRQ 280
gp|AX647081|28800069    TVLVFLLCGLPFGIQWALFSRIHLDWKVLFCHVHLVSIFLSALNSSANPIIYFFVGSFRQ 280
tr|Q8TDE1               TVLVFLLCGLPFGIQWALFSRIHLDWKVLFCHVHLVSIFLSALNSSANPIIYFFVGSFRQ 280
tr|Q8TDE0               TVLVFLLCGLPFGIQWALFSRIHLDWKVLFCHVHLVSIFLSALNSSANPIIYFFMGSFRQ 280
gp|AX657514|29160254    TVLVFLLCGLPFGIQWALFSRIHLDWKVLFCHVHLVSIFLSALNSSANPIIYFFVGSFRQ 258
tr|Q96LB2               TVLVFLLCGLPFGIQFFLFLWIHVDREVLFCHVHLVSIFLSALNSSANPIIYFFVGSFRQ 280
tr|Q8TDD8               TVLVFLLCGLPFGIQFFLFLWIHVDREVLFCHVHLVSIFLSALNSSANPIIYFFVGSFRQ 280
tr|Q8TDD9               TVLVFLLCGLPFGIQFFLFLWIHVDREVLFCHVHLVSIFLSALNSSANPIIYFFVGSLRQ 280
tr|Q96LA9               TVLVFLLCGLPFGILGALIYRMHLNLEVLYCHVYLVCMSLSSLNSSANPIIYFFVGSFRQ 280
gp|AX646849|28799318    TVLVFLLCGLPFGILGALIYRMHLNLEVLYCHVYLVCMSLSSLNSSANPIIYFFVGSFRQ 280
tr|Q8TDD6               TVLVFLLCGLPFGILGALIYRMHLNLEVLYCHVYLVCMSLSSLNSSANPIIYFFVGSFRQ 280
tr|Q8TDD7               TVLVFLLCGLPFGILGALIYRMHLNLEVLYCHVYLVCMSLSSLNSSANPIIYFFVGSFRQ 280
gp|AX657510|29160250    TVLVFLLCGLPFGILGALIYRMHLNLEVLYCHVYLVCMSLSSLNSSANPIIYFFVGSFRQ 258
tr|Q7TN45               TVLVFIFFGLPIGICVFLLPWIHMMLSSFF---YEMVTLLSCVNSCANPIIYFFVGSIRH 280
tr|Q91ZC0               TVVVFIYFGMPFGICWFLLSRIMEFDSIFFNNVYEIIEFLSCVNSCANPIIYFLVGSIRQ 282
tr|Q91ZC3               TVLVFLIFGLPYGIYWFLLEWIREFHDNKPCGFRNVTIFLSCINSCANPIIYFLVGSIRH 286
tr|Q7TN48               TVLVFLLCGLPYGIYWFLLEWTEKFNYNLPCGFHPVTVLLSCVNSCANPIIYFLVGSIRH 264
tr|Q91ZC2               TVLVFLIFGLPFGIYWILYQWISNFYYVEICNFYLEILFLSCVNSCMNPIIYFLVGSIRH 294
tr|Q8CDY4               TVLVFLIFGLPFGIYWILYQWISNFYYVEICNFYLEILFLSCVNSCMNPIIYFLVGSIRH 294
                        :::*::  *:* *:   *                   *:.:**. ******::**:*:


mMrgC11                 H-RKHRSLKR---VLKRALEDTPEEDEYTDSHL-HKTTEISESRY---------- 322
tr|Q91YB7               --QKHQTLKM---VLQRALQDTPETAEN--------TVEMSSSKVEP-------- 331
tr|Q7TN49               --QKHQTLKM---VLQRALQDTPETAEN--------TVEMSSSKVEP-------- 304
tr|Q91WW5               R-LKHQTLKM---VLQNALQDTPETAKI--------MVEMSRSKSEP-------- 304
tr|Q91ZC6               R-LKHQTLKM---VLQSALQDTPETPEN--------MVEMSRNKAEL-------- 301
tr|Q91WW3               R-LNKQTLKM---VLQKALQDTPETPEN--------MVEMSRNKAEP-------- 302
tr|Q8R4G1               R-KKHRSLKM---VLKRALEETPEEDEYTDSHV-QKPTEISERRC---------- 337
tr|Q7TN42               R-KKHRSLKM---VLKRALEETPEEDEYTDSHV-QKPTEISERRC---------- 323
tr|Q96LB0               R-QNRQNLKL---VLQRALQDTPEVDEGGGWLP-QETLELSGSRLEQ-------- 322
gp|AX923125|40216229    R-QNRQNLKL---VLQRALQDTPEVDEGGGWLP-QETLELSGSRLEQ-------- 322
gp|AX647081|28800069    R-QNRQNLKL---VLQRALQDTPEVDEGGGWLP-QETLELSGSRLEQ-------- 322
tr|Q8TDE1               R-QNRQNLKL---VLQRALQDTPEVDEGGGWLP-QETLELSGSKLEQ-------- 322
tr|Q8TDE0               L-QNRKTLKL---VLQRDLQDTPEVDEGGWWLP-QETLELSGSRLEQ-------- 322
gp|AX657514|29160254    R-QNRQNLKLDSMCRRTALYKTIRSRESYSLSREQQREDPTHDSILS-------- 304
tr|Q96LB2               R-QNRQNLKL---VLQRALQDASEVDEGGGQLP-EEILELSGSRLEQ-------- 322
tr|Q8TDD8               R-QNRQNLKL---VLQRALQDASEVDEGGGQLP-EEILELSGSRLEQ-------- 322
tr|Q8TDD9               R-QNRQNLKL---VLQRALQDTPEVDEGGGWLP-QETLELSGSRLEQ-------- 322
tr|Q96LA9               R-QNRQNLKL---VLQRALQDKPEVDKGEGQLP-EESLELSGSRLGP-------- 322
gp|AX646849|28799318    R-QNRQNLKL---VLQRALQDKPEVDKGEGQLP-EESLELSGSRLGP-------- 322
tr|Q8TDD6               R-QNRQNLKL---VLQRALQDKPEVDKGEGQLP-EESLELSGSKLGP-------- 322
tr|Q8TDD7               R-QNRQNLKL---VLQRALQDKPEVDKGEGQLP-EESLELSGRRLGP-------- 322
gp|AX657510|29160250    R-QNRQNLKL---VLQRALQDKPEVDKASATRS-RTRTTSTSSASTPPRPT---- 304
tr|Q7TN45               HRLQRQTLKL---LLQRAMQDTPEEE-GGERGPSQKSEDLEVVRCSS-------- 323
tr|Q91ZC0               HRLRWQSLKL---LLQRAMQDTPEEE-SGERGPSQRSGELETV------------ 321
tr|Q91ZC3               HRFQRKTLKL---LLQRAMQDSPEEEECGEMGSSRRPREIKTVWKGLRAALIRHK 338
tr|Q7TN48               HRFQRKTLKL---LLQKAMQDTPEEEECGEMGS---------------------- 294
tr|Q91ZC2               RRFRRKTLKL---LLQRAMQDTPEEEQSGNKSSSEHPEELETVQSCS-------- 338
tr|Q8CDY4               RRFRRKTLKL---LLQRAMQDTPEEEQSGNKSSSEHPEELETVQSCS-------- 338
                           . :.**      :  : .  .
```

Figure S2.1 (*continued*)

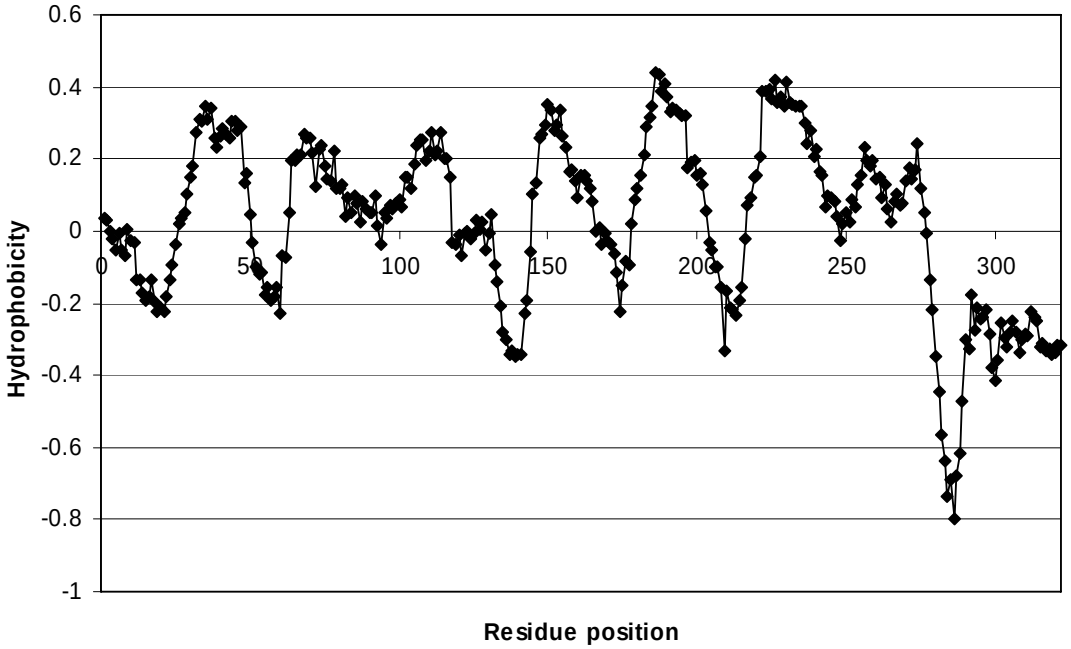

Figure S2.2 Hydrophobicity profile for mMrgC11 sequence set (window size = 12).

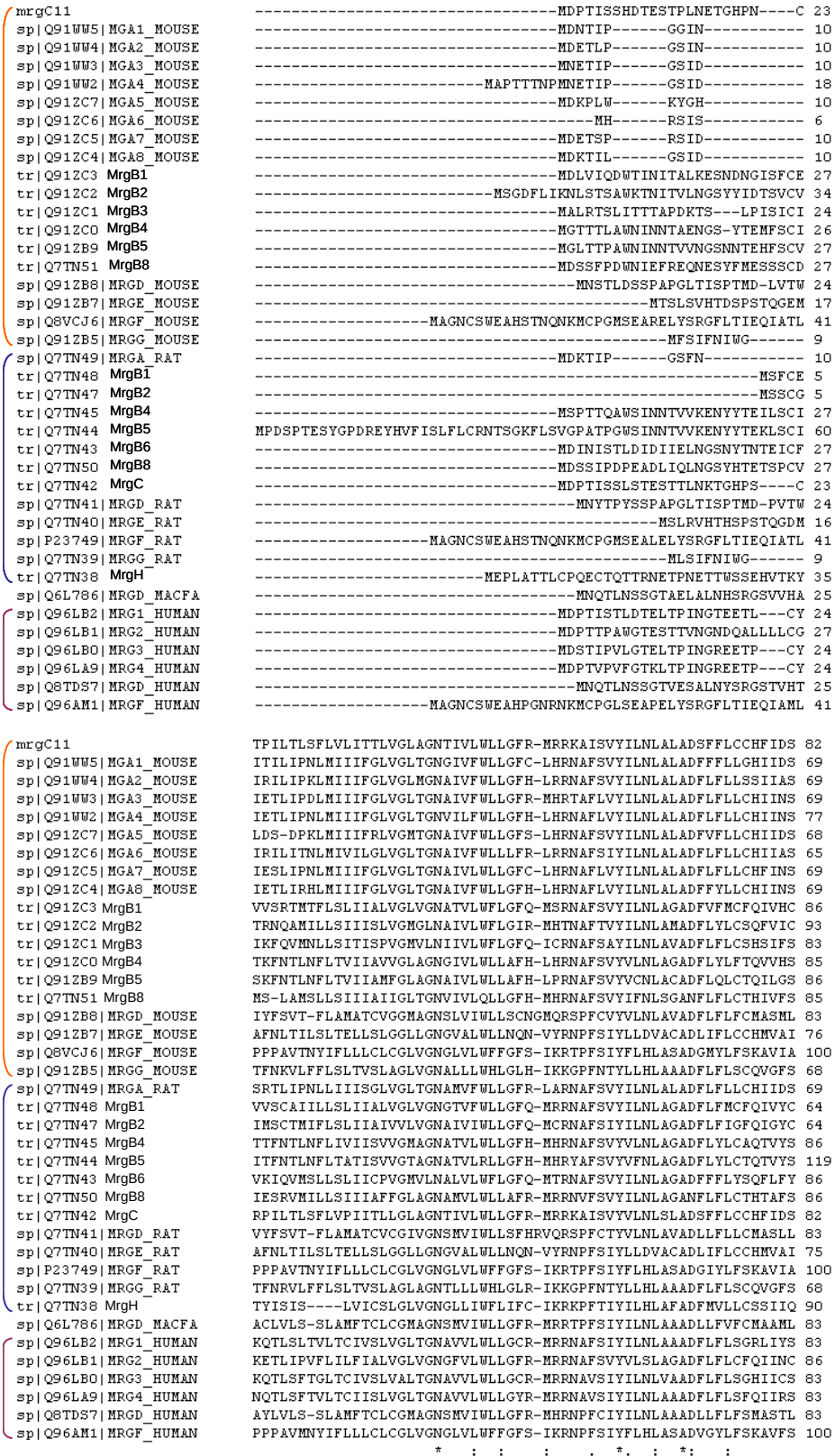

Figure S2.3 Multiple alignment of 39 verified Mrg sequences, including 19 mouse (orange), 13 rat (navy), 1 monkey and 6 human (violet) receptors. The positions of six key residues are specified in red boxes.

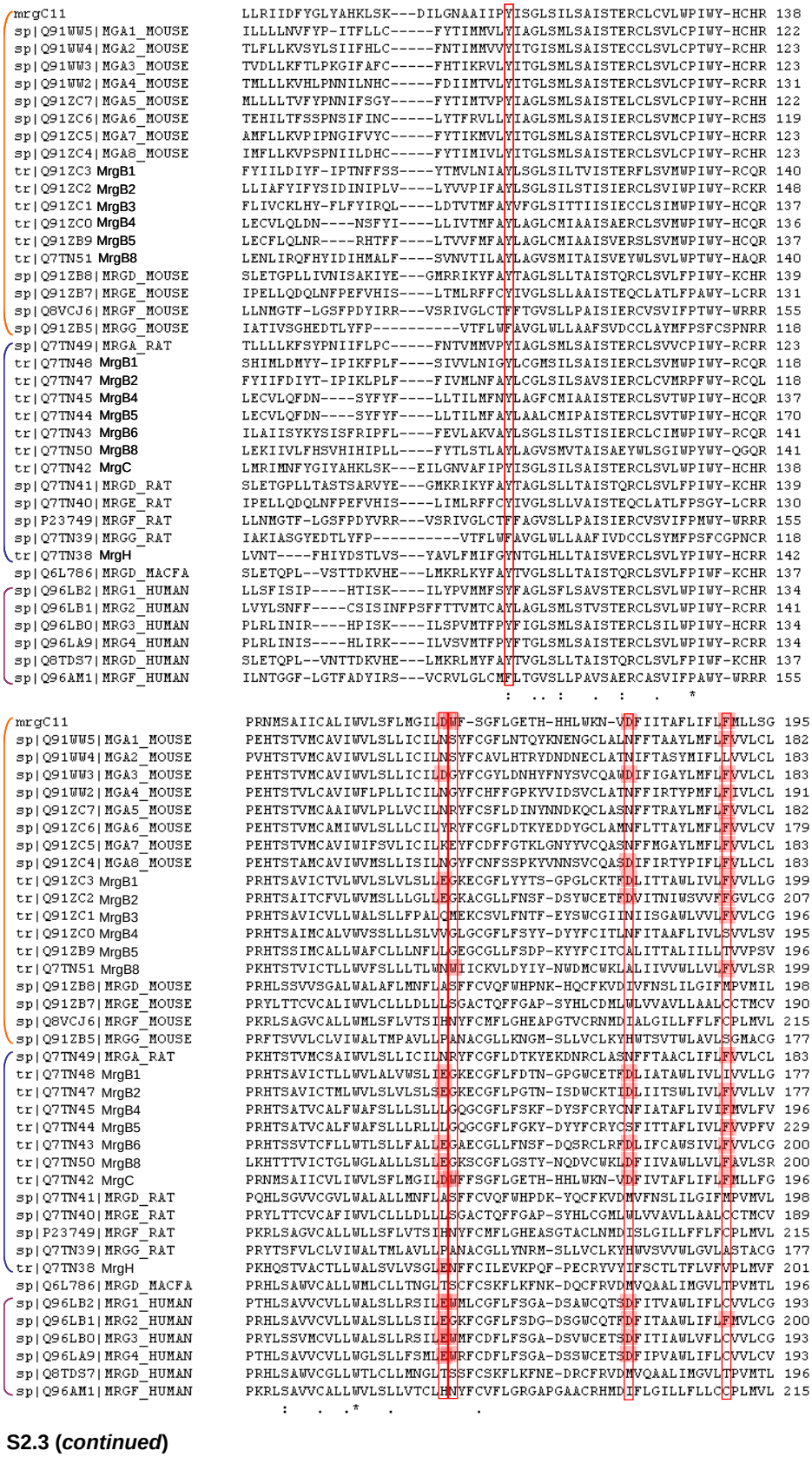

Figure S2.3 (*continued*)

Figure S2.3 (*continued*)

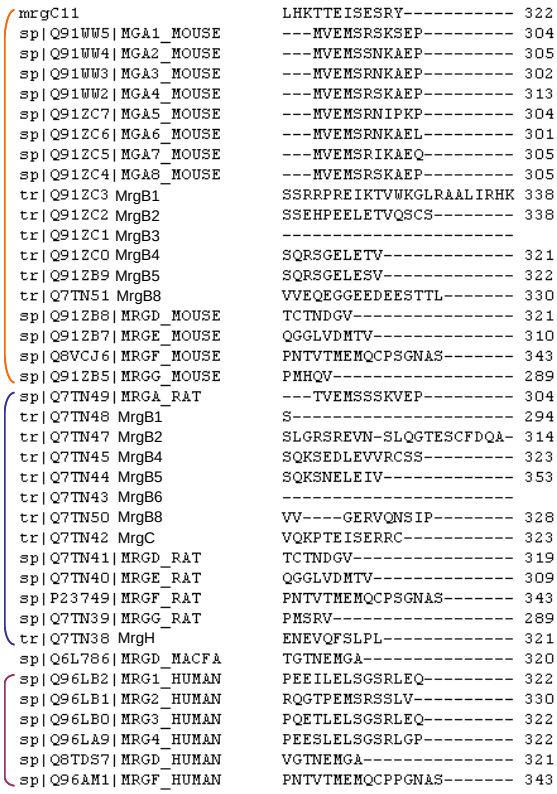

Figure S2.3 (*continued*)

Table S2.1 Hit sequences from independent BLAST search of each TM

Sequence	Identity (%)[a]	TM1	TM2	TM3	TM4	TM5	TM6	TM7
tr|Q91YB7	46	x	x	x	x		x	x
tr|Q7TN49	49	x	x	x	x		x	
tr|Q91WW5	46	x	x	x	x		x	x
tr|Q91ZC6	47	x	x		x	x	x	x
tr|Q91WW3	47	x	x	x			x	
tr|Q8R4G1	88	x	x	x	x	x	x	x
tr|Q7TN42	88	x	x	x	x	x	x	x
tr|Q96LB0	51	x	x	x			x	x
gp|AX923125|40216229	51			x	x		x	x
gp|AX647081|28800069	51	x	x	x	x		x	
tr|Q8TDE1	51	x	x	x	x		x	x
tr|Q8TDE0	49	x	x		x		x	x
gp|AX657514|29160254	46	x	x	x	x		x	x
tr|Q96LB2	53	x			x		x	x
tr|Q8TDD8	53	x			x		x	x
tr|Q8TDD9	53	x			x		x	x
tr|Q96LA9	51	x		x	x		x	x
gp|AX646849|28799318	50	x		x	x		x	x
tr|Q8TDD6	50	x		x	x		x	x
tr|Q8TDD7	50	x		x	x		x	x
gp|AX657510|29160250	50	x		x	x		x	x
tr|Q7TN45	43	x	x			x	x	x
tr|Q91ZC0	42	x						x
tr|Q91ZC3	44	x	x	x	x	x	x	x
tr|Q7TN48	48	x	x				x	x
tr|Q91ZC2	42		x		x		x	x
tr|Q8CDY4	42		x		x		x	x
tr|Q91ZB9	41	x						x
tr|Q96LB1	49	x	x		x	x	x	x

[a] w.r.t. the sequence of mMrgC11

Table S2.1 (*continued*)

Sequence	Identity (%)[a]	TM1	TM2	TM3	TM4	TM5	TM6	TM7
tr|Q91ZC5	45	x	x	x			x	x
tr|Q7TN39	23	x						
tr|Q91ZC4	42	x	x	x	x		x	x
tr|Q91WW4	42	x	x	x	x		x	
tr|AAH64040	40	x	x	x	x		x	
tr|Q8NGK7	37	x					x	
tr|Q8TDS7	37	x						
tr|Q91ZB8	35		x					
tr|Q91ZC7	45		x	x			x	
tr|Q91WW2	42	x	x	x			x	x
tr|Q7TN47	42	x					x	x
tr|Q7TN50	39	x	x					
tr|Q7TN41	34		x					
tr|Q91ZB5	23	x						
tr|Q91ZB7	32		x		x		x	
tr|Q7TN40	30		x					
tr|Q7TN43	44		x	x				
tr|Q91ZC1	36		x					
tr|Q7TN51	34		x		x			
sp|MRG_HUMAN	29		x					
tr|Q7TN44	40		x					
tr|Q8IXE2	31				x			x
tr|Q7TN46	42				x		x	x
tr|Q8N7J6	33				x			x
sp|MRGF_HUMAN	31				x			x
sp|MRGF_RAT	31							x
sp|MRGF_MOUSE	31							x

[a] w.r.t. the sequence of mMrgC11

Table S2.2 Calculated energies (in kcal/mol) of configurations generated in combinatorial rotations of TM3, 5 and 6; the rotational angle of TM3 was scanned for 360 degrees (in 30 degree increments)

TM5	TM6	-150	-120	-90	-60	-30	0	30	60	90	120	150	180
0	**0**	613.5	603.1	627.3	644.4	631.5	575.9	615.7	604.7	656.5	702.2	674.7	645.0
0	**-30**	632.9	585.6	626.5	633.4	652.2	579.3	593.9	593.6	635.3	706.1	683.0	593.0
0	**30**	610.7	597.1	633.8	618.1	642.8	589.1	608.6	593.7	634.9	679.9	670.2	621.6
0	**-60**	627.0	606.7	669.5	641.1	686.5	598.7	638.6	641.9	657.2	742.7	678.5	663.4
0	**60**	609.9	588.9	653.4	626.7	631.6	586.4	694.7	593.8	636.3	676.6	729.2	612.2
-30	**0**	514.4	566.8	599.9	568.0	595.5	555.5	594.9	571.3	570.8	623.7	624.1	616.5
-30	**-30**	552.1	569.9	580.8	557.2	612.1	566.0	567.9	569.3	571.6	604.1	606.5	561.1
-30	**30**	545.4	549.1	605.0	542.8	597.2	569.4	593.8	574.5	581.1	624.9	629.9	577.5
-30	**-60**	555.4	560.5	643.8	552.6	610.8	596.2	593.6	593.3	643.2	686.4	678.4	609.2
-30	**60**	534.2	558.9	610.2	553.3	600.0	547.8	599.5	569.3	560.0	627.8	631.9	580.9
30	**0**	593.3	554.0	616.0	581.1	571.7	551.1	585.6	594.1	607.7	645.8	649.7	598.3
30	**-30**	609.5	538.9	609.6	531.6	551.6	509.1	573.2	600.4	576.0	606.7	607.8	567.2
30	**30**	595.0	539.8	609.3	553.5	555.3	546.1	640.1	597.0	581.9	624.6	618.1	585.3
30	**-60**	841.2	569.8	646.4	596.6	583.8	566.5	618.5	604.0	594.8	650.5	745.2	661.6
30	**60**	624.0	575.4	627.3	555.8	572.1	571.7	583.4	658.1	555.4	635.2	667.7	607.3
-60	**0**	550.2	526.2	546.3	547.7	536.7	478.3	584.3	561.9	583.6	634.3	587.1	565.2
-60	**-30**	565.5	479.0	513.7	544.1	537.9	511.8	540.8	552.3	596.1	580.7	564.4	531.2
-60	**30**	558.0	551.4	532.3	519.8	555.2	488.3	560.7	555.2	557.7	612.9	569.3	543.0
-60	**-60**	557.4	524.0	562.2	553.9	587.8	574.4	615.6	558.8	606.9	609.3	573.2	601.3
-60	**60**	520.5	556.1	545.6	533.6	545.3	536.8	561.1	571.6	588.4	621.0	606.1	586.5
60	**0**	605.5	564.7	632.0	609.2	588.3	551.0	600.1	575.0	604.0	673.8	655.2	634.4
60	**-30**	611.1	542.6	601.8	549.0	569.0	563.0	1082.5	557.3	580.7	747.9	641.6	560.2
60	**30**	614.4	617.6	626.6	545.2	592.0	556.3	698.1	574.1	578.9	693.8	631.5	624.1
60	**-60**	613.3	588.0	664.1	554.4	619.8	587.1	616.1	583.0	591.5	694.0	658.7	634.5
60	**60**	616.3	638.9	681.2	569.1	599.2	607.3	606.9	591.2	590.9	704.2	658.1	623.5

Table S2.3 Calculated binding energy (in kcal/mol) and its component contribution for ligands in mMrgC11; the binding energy of RFa in mMrgA1 is also included on the last row for comparison

Ligand	B.E.	Coulomb	VDW	Hbonds	Desolvation	EC50, nM
FMRFa	-109	-83	-34	-81	90	168 ± 26
dFMRFa	-103	-75	-35	-79	86	276 ± 113
FdMRFa	-117	-90	-33	-82	88	113 ± 37
FMdRFa	-78	-46	-51	-56	75	inactive
FMRdFa	-80	-60	-43	-61	84	inactive
acetylated RFa	-97	-67	-19	-45	34	
acetylated dRFa	-82	-48	-25	-40	31	
acetylated RdFa	-75	-49	-27	-29	30	
RF	-71	-80	-19	-47	75	1255 ± 478
RFa	-74	-86	-19	-62	94	682 ± 371
RFa/mMrgA1	58	-0.03	-15	-20	93	inactive