Prediction of the 3D Structure of FMRF-amide Neuropeptides Bound to the Mouse MrgC11 GPCR and Experimental Validation

ARTICLE *in* CHEMBIOCHEM · SEPTEMBER 2007

Impact Factor: 3.09 · DOI: 10.1002/cbic.200700188 · Source: PubMed

6 AUTHORS, INCLUDING:

Peter M Kekenes-Huskey
University of Kentucky
38 PUBLICATIONS **293** CITATIONS

SEE PROFILE

William A. Goddard
California Institute of Technology
1,329 PUBLICATIONS **67,304** CITATIONS

SEE PROFILE

DOI: 10.1002/cbic.200700188

Prediction of the 3D Structure of FMRF-amide Neuropeptides Bound to the Mouse MrgC11 GPCR and Experimental Validation

Jiyoung Heo,[a] Sang-Kyou Han,[b] Nagarajan Vaidehi,[a, c] John Wendel,[a] Peter Kekenes-Huskey,[a] and William A. Goddard, III*[a]

We report the 3D structure predicted for the mouse MrgC11 (mMrgC11) receptor by using the MembStruk computational protocol, and the predicted binding site for the F-M-R-F-NH_2 neuropeptide together with four singly chirally modified ligands. We predicted that the R-F-NH_2 part of the tetrapeptide sticks down into the protein between the transmembrane (TM) domains 3, 4, 5, and 6. The Phe (F-NH_2) interacted favorably with Tyr110 (TM3), while the Arg makes salt bridges to Asp161 (TM4) and Asp179 (TM5). We predicted that the Met extends from the binding site, but the terminal Phe residue sticks back into an aromatic/hydrophobic site flanked by Tyr237, Leu238, Leu240, and Tyr256 (TM6), and Trp162 (TM4). We carried out subsequent mutagenesis experiments followed by intracellular calcium-release assays that demonstrated the dramatic decrease in activity for the Tyr110Ala, Asp161Ala, and Asp179Ala substitutions, which was predicted by our model. These experiments provide strong evidence that our predicted G protein-coupled receptor (GPCR) structure is sufficiently accurate to identify binding sites for selective ligands. Similar studies were made with the mMrgA1 receptor, which did not bind the R-F-NH_2 dipeptide; we explain this to be due to the increased hydrophobic character of the binding pocket in mMrgA1.

Introduction

The family of G protein-coupled receptors (GPCR) exhibit cell-signaling functions that range from receptors for sensory signals (odors, tastes, or light) and for endogenous signals (including peptides, lipids, neurotransmitters, or nucleotides), which makes them a most important class of membrane proteins. The physiological significance of these ligands has made many of these GPCRs prime drug targets for the pharmaceutical industry. At least 720 human genes belong to the GPCR superfamily and about half of them are thought to encode sensory receptors.[1] Among the remaining ~360 genes, the endogenous ligand has been identified for around 210 receptors, which leaves ~150 receptors for which the ligands remain unknown (orphan receptors). These orphan receptors might play important, albeit unknown, functions in various cells, and some could be useful as new drug targets. To understand their possible significance, it is important to characterize these orphan receptors by identifying their endogenous ligands and small-molecule ligands that can selectively agonize or antagonize them.

Recently Dong et al.[2] and Lembo et al.[3] identified a novel family of GPCRs called the Mas-related gene (Mrg) receptor for mice, or the sensory neuron specific receptor (SNSR) in mice and human. A subset of these receptors, which include the mouse MrgA1 (mMrgA1) and mouse MrgC11 (mMrgC11), is distributed mainly to isolectin $B4^+$, small diameter nociceptors in the dorsal root ganglia (DRG), which are suggested to be involved in pain sensation or modulation. Although their endogenous ligands and precise physiological function remain unknown, distinct and selective ligands that activate some of these receptors have been identified:

The neuropeptide RF amides are potent ligands for mouse Mrg receptors, for example, NPFF (actual sequence: FLFQPQRF-NH_2) for MrgA1 (EC_{50} ~200 nM) and MrgC11 (EC_{50} ~54 nM), and NPAF for MrgA4 (EC_{50} ~60 nM).[2,4] BAM22 (with the QKRYG carboxy terminus) derived from preproenkephalin A, one of the endogenous opioid peptides, activates SNSR3 (EC_{50} ~13 nM) and SNSR4 (EC_{50} ~16 nM).[3] More recently Grazzini et al. observed that γ2-MSH (with a carboxyl terminus of WDRFG) is highly potent in rat MrgC receptor and the active moiety rec-

[a] *Dr. J. Heo, Dr. N. Vaidehi, J. Wendel, P. Kekenes-Huskey, Prof. W. A. Goddard, III*
Materials and Process Simulations Center (139-74)
Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, CA 91125 (USA)
Fax: (+1)626-585-0918
E-mail: wag@wag.caltech.edu

[b] *Dr. S.-K. Han*
Division of Biology (147-75)
California Institute of Technology
Pasadena, CA 91125 (USA)

[c] *Dr. N. Vaidehi*
Present address:
City of Hope Graduate School of Biological Sciences
Division of Immunology, Beckman Research Institute of City of Hope
Duarte, CA 91010 (USA)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

ognized by rat MrgC receptor is the C-terminal RF-amide motif of γ2-MSH.[5] Cortistatin has been identified as a highly potent ligand for human MrgX2 (EC_{50} ~25 nM).[6] In addition adenine shows high affinity (K_i ~18 nM) and potency for rat MrgA receptor.[7] Recent studies have also shown that MrgD receptors specifically respond to micromolar concentrations of β-alanine.[8] The ability of MrgC11 to be agonized by F-M-R-F-NH_2 (FMRFa) and R-F-NH_2 (RFa) is interesting because these neuropeptides play an important neurosensory role in insects, but a role in mammals is not known.

Clearly, it would be most useful to have the 3D structure of the receptor to help select the most promising new ligands for experimental assays. Unfortunately, it has not been possible to obtain the crystals required for crystallography for any of these systems and hence there is no information about the structure or binding site from experiment. To provide such structures we used the MembStruk computational method[9,10] to predict the 3D structure of the mMrgC11 and mMrgA1 receptors, and we used the HierDock computational method[11] to predict the binding sites of the di- and tetrapeptide ligands that contain the RF-amide motif, which have been identified as agonist for these receptors.[4] These methods have predicted reliable structures and accurate binding sites for other GPCRs for which mutation and binding studies were available to provide empirical information about the binding site.[12,13] No such information was available for mMrgC11 and mMrgA1, which provided a much more stringent test of the methods.

Our structure predictions identified three residues, Tyr110 (TM3), Asp161 (TM4), and Asp179 (TM5), that play key roles in binding of the FMRFa and RFa ligands. Based on these predictions, we carried out mutagenesis experiments in which each of these three residues was mutated to Ala, and measured the intracellular calcium release elicited from agonist binding. In addition, we carried out experiments for other mutations (Tyr110Trp and Tyr110Phe), which were expected to have more modest effects on binding. These experiments validated our predictions; this suggests that the 3D structure of these proteins and the predicted binding site for the ligands were reasonably accurate. Armed with the binding site, one can now search for small-molecule nonpeptide agonists or antagonists that might be useful for elucidating the pharmacology and function of these systems, and perhaps to identify the endogenous ligand.

Results and Discussion

Characteristics of the mMrgC11 receptor structure

The predicted TM regions for mMrgC11 are given in Figure 1 and the predicted 3D structure of the mMrgC11 receptor is shown in Figure 2. We predicted that Pro223 (TM6) and Pro271

```
                                                          TM1
mMrgC11 MDPTISSHDTESTPLN-ETGHPNCTPILTLSFLVLITTLVGLAGNTIVLWLLGFRMRRKA 59
mMrgA1  --------------MDNTIPGGINITILIPNLMIIIFGLVGLTGNGIVFWLLGFCLHRNA 46
            TM2                                                TM3
mMrgC11 ISVYILNLALADSFFLCCHFIDSLLRIIDFYGLYAHKLSKDILGNAAIIPYISGLSILSA 119
mMrgA1  FSVYILNLALADFFFLLGHIIDSILLLLNVF--YP-ITFLLCFYTIMMVLYIAGLSMLSA 103
                                  TM4
mMrgC11 ISTERCLCVLWPIWYHCHRPRNMSAIICALIWVLSFLMGILDWF-SGFLGETHHH-LWKN 177
mMrgA1  ISTERCLSVLCPIWYHCHRPEHTSTVMCAVIWVLSLLICILNSYFCGFLNTQYKNENGCL 163
            TM5                                                TM6
mMrgC11 -VDFIITAFLIFLFMLLSGSSLALLLRILCGPRRKPLSRLYVTIALTVMVYLICGLPLGL 236
mMrgA1  ALNFFTAAYLMFLFVVLCLSSLALVARLFCGTGQIKLTRLYVTIILSILVFLLCGLPFGI 223
                          TM7
mMrgC11 YLFLLYWFGVHLHYPFCHIYQVTAVLSCVNSSANPIIYFLVGSFRQHRKHRSLKRVLKRA 296
mMrgA1  HWFLLFKIKDDFHVFDLGFYLASVVLTAINSCANPIIYFFVGSFRHRLKHQTLKMVLQNA 283

mMrgC11 LEDTPEEDEYTDSHLHKTTEISESRY 322
mMrgA1  LQDTPET---AKIMVEMSRSKSEP-- 304
```

Figure 1. The predicted transmembrane (TM) regions. The sequence alignment of mMrgC11 and mMrgA1 was based on the alignment with the entire set of sequences obtained by using the BLAST search with the mMrgC11 sequence (see Figure S1 in the Supporting Information). The hydrophobic center of each TM is indicated with an arrow. The residues involved in the mutagenesis experiments are in bold font and highlighted.

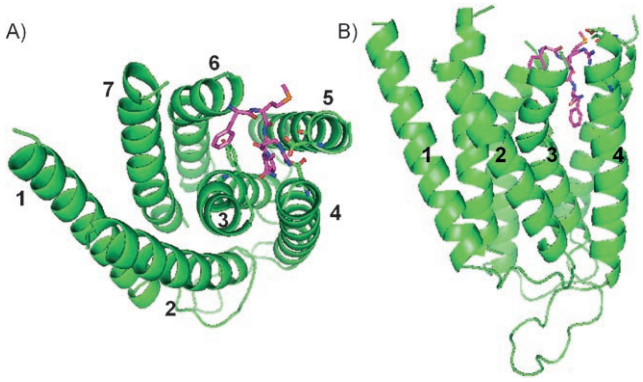

Figure 2. The predicted 3D structure for the FMRFa–mMrgC11 complex. The three key residues (Tyr110, Asp161, and Asp179) are shown in a stick format, and the TM regions are numbered. A) Top view from extracellular region; B) side view with the extracellular region at the top.

(TM7) bend the corresponding helices at 28° and 15°, respectively. These prolines are highly conserved amongst family A GPCRs, which include bovine rhodopsin (which has bends of 24° and 33°, respectively). Moreover, Pro109 in the middle of TM3 leads to a 23° bend (13° in rhodopsin at the Gly–Gly site). We predicted 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 only GPCR for which there is an experimental 3D X-ray crystal structure is bovine rhodopsin[14] (PDB ID: 1U19, 2.2 Å resolution). As expected from the low (22%) sequence identity for the TM regions of mMrgC11 with rhodopsin, our predicted structure for mMrgC11 differs significantly from bovine rhodopsin, with a coordinate root mean square (CRMS) difference in the TM Cα atoms of 3.75 Å.

Several conserved residues participated in the interhelical hydrogen bonds to maintain the stability of the mMrgC11. Thus:

- Asn44 (TM1; highly conserved in the whole of family A GPCRs) formed a hydrogen bond with the Ser268 carbonyl group of the backbone in TM7 (Figure 3A). In the inactive form of rhodopsin[15] Asp71 (TM2; Asp83 in rhodopsin) forms an interhelical hydrogen bond with this Asn44 (Asn55 in rhodopsin). However, in MrgC11 we predicted that Asp71 is in proximity, but not in hydrogen-bond contact, with Asn44. Such differences are plausible since Miura et al. have reported TM2 movement from activation in angiotensin II type 1 receptor (by using substituted cysteine accessibility mapping).[16] Thus, Asp in TM2 might interact differently for an active GPCR, compared to one in the inactive rhodopsin structure; this suggests that our predicted structure might be in the active state.
- Asn66 (TM2) and Trp151 (TM4; both highly conserved in the family A GPCRs) formed a hydrogen bond just like the analogous pair in rhodopsin (Asn78–Trp161).
- Tyr63 (TM2; a highly conserved residue in the Mrg receptor family across the 39 sequences available on Swiss-Prot and TrEMBL) participated in hydrogen bonding with Ser112 (TM3; Figure 3A).
- Another conserved residue among Mrg receptors, Ser143 (TM4), formed a hydrogen bond with the hydroxyl group in Thr122 (TM3; Figure 3A).
- Arg215 (TM6) contacted the backbone carbonyl group of Val277 in TM7 (Figure 3A).

We predicted several other interhelical hydrogen bonds that involved nonconserved hydrophilic residues. Asp179 (TM5) and Asp161 (TM4), which we identified as key residues for ligand binding, were in contact with Lys99 (TM3) and Thr183 (TM5), respectively, in the apo protein. Most other hydrogen bonds were found in the TM regions near the intracellular loop. These regions packed more compactly, while the near-extracellular regions were wider—as is appropriate for ligand binding.

No direct contact (hydrogen bond or salt bridge) was found between TM3 and TM6 or between TM3 and TM7 in the TM regions of MrgC11. However, these TM helices interacted with each other through well-stacked aromatic rings (Figure 3B). These included: Tyr227–Phe190, Phe190–Tyr110, Tyr110–Tyr237, and Tyr237–Tyr256. Tyr110 (TM3), one of the aromatic residues that participated in these interactions, is conserved among the Mrg receptors (5 of 39 have Phe at this position instead of Tyr). In addition, Trp in TM6, which is known to be involved in activating rhodopsin, is not found in the mMrgC11 receptor, in which it is replaced with Gly. Thus, activation in mMrgC11 might involve a different mechanism than that of rhodopsin. As discussed later, we found that the agonists to MrgC11 bind in the pocket located between TM3, 4, 5, and 6. Possibly the binding of the ligand to this pocket affects these aromatic–aromatic interactions to help induce activation.

Description of the binding sites

The predicted RFa binding site is located between TM3, TM4, TM5, and TM6. In contrast to the horizontal orientation of 11-*cis* retinal in rhodopsin, we predicted that RFa orients vertically in the binding pocket. As discussed above, the interhelical stacked aromatic rings present between TM3 and TM6 do not allow the ligand to cross into the TM1-2-3-7 pocket. A similar orientation has been suggested for the formylated peptide, fMLF,[17] which binds parallel to the average helix of the formyl peptide receptor (FPR). Since RFa is a small peptide ligand (like fMLF) it could also have fitted parallel in the pocket but for the longer peptides that bind to MrgC11, this horizontal placement would require the additional amino acids to kink towards TM2 and TM7; this leads to contact with these TMs mainly in the loop regions.

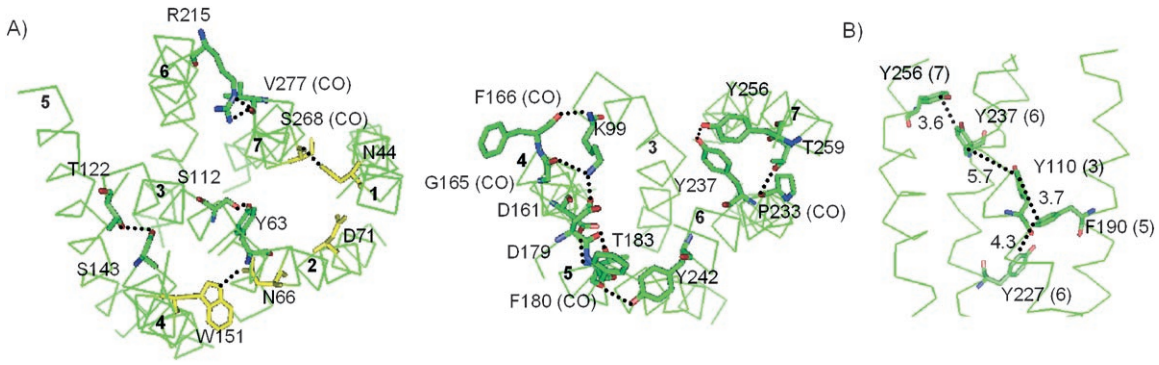

Figure 3. A) The interhelical hydrogen bond network in the mMrgC11 receptor (dashed lines). The highly conserved residues in the family A of GPCRs that form interhelical hydrogen bonds in rhodopsin are in yellow. Left: viewed from the intracellular region; right: viewed from the extracellular region. The HBPLUS[37] program was used to calculate hydrogen bonds (maximum D–A distance = 3.9 Å, minimum D–H–A angle = 90.0°; where D is the donor heavy atom, H the hydrogen, and A the acceptor). The TM regions are indicated in bold numbers. B) Interhelical aromatic π-stacking interactions in TM regions of mMrgC11 receptor. The closest C–C distance between two benzyl rings is shown in Å. The TM regions are indicated in brackets.

Predicted binding site of the dipeptides: The detailed interactions of bound dipeptides with mMrgC11 receptors are described in Figure 4. The binding mode of R-F-OH (RF) was similar to R-F-NH_2 (RFa) although the rotamer conformations of

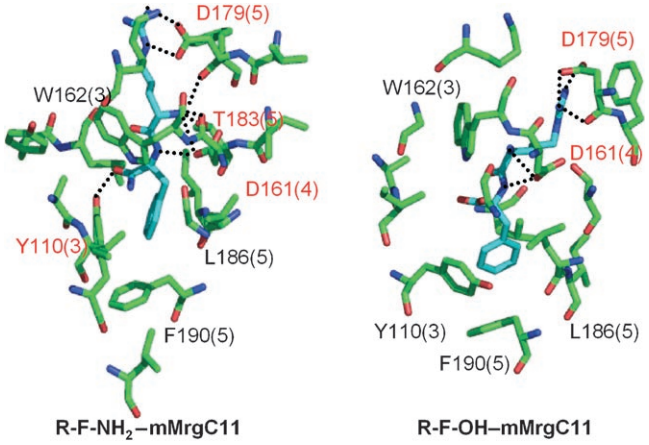

Figure 4. The predicted 5 Å binding pocket of the RFa and RF dipeptide agonists to mMrgC11. The intermolecular hydrogen bonds, which were calculated with explicit hydrogens by using the same criteria as in Figure 3A, are indicated by dotted lines.[38] A residue the side-chain of which participates in the hydrogen bond is specified in red, while one the backbone of which is involved is in blue. Residues that show good hydrophobic interactions are specified in black. The top of the picture corresponds to the extracellular regions; the TM regions are indicated in brackets.

certain residues were different. The common features were that the positively charged moieties were stabilized through the salt bridges and other hydrophilic interactions. Thus:

- The Arg of the ligand had a good electrostatic interaction with Asp179 (TM5).
- The N terminus had a good electrostatic interaction with Asp161 in TM4 ($d(N^{+}\text{-}O(COO^{-}))=2.97$ or 3.02 Å).
- The N terminus of RFa also formed a hydrogen bond with the hydroxyl group of Thr183 (TM5).
- In addition, the C terminus of RFa formed a hydrogen bond with the hydroxyl group of Tyr110 (TM3).

The phenyl group of the Phe in the ligand was stabilized by several aromatic residues present in the binding pocket:

- Tyr110 interacted most closely with both ligands. For RFa the phenyl rings were in a sandwiched geometry with Tyr110 while for RF these two rings had the displaced herringbone or T shape.
- Phe190 (TM5) also had a good π–π interaction with Phe of the ligand, while Leu186 (TM5) also contributed 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: Using HierDock with the binding region identified from the RFa studies, we predicted the binding site for the three FMRFa tetrapeptides known to be good agonists (~100 nM) for mMrgC11.[4] We predicted a similar binding mode for all three, with the common C-terminal dipeptide part parallel to the average helical axis and the C terminus of the peptide toward the intercellular region. We calculated F-(D)M-R-F-NH_2 to bind strongest, and F-M-R-F-NH_2 and (D)F-M-R-F-NH_2 to have binding energies that were just 7 and 11% weaker, respectively. We will discuss them in the order of bond strength.

For F-(D)M-R-F-NH_2, denoted FdMRFa (Figure 5), in which the chirality of Met was right handed, the F–M region of the tetrapeptide stretched out horizontally toward TM6. Here:

- The amide at the C terminus formed hydrogen bonds with the side chain of Asp161 and the backbone carbonyl group of Gly158.
- The Phe at the C terminus resided in a good aromatic and hydrophobic environment formed by Tyr110, Phe190, and Leu186.
- The Arg was stabilized through electrostatic interactions with Asp161 (the distance between the N of the side chain in Arg and carboxylate oxygen of Asp161 was 2.87 Å) and Asp179. Thr183 also interacted with the side chain of Arg. In addition, Asp161 formed a hydrogen bond with a nitrogen atom of the backbone.
- The Met located in the peripheral region between TM5 and TM6 was near the hydrophobic residues Leu238, Phe239, and Ile187, but had no specific interaction.
- The N-terminal Phe was sandwiched between Trp162 (TM4) and Tyr237 (TM6); this led to favorable aromatic interactions. The N terminus was exposed to the extracellular region (Figure 5). 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.

For FMRFa, the overall binding mode was similar to FdMRFa. Some differences were that Thr183 no longer participated in hydrogen bonding with a peptide ligand and the side chain of the right-handed Met was closer to TM5 and interacted at the edge of the aromatic ring of Phe180 (S–C distance = 4.0 Å). Such a preference of S atoms at the edge of aromatic ring has been observed in the study of the nonbond interaction that involve the Met sulfur atom, by analyzing the protein crystal structures[18] (Figure 6).

The (D)F-M-R-F-NH_2 (dFMRFa) ligand showed similar binding interactions to FMRFa and FdMRFa. Although the N-terminal Phe had a different chirality from the latter two ligands, it led to a similar conformation of the side chain, which fit in between Trp162 and Tyr237. In this case the Met led to an intra-residue S···O interaction and an inter-residue interaction with Leu240, in which the sulfur atom behaved as an electrophile[18,19] (Figure 6).

Predicted binding sites of the nonagonists, F-M-(D)R-F-NH_2 and F-M-R-(D)F-NH_2: The experiments showed that the two other chirally modified FMRFa peptides, FMdRFa and FMRdFa, did not agonize mMrgC11 (up to 10 μM concentration) and could not compete with the known agonists. We predicted much

weaker binding energies (by 34% for FMdRFa and by 32% for FMRdFa) compared with FdMRFa, which is consistent with experimental data. Our predicted binding sites (residues within the 5 Å of the ligand) for these two cases are shown in Figure 7. Overall, these two nonagonist peptides showed similar binding characteristics to the agonist peptides. Points to note include:

- In both cases, the C-terminal Phe interacted with Tyr110 and Phe190, as seen for agonists.
- In FMdRFa, the side chain of Arg was located near Asp161 and Asp179, with favorable electrostatic interactions. However, the contact was less tight and the nonbond interaction energies with Asp161 and Asp179 decreased by 41 and 12%, respectively, compared with FdMRFa. We saw the intraresidue S···O interaction for Met in this case. The N-terminal Phe lost the π–π interaction with Trp162.
- In FMRdFa, the side chain of Arg was between Asp161 and Asp179, but the interaction was weaker than with FdMRFa. The sulfur of Met interacted with the backbone carbonyl group of Asp179. The N-terminal Phe was sandwiched with Trp162 and Tyr237.

Summary of binding sites: These studies identified several residues critical for ligand binding.

- The two aspartic acids, Asp161 and Asp179, contributed to favorable electrostatic interactions for the electropositive groups of the ligands: Arg for tetrapeptides and Arg and N terminus for dipeptides.
- Several aromatic residues contributed to favorable π–π interactions. Tyr110 and Phe190 had contact with the common C-terminal Phe of all five agonists.

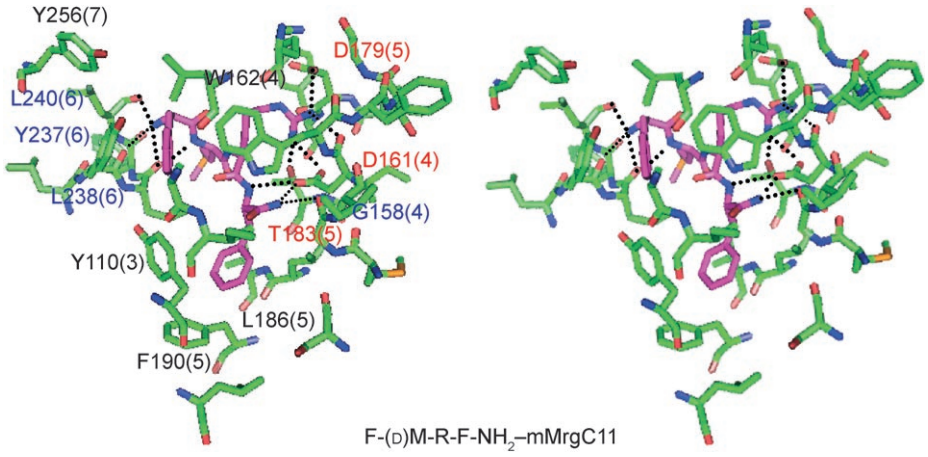

Figure 5. Stereo view of the detailed contacts in 5 Å binding pocket of F-(D)M-R-F-NH_2.[39] The TM regions are indicated in brackets.

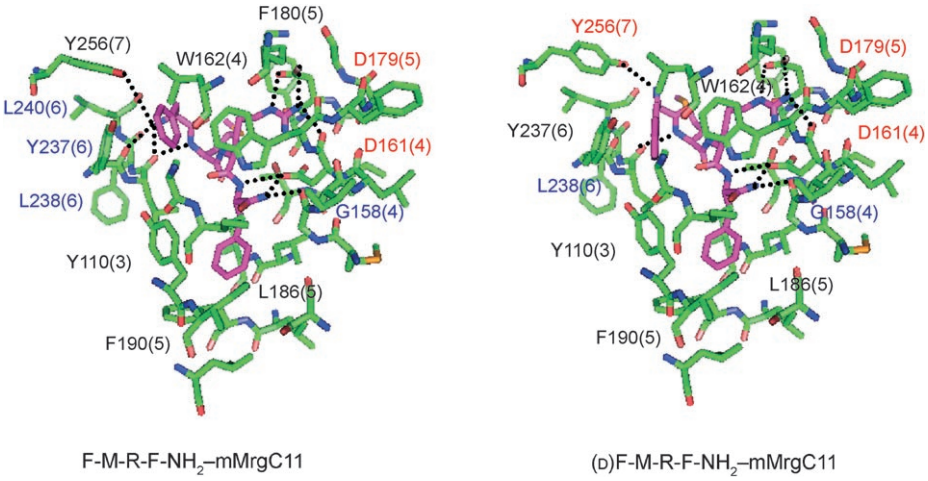

Figure 6. The predicted 5 Å binding site to mMrgC11 of the agonist tetrapeptides, F-M-R-F-NH_2 and (D)F-M-R-F-NH_2; the TM regions are indicated in brackets.

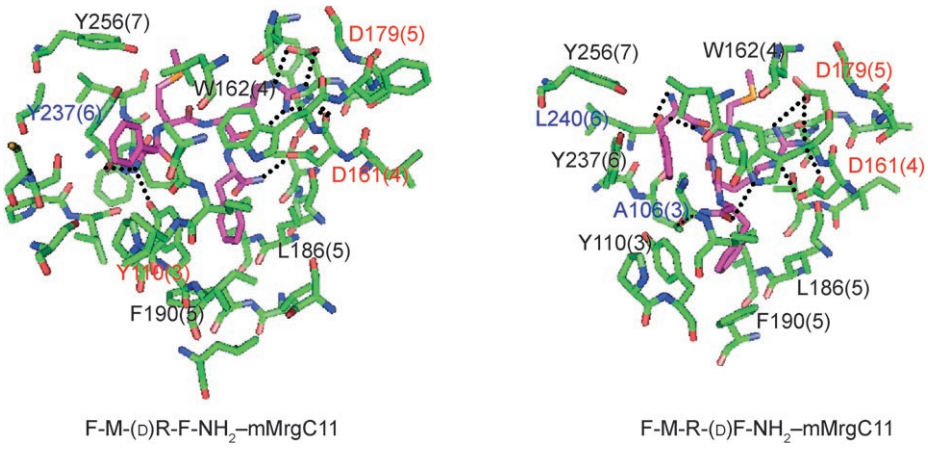

Figure 7. The predicted 5 Å binding pocket of the nonagonist tetrapeptides, F-M-(D)R-F-NH_2 and F-M-R-(D)F-NH_2; neither case was observed to experimentally bind even at 10 μM; the TM regions are indicated in brackets.

- In the tetrapeptide agonists, the additional phenyl group interacted with Trp162 and Tyr237. As mentioned previously, these aromatic residues were stacked well in the receptor in the absence of a ligand, and provided the interhelical interactions among TM3, 5, 6, and 7. It is plausible that this coupling with two phenyl groups of the tetrapeptide ligand along with the strong electrostatic interaction of Arg with Asp161 and Asp179 could induce the conformational change responsible for the activation.

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

Results from Mutagenesis experiments

Expression and localization of mMrgC11 wild-type and mutant receptors: We carried out three sets of experiments in which key residues were substituted to alanine: Tyr110Ala, Asp161Ala, and Asp179Ala. Figure 8A shows the images for mMrgC11; the wild-type and three mutant receptors were fused with green fluorescent protein (GFP) for visualization. All mutants, including the wild-type receptor predominantly appeared on the plasma membrane with some intracellular fluorescence mostly localized in presumable transport vesicles that exit from the endoplasmic reticulum. All mutant cells showed fluorescence signals that were as intense as the cells expressing the wild-type receptor. These images indicate that the mutant receptors were expressed at a level similar to that of the wild-type receptor, and were localized at the cell membrane.

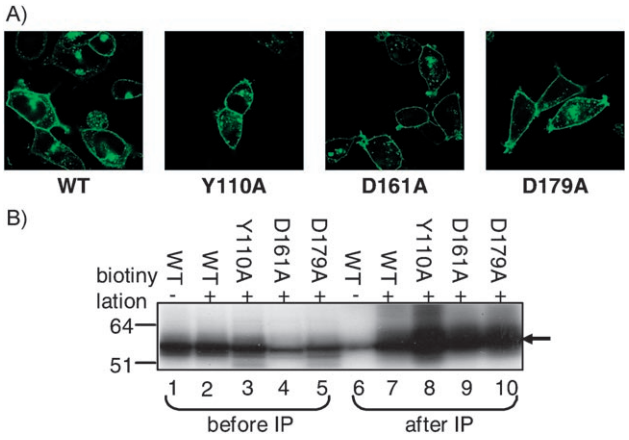

Figure 8. Expression of mMrgC11 wild-type and mutant receptors in Flp-In293 cells. A) Images of cells that expressed wild-type (WT) or mutant receptors; GFP was fused to the C terminus of the receptor. B) Biotinylation of the cell surface, where receptors are localized and folded. The biotinylated cell extract was blotted with anti-GFP after immunoprecipitation (IP) with streptavidin. Lanes 1–5: before IP, and lanes 6–10: after IP. The molecular weight markers are shown on the left in kDa.

To eliminate improper delivery of mutant receptors to the cell membrane as a possible cause for the lack of activity, we combined immunoprecipitation (IP) experiments with biotinylation. Lanes 1–5 in Figure 8B show total mMrgC11 receptor expression, including those that are not biotinylated but present in the cytosol, and those that have not crossed properly through the membranes. These blots again indicate that all three mutants were adequately expressed in the cells, although the expression levels of Asp161Ala and Asp179Ala mutants appeared to be slightly lower. The results of blots after IP with streptavidin (lanes 6–10) showed that the mutant receptors localized on the cell membrane and took apical positions at similar amounts to the wild-type receptor. Since the band corresponding to nonspecific binding of streptavidin in the wild-type cells (lane 6) was much weaker, we conclude that the major portion of the signals on the blots in lanes 7–10 come from biotin-specific binding. This suggests that the mutant proteins were properly inserted in the membrane, as the wild-type protein.

Dose-dependent intracellular calcium release assay with stably expressed MrgC11 receptors: Table 1 shows the EC_{50} values of various peptide ligands determined by intracellular calcium release assays with Flp-In293 cells that expressed the mMrgC11 receptors. Di- and tetrapeptides, and some longer peptide agonists, were selected from the ligands previously identified by Han et al.[4] We obtained slightly higher EC_{50} values in our cellular system compared to 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.[20] Nonetheless, the selectivity observed in this study is consistent with previous results, for example, FMdRFa and FMRdFa still showed no activity.

Table 1. The EC_{50} values [nM] of various peptide ligands measured by using the intracellular calcium-release assay with Flp-In293 cells that expressed the mMrgC11 receptor.

Peptide	Sequence	This work	Ref. [4]
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

[a] Inactive: activation was not detected up to the highest concentration tested (10 μM). Data represent the mean (± SEM) of four independent experiments. [b] Comparison was made to previous results by Han et al. in ref. [4] (different cell line).

Results for Tyr110Ala, Asp161Ala, and Asp179Ala mutant receptors: Out of the twelve ligands tested for the wild-type receptor, we selected the six most potent ligands to measure the

potencies for Tyr110Ala, Asp161Ala, and Asp179Ala mutant receptors (Table 2). Key results are:

- We found that the Tyr110Ala mutant was not activated by any of the six tested ligands (up to a concentration of 33 μM); this indicates that Tyr110 is critical for activation.
- The Asp179Ala mutant showed no potency for the three tetrapeptide ligands, while the other three ligands were only activated with ten-times higher concentration of ligand.
- For the Asp161Ala mutant we found that four of the six ligands no longer caused activation, while the other two only activated with 100-times higher concentrations.

These results, which show that mutation of Tyr110, Asp161, or Asp179 very strongly reduce or eliminate the activity of the mMrgC11 receptor, are consistent with the predictions that these residues are involved in ligand binding. This provides strong support for the overall 3D structure predicted for mMrgC11.

Control experiments: For a positive control experiment, we transiently expressed the Phe74Ala (TM2) and Asp81Ala (TM2) mutants in HEK293 cells along with the Tyr110Ala, Asp161Ala, and Asp179Ala mutant receptors, which were also transiently expressed under the same condition. Phe74 and Asp81 were chosen since they are aromatic or negatively charged residues predicted to be in the pocket but without contact to the ligand. The intracellular calcium-release assay was then carried out with FMRFa (0.33 μM). The activity of Phe74Ala and Asp81Ala mutants was intact, and the other three mutants showed no activity. This is consistent with binding being restricted to the pocket that includes the Tyr110, Asp161, and Asp179 residues of TM3, 4, and 5.

Role of OH in Tyr110: We also investigated the implication of the hydroxyl group in Tyr110 in ligand recognition by substituting this tyrosine to phenylalanine or tryptophan. For the three tetrapeptide agonists Tyr110Phe or Tyr110Trp mutations led to a factor of 5–10 reduction in potency (Table 2). This is consistent with our predicted structure (Figures 5 and 6) in which the hydroxyl group of Tyr110 did not interact with the ligand, but instead formed a hydrogen bond with the carbonyl group of the backbone. The missing hydroxyl group should result in a dangling hydrogen bond donor that might induce a slight overall conformational change in the binding pocket; this explains the slight loss in activity. Since substitution of Tyr110 to Ala totally extinguished the activity for all six ligands, we conclude that the aromatic ring must be significant for all six cases.

The potencies of the γ1- and γ2-MSH ligands were not affected by the absence of the hydroxyl group; this indicates that the hydroxyl group does not contribute to ligand activation for these ligands. This in turn suggests a slightly different binding conformation.

Competition by FMdRFa and FMRdFa of agonist binding: To investigate whether the two nonagonist tetrapeptides, FMdRFa and FMRdFa, are antagonists or weak binders (or non-binders), we saturated the receptors either with FMdRFa or FMRdFa by using three concentrations, 3.3, 16, and 33 μM, and then measured the EC_{50} value for FdMRFa. The intensity of calcium signal remained on the same level as in the absence of FMdRFa or FMRdFa, and the EC_{50} values changed little (within standard deviation). This result shows that FMdRFa and FMRdFa did not block the efficacy of FdMRFa, and indicates that they did not bind to a site related to agonist binding or activation.

In summary, the experimental results validated that Tyr110, Asp161, and Asp179 are involved significantly in ligand potency, as expected from the predicted structures. Of course, loss of function in a mutant might be associated with other effects instead of the loss of interaction with a residue in the binding site, for example, changes in the interactions required to mediate conformational change for activation. However, except for Tyr110Ala, all mutations led to some activity for some ligands; this indicates that activation was still possible, albeit reduced by a factor of ~100. In addition, the Tyr110Trp and Tyr110Phe substitutions only led to reductions of factors of 5–10, as could be expected from such slight changes in structure. Thus we conclude that these three residues are involved in ligand binding, and hence the loss of activity observed in our experiment was related with the decrease of binding affinity due to the mutations.

Thus, we consider that the experiments validate the accuracy of our predicted 3D structure for MrgC11, and the accuracy in the predicted ligand-binding site.

Table 2. The EC_{50} values [nM] measured for mutant mMrgC11 receptors by using the intracellular calcium-release assays.

Peptide	Binding[a]	Wild-type	Tyr110Ala[b]	Asp161Ala[b]	Asp179Ala[b]	Tyr110Phe[c]	Tyr110Trp[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 = 117 kcal mol^{-1}). [b] The activity was measured in four independent experiments; + means that activation started at a given concentration (activation was obtained only once in the quadruplicate measurement). [c] Numbers in parentheses are the ratios with respect to the EC_{50} value of wild-type receptor.

The mutations carried out here focused on testing the details of the dipeptide-binding region common to all agonists. The additional residues involved in the binding region for the tetrapeptide suggest new mutations that would test the details of the FM parts of the tetrapeptides. Thus, based on the predicted binding sites Trp162, Phe190, and Tyr237 all play important roles in binding to the terminal Phe in the ligand. Hence, mutation of any one of these to Ala should dramatically decrease binding. Since such mutations involve a big change in side-chain volume, it would also be reasonable to try larger hydrophobic residues, such as Leu or Ile.

For additional validation, radiolabeled ligand-binding assays for either the mutant receptor or mutant peptide ligand would be useful. Such studies would further improve our understanding of the receptor structure and ligand-binding site.

Molecular dynamics (MD) simulation of the fully solvated mMrgC11–F-(D)M-R-F-NH_2 complex in the membrane

Our predictions of the structure of the mMrgC11 protein and of the ligand-binding sites use many simplifications to allow sampling of all possible orientations of the TM domains and of all possible binding sites of the ligands. These involved only two layers of lipid and implicit solvent approximations. In addition, the structures were obtained by minimization of structures obtained from our Monte Carlo sampling. With such a simple description of solvent and membrane we would not use MD simulations at normal temperatures for fear that the ligand would rearrange.

In order to test the stability of our predicted structure, we embedded this predicted structure in an infinite periodic membrane and built the resulting structure into a fully equilibrated water solvent box, much as we reported previously for β2 adrenergic receptor.[21] The result is a box of dimensions 40×40×100 Å with ~47000 atoms. After minimizing these structures, we carried out MD simulations at 300 K for 7 ns. We found no significant changes in the binding site or the positioning of the ligand in this site.

The binding mode of the tetrapeptide F-(D)M-R-F-NH_2 after 7 ns equilibration is shown in Figure 9A. The C terminus amide group maintained the hydrogen bond with the side chain of Asp161 (TM4). The C terminus Phe of the ligand remained positioned in a stabilizing aromatic and hydrophobic environment formed by Tyr110 (TM3), Phe190 (TM5), Leu186 (TM5), and additionally Ile107 (TM3; the latter two are not shown for clarity). The Arg was stabilized through the electrostatic interaction with Asp161 and an additional hydrogen bond with the side chain of Thr183 (TM5). Asp179 in TM5 moved slightly from the Arg due to intervention of water molecules to mediate the hydrogen bond between the side chain of Asp179 and the backbone carbonyl group of the ligand (Figure 9B). The carboxylate group of Asp179 was partially solvated with water molecules and stabilized by the positively charged quaternary amine group of a lipid molecule (distance between N of quaternary amine and carboxylate oxygen of Asp179 = 4.16 Å).

The N-terminal Phe of the ligand remained sandwiched between Trp162 (TM4) and Tyr237 (TM6; the closest C–C distances between two aromatic rings were 4.64 Å and 3.39 Å, respectively). Two more aromatic residues in TM6, Tyr242 and Phe244 (located close to the extracellular loop, not shown here), came into the binding pocket (the closest C–C distances between two aromatic rings were 4.22 Å and 3.40 Å, respectively), and yielded the additional favorable aromatic interaction with the N-terminal Phe of the ligand.

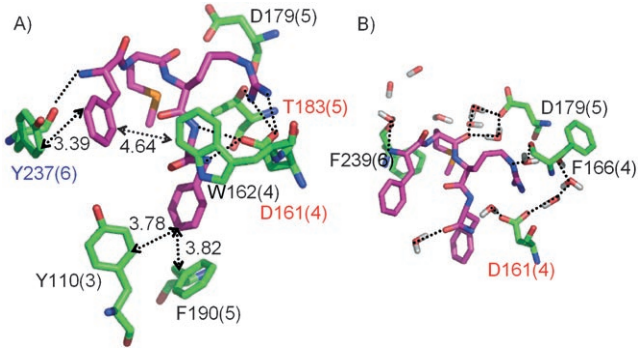

Figure 9. The 5 Å binding site of F-(D)M-R-F-NH_2 in mMrgC11 receptor after 7 ns of MD at 300 K. A) Interactions with key residues: the shortest C–C distance between two aromatic rings is indicated by dotted arrows. The intermolecular hydrogen bonds (calculated with explicit hydrogens by using the same criteria as in Figure 3A) are indicated with dotted lines. B) Water molecules in the 5 Å binding pocket participated in hydrogen bonds with the ligand or receptor. Some water molecules bridged the receptor and ligand. The TM regions are indicated in brackets.

To summarize, we observed that the initial interactions of the ligand with the key residues (Asp161 and four aromatic residues, Tyr110, Phe190, Trp162, and Tyr237) were preserved throughout the entire MD simulation. Asp179 in TM5 now interacted with the ligand through water-mediated hydrogen bonds. Thus, the bound complex remained stable; this indicates that the MembStruk predicted structures are reasonably accurate.

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

The experiments[4] showed that mMrgA1 has a very different binding pattern from mMrgC11. It is not agonized by RFa or RF_{OH} and its binding to the tetrapeptide agonists was 4–100-times weaker than for MrgC11. To provide an additional test of our computational methods we predicted the 3D structure of mMrgA1 and the binding of the ligands using the same methods as for mMrgC11.

3D structure of the protein: We used MembStruk in the same way as with mMrgC11. The sequence identity between MrgC11 and MrgA1 was 53% for the TM regions and 46% for the entire sequence. Since mMrgA1 was included in the set of homologous sequences for the TM prediction of mMrgC11, the same TM prediction was used as shown in Figure 1. The coordinate root mean square deviation (CRMSD) of the Cα atoms between mMrgC11 and mMrgA1 was 2.49 Å in the TM regions, and 4.94 Å if the loops were included.

Binding of ligands to mMrgA1: Since RFa was not observed to bind to mMrgA1, we could not expect that application of

HierDock would find the same binding site as for mMrgC11. Since we wanted to understand why RFa would not bind to this site, we wanted to compare the binding of RFa to the site corresponding to the binding site in mMrgC11. Thus, we matched the Cα atoms of mMrgA1 with the corresponding atoms of the RFa–mMrgC11 complex and then positioned the RFa ligand to generate the spheres for docking. We then docked the RFa ligand in the corresponding pocket surrounded by TM3, 4, 5, and 6 (Figure 10A). We found that Tyr94, Phe177, and Leu173 (homologous residues of Tyr110, Phe190, and Leu186 in mMrgC11) formed a hydrophobic pocket for Phe, as in mMrgC11, but they were located slightly farther (the closet C–C distance between aromatic rings is 4.5 to 5.5 Å). Asn145, the homologous residue of Asp161 in mMrgC11, was involved in hydrogen bonding with the C terminus. The Arg side chain was surrounded by hydrophobic residues (Leu83, Leu84, and Tyr87) and did not have any favorable interactions with the receptor. We calculated a negative binding energy; this indicates that RFa did not bind to the mMrgA1 receptor, which is consistent with experiment.

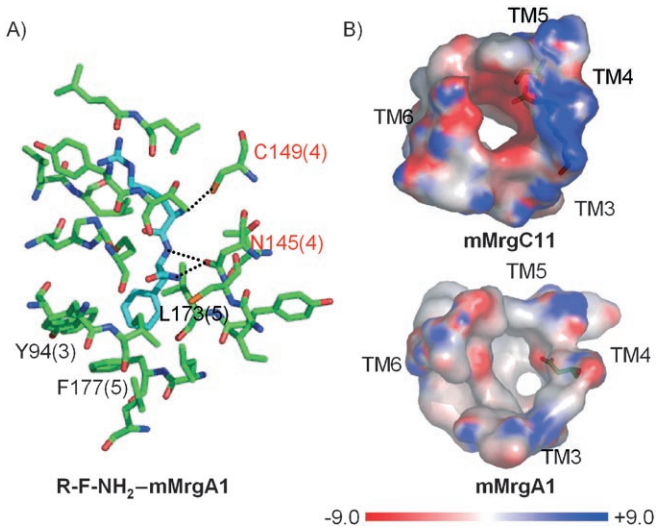

Figure 10. The predicted structure of the mMrgA1 receptor. A) The 5 Å binding pocket of RFa; the TM regions are indicated in brackets. B) The electrostatic-potential map of the binding pocket in mMrgC11 and mMrgA1. The residues within 5 Å from RFa ligand were selected for visualization. Asp161, Asp179, and Lys99 of mMrgC11 and Asn145 of mMrgA1 are specified in stick formats. The electrostatic potential was computed by using APBS and visualized with PyMOL. The van der Waals radii of DREIDING forcefield were used for APBS calculation.

Figure 10B shows the electrostatic potential maps of the binding pocket in mMrgC11 and of the corresponding region in mMrgA1. This electrostatic potential was calculated for the entire receptor by using adaptive Poisson–Boltzmann solver (APBS).[22] The electrostatic potential maps clearly show the difference in the binding pockets between mMrgC11 and mMrgA1. The pocket of mMrgA1 was much more hydrophobic than that of mMrgC11. Thus, it is plausible that highly polar peptides that contain an Arg residue would be unfavorable in the hydrophobic pocket in mMrgA1. This explains why the RFa ligand failed to bind strongly to the mMrgA1 receptor, as observed experimentally. These results provide additional validation of our prediction methods. We expect that the potency for these ligands to mMrgA1 would increase if either Asn145 or Asn166 residues is substituted to Asp.

To determine new ligands expected to bind strongly to mMrgA1, we started with the docked RFa described above and computed the binding for all 400 possible dipeptides matched to this site, and then optimized by energy minimization (following reassignment of the side chains by using SCREAM). We found dramatically improved binding affinity for cases in which Arg was replaced with such aromatic or less polar residues as Trp, Tyr, Phe, Ile, Leu, and Val, and in which the N terminus was replaced with aromatic residues (Trp, Tyr, Phe). In addition, replacement with Arg with the polar residue Gln led to hydrogen bonding with Thr88 (TM3), which enhanced the affinity. The top three candidates were: W-W-NH_2, W-F-NH_2, and Q-W-NH_2.

Conclusions

We have predicted the 3D structure of the mMrgC11 receptor and used it to predict the binding sites for a number of di- and tetrapeptide ligands. We predicted that in each case the peptide ligand binds in a pocket defined by TM3, 4, 5, and 6 and oriented parallel to the helical axis. These predictions suggested that Tyr110, Asp161, and Asp179 play key roles in ligand binding.

To test these predictions, we then carried out several mutagenesis experiments. For the six ligands that exhibited EC_{50} of 100–400 nM with wild-type receptor, we found that the EC_{50} for the Tyr110Ala, Asp161Ala, and Asp179Ala mutant receptors were higher than 33 μM for 14 of 18 combinations and 50–100-times higher for the other four combinations. This result validated the implication of these residues for the activation or possibly binding of the ligand, and provided very strong validation of the predicted 3D structure of MrgC11 and of the predicted binding site.

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

The stability of these ligand–GPCR structures were studied with MD simulations at 300 K for the full ligand–protein complex embedded in a periodic infinite lipid bilayer membrane fully solvated with water. For mMrgC11–F-(D)M-R-F-NH_2, we found that all key residues were preserved throughout the entire MD trajectory while the tetrapeptide remained solidly in the same site. These results provide additional validation of our structure-prediction methods.

This study indicates that computational methods are now adequate to predict useful 3D structures for GPCRs, which can subsequently be validated with a modest number of experiments. This structural information about the binding sites should provide insight into the functional characterization of these Mrg receptors and help to determine how they function. This should help 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, and indicate that a systematic combination of computational tools along with biochemical experiments can provide the structural data to provide an increased understanding of membrane protein receptors and their activation.

Computational Methods

All calculations were carried out by using DREIDING force field (FF)[23] with the charges from CHARMM22.[24] All energy and force calculations were carried out by using MPSIM.[25] The cell multipole method[26] 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.

Structure predictions of the Mrg receptor: The 3D structure of the mMrgC11 and mMrgA1 receptors were predicted independently by using MembStruk (version 4.05). The details of MembStruk (version 3.5) were described in ref. [10]. Here we outline the procedure, and highlight the aspects that were either relevant to Mrg receptors or improved in version 4.05.

Prediction of transmembrane regions: The TM regions were predicted by using the TM2ndS method.[10] We used NCBI BLAST[27] to search the nonredundant protein database for sequences homologous to the mMrgC11 receptor with bit scores of >200. Twenty seven sequence hits had sequence identities to mMrgC11 that ranged from 41 to 88%. This set of sequences (see Table S1 in the Supporting Information) included the mMrgA1 receptor, the sequence identity of which to mMrgC11 was 44%. Twenty two of these 27 sequences belonged to the Mrg receptor family, and the remaining five corresponded to unnamed GPCRs. We then carried out a multiple-sequence alignment with these 27 sequences using ClustalW.[28] These results (Figure S1 in the Supporting Information) were used as input to TM2ndS. The hydrophobicity profile (Figure S2) resulting from TM2ndS had no deep valley with negative values between TM2 and TM3; this led 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 calculated the hydrophobicities from a second round of seven TM predictions in each of which the sequence of each TM core (15 amino acids around the hydrophobic center) was used as template. This second set of independent BLAST searches was executed under high-gap penalty within 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 by using the multiple-sequence alignment of these seven sets of sequences, (Table S1). The final refined TM region and its hydrophobic center for each of the seven TM domains was 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 a fully extended conformation of side chains. These were assembled such that the seven predicted hydrophobic centers were all in the *xy* plane with the (*x*,*y*) coordinates adapted from the 7.5 Å electron-density map of frog rhodopsin.[29] 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.[29]

Then we carried out 200 ps MD simulations at 300 K without solvent or lipid. This allowed 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. Then, by using this conformation the net hydrophobic moment was calculated for the middle 15 residues around the hydrophobic center for each helix. Each helix was rotated again so that its hydrophobic moment faced toward the membrane. This hydrophobicity-based rotation worked well for the six TM helices with extensive contacts to lipid bilayers. This process placed the highly conserved Asn44 (TM1)–Asp71 (TM2) pair and Asn66 (TM2)–Trp151 (TM4) pair close enough to form hydrogen bonds. Thus, we concluded that the orientations of TM1, 2, and 4 were fine. Then we carried out combinatorial rotations of TM3, 5, and 6 with energy minimization to select the rotational orientation with the best energy. The results of these scans are shown in Table S2. The rotational orientation of TM7 was scanned over 360° in 5° increments; for each angle all atoms were optimized. In fact, the initial orientation showed the best energy. For every rotation we reassigned the side-chain conformation using SCWRL 3.0[30] before energy minimization.

Rigid-body dynamics in lipid bilayers and addition of loops: Next we added two layers of explicit lipid molecules (52 molecules of dilauroylphosphatidyl choline lipid) that surrounded the TM bundle. The initial structures for these studies 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 by using rigid body MD for 50 ps, in which each helix and lipid molecule was treated as a rigid body with just six 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 reassigned by using SCWRL, and the bundle (helices plus lipid) was minimized to a root mean square (RMS) force of 0.5 kcal mol^{-1} Å^{-1} by using conjugate gradients.

The loops were added to the helices by using MODELLER6v2.[31] The side chains were reassigned by using SCWRL, and subsequently a full atom conjugate gradient minimization of the receptor was performed.

In many GPCRs (including bovine rhodopsin and catecholamine 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, which leads 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 by using annealing MD while keeping the coordinates of TM helices fixed. In this process, the system was heated from 50 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 by using the conjugate gradients. This final structure (Figure 2, top and side views) was used for all docking studies.

Docking predictions with peptide ligands: We used a refined version (MSC-Dock) of the docking procedure described by Cho et al.[32] Since peptide ligands are highly flexible we modified the step in HierDock 2.0 (described by Vaidehi et al.[9]) to include the scan of the entire receptor with RFa to locate the binding site.

Scanning of the binding sites: In order to locate the putative binding site the entire receptor was scanned with the R-F-NH_2 (RFa) dipeptide known to agonize the receptor (EC_{50} = 460 nM). First, the molecular surface was created by using the autoMS utility in DOCK 4.0[33] with the default values for surface density (= 3.0 dots per $Å^2$) and probe radius (= 1.4 Å). Then, we generated spheres from each that filled the void space in the receptor. To do this we used SPHGEN in DOCK 4.0. We then constructed a total of 40 cubic boxes (sides of 10 Å) and spaced by 8 Å that covered this set of spheres.

Assuming that the ligand binds inside the bundle from the extracellular region, we analyzed these spheres by using a buried surface criterion to preselect four nonperipheral regions, located on the upper half of the receptor (Figure S3). The spheres inside each box were used to define the docking region as input to DOCK 4.0.

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 ten torsions), we wanted to ensure that this extensive conformation space was sampled in the docking. Thus for each peptide ligand we used the Metropolis Monte Carlo (MC) method in Cerius2[34] (with a MC temperature of 5000 K in vacuum) to generate a set of 1000 conformers that had a diversity of CRMS = 1.0 Å after being ranked according to energy values.

Level 0: Then for each of the 1000 conformers we used DOCK 4.0 to generate a set of 3000 configurations within each of the four binding regions of the receptor. From these we selected the 100 best configurations for each of the 1000 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 option was used. The bump filter option was turned on (maximum bump = 10) and the reduced (to 75%) van der Waals radius was used.

Level 1: The configurations from level 0 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 DOCK energy. This ordered list was trimmed by using a diversity criterion of CRMS = 0.6 Å and the top 100 configurations were selected. Each of these was conjugate-gradient minimized in MPSIM with 100 steps while the receptor coordinates were fixed.

Level 2: The ten best configurations according to energy were selected from level 1 and the whole ligand–protein complex was minimized in MPSIM with 100 steps by using the all-atom DREIDING energy function to evaluate the energy for the ligand–receptor complex. The side chains for all residues of the receptor within 5 Å of the ligand were reassigned by using the SCREAM side-chain replacement program (Kam, Vaidehi, and Goddard, unpublished results), which uses a side-chain rotamer library of 1478 rotamers with a diversity of 1.0 Å.

Level 3: The binding energies were then calculated for these ten optimized ligand–receptor complex structures. Here, the energy of the free ligand was calculated by using its docked conformation (snap-bind energy) and its solvation energy was evaluated by using the Analytical Volume Generalized Born (AVGB) continuum solvation method.[35] In these calculations, the dielectric constant was set to 78.2 for the exterior region and to 1.3 for the interior region. The energy for the ligand–receptor complex was calculated in the gas phase with the receptor fixed, and only the ligand–receptor interaction energy and the internal energy of the ligand were included. The final binding energy was obtained by subtracting the energy of free ligand (solvation energy included) from the energy of the ligand–receptor complex. The final best structure was selected as the one with the highest binding energy.

Among four regions we found that the RFa ligand had the highest binding energy in the region that involved TM3, 4, 5, and 6 (blue in Figure S3 B)—the putative binding site. The final structure of the RFa–receptor complex was further refined by using annealing MD as described previously. Here, only the ligand and the residues within 10 Å of the binding pocket (including backbone atoms) were allowed to move. At the end of the annealing cycle, the system was minimized to an RMS force of 0.3 kcal mol^{-1} $Å^{-1}$ and the side chains of the residues in the receptor within 4 Å from the ligand was reassigned again with SCREAM. Then the side chain of peptide ligand was also reassigned with SCREAM. The spheres were redefined with this final optimized RFa–receptor complex and were used for docking of other peptide ligands.

For the five F-M-R-F tetrapeptide stereoisomers, we first docked the R-F amide part of the C terminus. This motif is common to most peptide agonists of mMrgC11.[4] Indeed the efficacy results for the five chiral modified F-M-R-F-amides (Table 1) showed that the chirality of the R-F part dramatically affected 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 to build the remaining F-M amino acids for constructing the docked tetrapeptide.

For docking the three dipeptides, we used the Dock-Diversity Completeness protocol (DDCP) described by Cho et al.[32] to improve completeness in searching the configurations in DOCK (level 0). Briefly, DDCP attempts to generate a complete set of ligand configurations families with a fixed diversity (1.0 Å). Completeness is defined as the point at which the fraction of new configuration that belongs 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 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^{-1} Å of RMS force) with the ligand atoms movable and the receptor atoms fixed. Then, the ten 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 by using surface generalized Born model (SGB).[36] The side-chain conformations 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 ten complex structures were minimized (100 steps or 0.1 kcal mol^{-1} Å of RMS force) with all atoms movable.

The above docking procedure was applied to each of the 1000 conformers of each peptide ligand generated with the MC method in Cerius2 (with a MC temperature of 5000 K in vacuum) by using

a diversity of CRMS = 2.0 Å. Prior to docking the structures of the 1000 conformers were minimized in gas phase and ordered according to energy. Then, they were reclustered 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 to at least ten family heads for the R-F dipeptides and over 20 family heads for acetylated R-F dipeptides.

For each such structure, the docking process ended up with ten 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 HierDock was further refined by annealing MD as previously described. 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^{-1} Å^{-1}.

Building the terminal F-M residues from the bound acetylated R-F-NH_2: The conformations of the terminal F-M residues were sampled by using moleculeGL—a recursive, Metropolis Monte Carlo-based rotamer design technique (Kekenes-Huskey, Vaidehi, and Goddard, unpublished results) from the R-F-NH_2 dipeptide docked in the mMrgC11 receptor with the extracellular loops removed. Either the psi angle of Met or the phi angle of Arg was defined as 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^{-1} Å^{-1} RMS force) with the coordinates of receptor fixed. Then the side chains of residues within 5 Å of the ligand were assigned by using SCREAM and the structure of the whole complex was minimized. The final best structure was refined by annealing as described in the previous section.

Molecular dynamics simulation for ligand–protein in a solvated full membrane system: After predicting the structure of the mMrgC11–F-(D)M-R-F-NH_2 complex, we embedded this structure in a periodic infinite membrane and solvated the system with explicit water to form a system in which we could equilibrate with MD at 300 K for 7 ns. The procedure was similar to that described in ref. [21] for β2 adrenergic receptor.

The periodic box started at 40×40×100 Å^3 with 47651 atoms, which included 4180 receptor atoms, 74 ligand atoms, 4288 lipid atoms, 39087 water atoms, and ten chlorine atoms. We used palmitoyl-oleoyl-phosphatidylcholine (POPC) to form the lipid bilayers. Prior to full MD, the system was subjected to 5000 steps of conjugate gradient energy minimization, followed by 100 ps of equilibration, while the coordinates of the receptor–ligand complex were fixed. Then the system was gradually heated from 0 to 310 K by using Langevin dynamics with a damping coefficient of 5 ps^{-1}, and the target temperature was reached after ~7 ps. The system was again subjected to 5000 steps of conjugate gradient energy minimization without any restraints. During energy minimization, the nonbonded and electrostatic interactions were computed for every time step. Finally, the full MD simulation was carried out for 7 ns while a constant pressure of 1 atm was maintained by using the Langevin piston method. The final structure after a 7 ns equilibration was minimized with conjugate gradient for 5000 steps.

Experimental Section

In vitro mutagenesis: Each point mutation was incorporated into the mMrgC11–GFP coding sequence in pcDNA3.1/Zeo(+) plasmid (Invitrogen) by using the QuickChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). The mutagenic oligonucleotide primers were synthesized and purified at the oligonucleotide synthesis center at Caltech. All mutant constructs were verified by DNA sequencing. Later the wild-type and mutant genes in pcDNA3.1/Zeo(+) were subcloned into the pcDNA5/FRT expression vector (Invitrogen) for the generation of stably expressing cell lines.

Cell culture and transfection: Flp-In™-293 cells (Invitrogen) were cotransfected with the mMrgC11–GFP gene in pcDNA5/FRT vector and pOG44 plasmid (Invitrogen) by 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 fetal bovine serum (FBS; 10%), penicillin/streptomycin (1%) and L-glutamine (1%). The cells were passaged into fresh medium 48 h after transfection and then selected with hygromycin (400 μg mL^{-1}). After a two-week period of selection, hygromycin-resistant clones were picked and maintained in selective medium with hygromycin (200 μg mL^{-1}).

Biotinylation and immunoprecipitation: Flp-In™-293 cells that stably expressed wild-type and mutant receptors were placed into a culture dish (10 cm) coated with poly-L-lysine, and incubated for 24 h, at 37 °C in a humidified atmosphere with 5% CO_2. The cells were washed twice with ice-cold phosphate buffered saline (PBS) and incubated with Sulfo-NHS-LC-biotin (3 mL, 0.5 mg mL^{-1}; Pierce Biotechnology, Inc.) in PBS supplemented with HEPES (0.1 mM, pH 7.5) at room temperature for 30 min. The biotinylation reaction was quenched by washing the cells three times with Tris-buffered saline (10 mM Tris, pH 7.5, 154 mM NaCl). The washed cells were incubated with cold lysis buffer (5 mL; 10 mM HEPES, pH 7.4, 1 mM EGTA) supplemented with 4-(2-aminoethyl)-benzene sulfonyl fluoride hydrochloride (100 μM) 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 750 *g* for 10 min at 4 °C to remove the nuclei and cell debris. The resulting supernatant was centrifuged at 75000 *g* for 30 min at 4 °C. The membrane pellet was solubilized in ice-cold TX/G buffer (500 μL; 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 1 h. Insoluble material was removed by centrifugation at 10000 *g* for 15 min at 4 °C. The protein concentration was estimated by using the DC protein assay kit (Bio-Rad).

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

Intracellular calcium-release assay: The cells were placed into a 96-well cell culture plate coated with MATRIGEL matrix (BD Biosci-

ences). After 16–24 h incubation, the cells were washed twice with assay buffer (Hank's balanced salt solution supplemented with 10 mM D-glucose, 20 mM HEPES, and 1.6 mM NaOH) and loaded with fura-2/AM (2 μM; Molecular Probes) in assay buffer at room temperature for 20 min. Then the cells were washed four times with assay buffer to eliminate residual fura-2/AM outside the cell membranes. The fluorometric imaging plate reader (FLIPR) assay was carried out at various concentrations of peptide ligands (1 nM–10 μM) with the FlexStation II system (Molecular Devices). The fluorescence emitted from the excitation at 340 nm and 380 nm was measured as a function of time, and the ratio of emission at these two excitation wavelengths was evaluated. 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} values.

Acknowledgements

We thank Professor Melvin I. Simon for suggesting this project, for helpful discussions, and for use of his experimental facilities (funded by a grant from the National Institutes of Health (NIH-NS048499)). J.H. thanks J.-I. Hwang and K.-J. Shin for helpful comments on the experiments. The computational research was supported partially by funding from the NIH (MH073910-01, CA 112293-01) and the computational facilities were provided by DURIP grants from ARO and ONR.

Keywords: FMRFamide neuropeptide · G protein-coupled receptors · molecular modeling · Mrg receptor · mutagenesis

[1] A. Wise, S. C. Jupe, S. Rees, *Annu. Rev. Pharmacol. Toxicol.* **2004**, *44*, 43–66.
[2] X. Z. Dong, S. K. Han, M. J. Zylka, M. I. Simon, D. J. Anderson, *Cell* **2001**, *106*, 619–632.
[3] P. M. C. Lembo, E. Grazzini, T. Groblewski, D. O'Donnell, M. O. Roy, J. Zhang, C. Hoffert, J. Cao, R. Schmidt, M. Pelletier, M. Labarre, M. Gosselin, Y. Fortin, D. Banville, S. H. Shen, P. Strom, K. Payza, A. Dray, P. Walker, S. Ahmad, *Nat. Neurosci.* **2002**, *5*, 201–209.
[4] S. K. Han, X. Z. Dong, J. I. Hwang, M. J. Zylka, D. J. Anderson, M. I. Simon, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14740–14745.
[5] E. Grazzini, C. Puma, M. O. Roy, X. H. Yu, D. O'Donnell, R. Schmidt, S. Dautrey, J. Ducharme, M. Perkins, R. Panetta, J. M. A. Laird, S. Ahmad, P. M. C. Lembo, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7175–7180.
[6] N. Robas, E. Mead, M. Fidock, *J. Biol. Chem.* **2003**, *278*, 44400–44404.
[7] E. Bender, A. Buist, M. Jurzak, X. Langlois, G. Baggerman, P. Verhasselt, M. Ercken, H. Q. Guo, C. Wintmolders, I. Van den Wyngaert, I. Van Oers, L. Schoofs, W. Luyten, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8573–8578.
[8] T. Shinohara, M. Harada, K. Ogi, M. Maruyama, R. Fujii, H. Tanaka, S. Fukusumi, H. Komatsu, M. Hosoya, Y. Noguchi, T. Watanabe, T. Moriya, Y. Itoh, S. Hinuma, *J. Biol. Chem.* **2004**, *279*, 23559–23564.
[9] N. Vaidehi, W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, G. Zamanakos, W. A. Goddard, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12622–12627.
[10] R. J. Trabanino, S. E. Hall, N. Vaidehi, W. B. Floriano, V. W. T. Kam, W. A. Goddard 3rd, *Biophys. J.* **2004**, *86*, 1904–1921.
[11] W. B. Floriano, N. Vaidehi, G. Zamanakos, W. A. Goddard, *J. Med. Chem.* **2004**, *47*, 56–71.
[12] M. Y. S. Kalani, N. Vaidehi, S. E. Hall, R. J. Trabanino, P. L. Freddolino, M. A. Kalani, W. B. Floriano, V. W. T. Kam, W. A. Goddard, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3815–3820.
[13] P. L. Freddolino, M. Y. S. Kalani, N. Vaidehi, W. B. Floriano, S. E. Hall, R. J. Trabanino, V. W. T. Kam, W. A. Goddard, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2736–2741.
[14] T. Okada, M. Sugihara, A. N. Bondar, M. Elstner, P. Entel, V. Buss, *J. Mol. Biol.* **2004**, *342*, 571–583.
[15] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, *289*, 739–745.
[16] S. Miura, S. S. Karnik, *J. Biol. Chem.* **2002**, *277*, 24299–24305.
[17] J. S. Mills, H. M. Miettinen, D. Cummings, A. J. Jesaitis, *J. Biol. Chem.* **2000**, *275*, 39012–39017.
[18] D. Pal, P. Chakrabarti, *J. Biomol. Struct. Dyn.* **2001**, *19*, 115–128.
[19] R. E. Rosenfield, R. Parthasarathy, J. D. Dunitz, *J. Am. Chem. Soc.* **1977**, *99*, 4860–4862.
[20] D. Meyer zu Heringdorf, N. Niederdräing, E. Neumann, R. Fröde, H. Lass, C. J. Van Koppen, K. H. Jakobs, *Eur. J. Pharmacol.* **1998**, *354*, 113–122.
[21] P. Spijker, N. Vaidehi, P. L. Freddolino, P. A. J. Hilbers, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4882–4887.
[22] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041.
[23] S. L. Mayo, B. D. Olafson, W. A. Goddard, *J. Phys. Chem.* **1990**, *94*, 8897–8909.
[24] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
[25] K. T. Lim, S. Brunett, M. Iotov, R. B. McClurg, N. Vaidehi, S. Dasgupta, S. Taylor, W. A. Goddard, *J. Comput. Chem.* **1997**, *18*, 501–521.
[26] H. Q. Ding, N. Karasawa, W. A. Goddard III, *J. Chem. Phys.* **1992**, *97*, 4309–4315.
[27] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
[28] J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
[29] V. M. Unger, P. A. Hargrave, J. M. Baldwin, G. F. X. Schertler, *Nature* **1997**, *389*, 203–206.
[30] A. A. Canutescu, A. A. Shelenkov, R. L. Dunbrack, *Protein Sci.* **2003**, *12*, 2001–2014.
[31] MODELLER6v2, University of California San Francisco, **2002**.
[32] A. E. Cho, J. A. Wendel, N. Vaidehi, P. M. Kekenes-Huskey, W. B. Floriano, P. K. Maiti, W. A. Goddard III, *J. Comput. Chem.* **2005**, *26*, 48–71.
[33] T. J. A. Ewing, I. D. Kuntz, *J. Comput. Chem.* **1997**, *18*, 1175–1189.
[34] Cerius2 Modeling Environment, Release 4.0, Accelrys Inc., San Diego, **1999**.
[35] G. Zamanakos, Ph.D. thesis, California Institute of Technology (Pasadena, USA), **2002**.
[36] A. Ghosh, C. S. Rapp, R. A. Friesner, *J. Phys. Chem. B* **1998**, *102*, 10983–10990.
[37] I. K. McDonald, J. M. Thornton, *J. Mol. Biol.* **1994**, *238*, 777–793.
[38] For the N terminus and the carboxylate oxygen of Asp161 in RFa, distance(D–A) = 2.92 Å, angle(D–H–A) = 162.6°, angle(AA–A–D) = 89.7° and angle(AA–A–H) = 92.4°, where D is the donor heavy atom, H the hydrogen, A the acceptor, and AA the acceptor antecedent.
[39] For the N of Arg side chain and carboxylate oxygen of Asp161, distance(D–A) = 2.87 Å, angle(D–H–A) = 159.7°, angle(AA–A–D) = 94.5°, and angle(AA–A–H) = 89.8°.

Received: April 12, 2007
Published online on July 23, 2007

