Chapter 5

Prediction of the 3D Structure of Rat MrgA G Protein-Coupled Receptor and Identification of its Binding Site[1]

5.1 Introduction

Rat MrgA is one of a few Mrg receptors for which the small molecular (non-peptide) agonists have been identified. It has been shown to be activated by adenine (and not guanine). Indeed adenine activates rMrgA with a K_i value of 18 nM, potentially identifying it as the endogenous ligand[1]. In this chapter we predict the 3D structure of the rMrgA receptor, and we report the ligand binding site for adenine and related ligands. This work builds upon our previous studies in which we first predicted the 3D structures of mouse MrgC11 (mMrgC11) and MrgA1 (mMrgA1) receptors using the MembStruk computational method[2, 3]. These structures were validated by predicting the binding sites and energies for several tetrapeptides, identifying key residues, and then experimentally confirming the expected changes in binding resulting from mutations of these residues, as described in chapter 2.

For this study on rMrgA, we use these validated mMrgC11 and mMrgA1 structures as templates to predict through homology modeling the 3D structure of rMrgA receptor (it is 49 % and 77 % sequence identical to the mMrgC11 and mMrgA1 sequences). Then we used this structure of rMrgA in conjunction with the HierDock computational procedure to predict the binding site of all nine ligands to the rMrgA receptor for which experimental data are available.

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

```
                TM1                                 TM2
rMrgA   RTLIPNLLIIISGLVGLTGNAMVFWLLGFRLARNAFSVYILNLALADFLFLLCHIIDSTL 60
mMrgA1  TILIPNLMIIIFGLVGLTGNGIVFWLLGFCLHRNAFSVYILNLALADFFFLLGHIIDSIL 60
mMrgC11 PILTLSFLVLITTLVGLAGNTIVLWLLGFRMRRKAISVYILNLALADSFFLCCHFIDSLL 60
          * .::::*   ****:** :*:***** : *:*:*********** :**  *:*** *
                              TM3                                 
rMrgA   LLLKF--SYPNIIFLPCFNTVMMVPYIAGLSMLSAISTERCLSVVCPIWYRCRRPKHTST 118
mMrgA1  LLLNV--FYP-ITFLLCFYTIMMVLYIAGLSMLSAISTERCLSVLCPIWYHCHRPEHTST 117
mMrgC11 RIIDFYGLYAHKLSKDILGNAAIIPYISGLSILSAISTERCLCVLWPIWYHCHRPRNMSA 120
         ::..   *.       : .  :: **:***:**********.*: ****:*:**.: *:
               TM4                 TM5                 
rMrgA   VMCSAIWVLSLLICILNRYFCGFLDTKYEKDNRCLASNFFTAACLIFLFVVLCLSSLALL 178
mMrgA1  VMCAVIWVLSLLICILNSYFCGFLNTQYKNENGCLALNFFTAAYLMFLFVVLCLSSLALV 177
mMrgC11 IICALIWVLSFLMGILDWF-SGFLGETHH--HLWKNVDFIITAFLIFLFMLLSGSSLALL 177
        ::*: *****:*: **: : .***.  :.  :     :*: :* *:***::*. *****:
                              TM6                                 
rMrgA   VRLFCGAGRMKLTRLYATIMLTVLVFLLCGLPFGIHWFLLIWIKIDYGKFAYGLYLAALV 238
mMrgA1  ARLFCGTGQIKLTRLYVTIILSILVFLLCGLPFGIHWFLLFKIKDDFHVFDLGFYLASVV 237
mMrgC11 LRILCGPRRKPLSRLYVTIALTVMVYLICGLPLGLYLFLLYWFGVHLHYPFCHIYQVTAV 237
         *::**. :  *:***.** *:::*:*:****:*:: ***  :  .       :* .: *
          TM7
rMrgA   LTAVNSCANPIIYFFVG 255
mMrgA1  LTAINSCANPIIYFFVG 254
mMrgC11 LSCVNSSANPIIYFLVG 254
        *:.:**.*******:**
```

Figure 5.1 Sequence alignment provided as an input for the homology modeling of rMrgA. The N-terminus (11 residues) and C-terminus (38 residues) were omitted because for such class A (rhodopsin-like) GPCRs especially for small ligands, they generally do not play a role in the binding of the ligand[4].

We also compare the putative binding site of rMrgA receptor with those of other known adenine-related GPCRs like adenosine receptors or purinergic receptors.

5.2 Materials and methods

5.2.1 Molecular modeling of receptor structure

We used MODELLER6v2[5] to build a homology model for the 3D structure of rMrgA receptor using the 3D structures for mMrgC11 and mMrgA1 as templates. The sequences of rMrgA receptor (TrEMBL accession number: Q7TN49) was aligned with mMrgC11 (TrEMBL accession number: Q8CIP3) and mMrgA1 (TrEMBL accession number: Q91WW5) using Clustal-W (version 1.82)[6] as shown in Figure 5.1. The sequence identity of rMrgA with mMrgC11 is 49%, while that for mMrgA1 is 77%, for the entire sequences. The TM regions have

44% to 76% identity (totaling 56%) between rMrgA and mMrgC11 and 77% to 88% identity between rMrgA and mMrgA1 (totaling 83%).

After predicting the overall 3D structure of rMrgA, the side chain conformations were re-assigned using the SCWRL3.0 side chain replacement program (~1.4 Å diversity)[7] and hydrogen atoms were added using the POLYGRAF software. The all-atom structure was optimized with the conjugate gradient minimization technique to an RMS in force of 0.5 kcal/mol/Å. Subsequently this minimized receptor structure was used as the starting point for gas phase NVT molecular dynamics (MD) simulations (using an internal dielectric constant of 2.5) at 300 K for 10 ps to account for changes in the backbone conformation. The conformation with the lowest total energy in the trajectory was selected and minimized to an RMS force of 0.5 (kcal/mol)/Å with conjugate gradients. All simulations used the DREIDING force field (FF)[8] with charges from CHARMM22[9] in the MPSim code[10]. The cell multipole method[11] was used for calculation of non bond interaction.

5.2.2 QM calculation of ligand tautomers

We docked to rMrgA the 9 molecules shown in Figure 5.2 (including adenosine phosphates), for all of which there are measured binding constants. The structures for these molecules were constructed using the Cerius2 build module[12]. The ligand conformations were minimized using conjugate gradients with the DREIDING FF and GASTEIGER charges[13]. For ligands with a significant number of torsions, such as 6-benzylaminopurine (6BAP), adenosine and adenosine phosphates, the X-ray crystal structures were obtained from the cambridge structural database and used as the starting conformation for docking without further optimization.

For 1-methyladenine (1MA) and 6BAP, several tautomeric forms are possible in addition to the direct substitution at N1 or N6 of adenine. For these systems we built all such tautomeric forms (see Figure 5.2) and calculated their relative stabilities using quantum mechanics (QM) (Jaguar v5.5 software[14]) to determine the dominant tautomeric form. The geometries were first

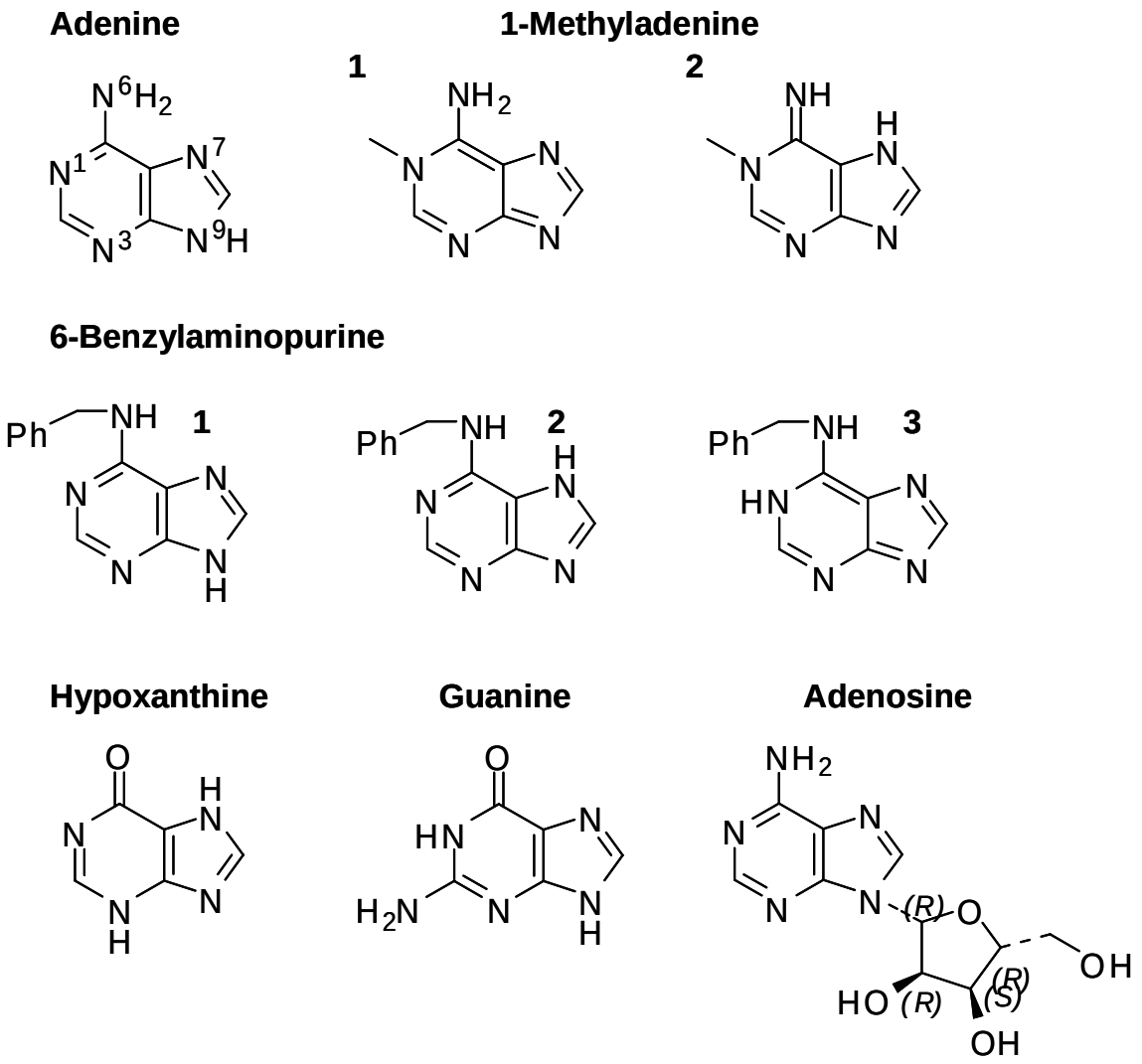

Figure 5.2 Ligand compounds used in docking studies for the rMrgA receptor. They are placed in order of experimental binding affinity from top-left to bottom-right. No binding was detected experimentally for the ligands of the third row. For 1-methyladenine (1MA) and 6-benzylaminopurine (6BAP), the most stable tautomeric forms are shown together.

optimized in the gas phase using the B3LYP flavor of Density Functional Theory with the 6-31G** basis set. The vibrational frequencies for thermodynamic quantities were calculated at the same level. The calculated frequencies were scaled by the factor 0.9614 appropriate for B3LYP/6-31G*. All thermodynamic quantities were computed at 298.15 K, based on standard ideal-gas statistical mechanics and the rigid-rotor harmonic oscillator approximations. We calculated the solvation energy in water using the Jaguar Poisson-Boltzmann methodology with

standard parameters (dielectric constant ε_{H2O} = 80.37, solvent probe radius R_{H2O} = 1.40 Å, and Dreiding van der Waals radii of atoms) for the final optimized QM structure. These results are in Table S5.1 of the supplementary information.

5.2.3 Prediction of the adenine binding site

Scanning the receptor to determine the putative binding region

To select the putative binding region, we used adenine (the best binder) to scan the entire receptor structure of rMrgA. To do this we first calculated the molecular surface using autoMS utility in DOCK4.0[15] with the default values for surface density (3.0 dots/Å^2) and probe radius (1.4 Å). Then we used SPHGEN in DOCK4.0 to generate spheres from each surface point to fill up the void space in the receptor. The receptor was partitioned into 41 cubic boxes each with sides of 10 Å such that all void spheres were included. The spheres inside each box were taken as an input for DOCK4.0 to define the docking region. The scoring energy grids of the protein were calculated using GRID in DOCK4.0, with a grid spacing of 0.3 Å and a nonbond cutoff distance of 10 Å. For each of the 41 regions, we performed rigid docking with the anchor search option in DOCK4.0. For each region, we sampled orientations until 100 passed the bump test and then we selected the ten top scoring orientations. For each of these 10 from each of the 41 boxes, we used MPSim to minimize the ligand conformation with the receptor coordinates fixed to obtain the final energy scores. Here we used the Dreiding FF. After scoring with MPSim, we calculated the percentage of buried surface for each of these 410 orientations using the Connolly MS program from Quantum Chemistry Program Exchange (QCPE). Of these, 103 had over 90 % of buried surface. From these we selected the best orientation for each box. Out of the 41 boxes, this led to seven possible binding regions with good energy and >90% buried surface. We then clustered the spheres near these seven regions, to obtain the two distinct putative binding sites shown in Figure 5.3.

Docking adenine and guanine into the predicted putative binding sites

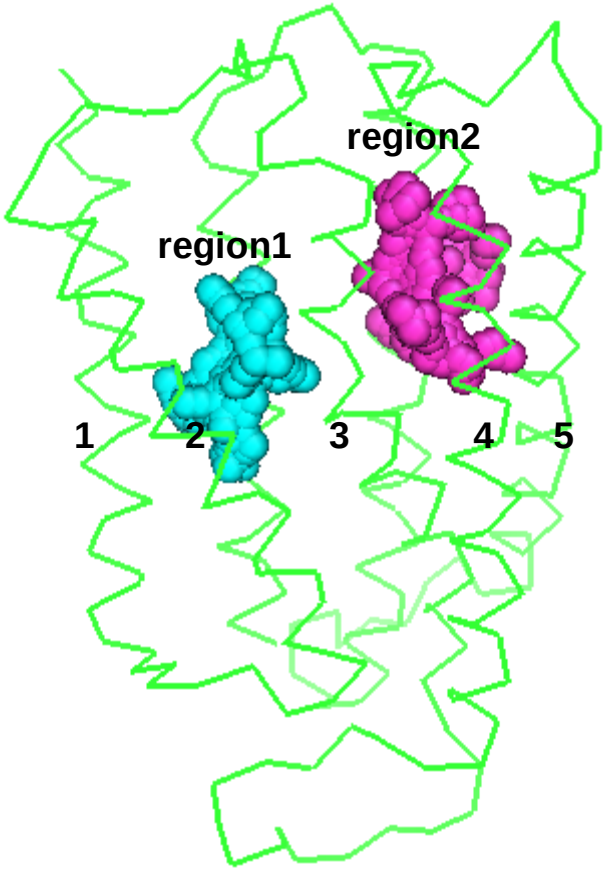

Figure 5.3 Putative binding sites predicted from the HierDock scanning procedure. Region 2 is in the TM3456 region that we find to bind adenine-like agonists. Region 1 is in the TM1237 region (it does not play a role in binding agonists, but might for antagonists).

The HierDock protocol was used to predict the binding site and energy of adenine to both binding regions. In the study on rMrgA we also used the modified HierDock protocol (MSC-Dock) described in chapter 2. Here we used a rejection ratio of 2.2 to define completeness (leading to 2,453 families that past the bump tests). We then enriched the top 75 families until there was an average of six members in each family (passing the bump tests). Then we scored these using MPSim (Dreiding FF) and selected the 30 best scoring family heads. These were minimized (conjugate gradients) using MPSim (50 steps or 0.1 kcal/mol/Å) with ligand movable and the receptor atoms fixed. Then the 5 best scoring ligands (total energy) were selected and the

side chain conformations of the residues of the receptor within 5 Å of the bound ligand were reassigned using the SCREAM side chain replacement program (This uses a side chain rotamer library of 1,478 rotamers with 1.0 Å resolution, with all atom DREIDING energy function to evaluate the energy for the ligand-receptor complex). The binding energies were then calculated for these 5 optimized ligand-receptor complex structures 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 analytical volume generalized Born (AVGB) continuum solvation method[16]. The dielectric constants for the continuum solvation method were set to 78.2 for the external region and to 1.3 for the internal region.

Guanine shows no binding in the experiments (worse than ~100 μM). We docked it to the two putative binding regions determined from scanning the receptor (shown in Fig. 5.3).

5.2.4 Refinement of the binding mode of adenine

To account for changes in the backbone structure of the receptor due to ligand binding, we started with the docked structure and carried out annealing MD simulations allowing the ligand and residues within 10 Å in the binding pocket to move (with other residues fixed). The procedure was to heat the system from 50 K to 600 K and then to cool it back down to 50 K in steps of 50 K. The system was equilibrated for 1ps between changes in temperature. 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 within 5 Å from the ligand was reassigned again with SCREAM.

5.2.5 Docking of other adenine derivatives

After optimizing the structure for adenine in the receptor, we re-clustered the spheres to define the binding site. Spheres within 1.0 Å from any atom in the docked adenine were selected out of the entire spheres generated for the final receptor structure that was previously optimized

with adenine. We then used the HierDock procedure described above to dock the adenine derivatives.

5.3 Results and discussion

5.3.1 Characteristics of receptor structure

The sequence identity of rMrgA receptor with bovine rhodopsin is ~18 % for TM regions (the averaged value obtained with the independent alignments for each TM). The RMSD of the coordinates of the Cα atoms between these two receptors is 3.72 Å in TM regions[17].

The RMSD of rMrgA with mMrgA1 (83 % sequence identity for TM regions) is 0.41 Å in the TM regions and the RMSD with mMrgC11 (56 % sequence identity for TM regions) is 2.59 Å in the TM regions. The predicted 3-D structure of rMrgA is shown in Figure 5.4(b) where it is superimposed with the predicted structures of mMrgA1 and mMrgC11.

Figure 5.5 shows the interhelical hydrogen bond network in TM regions formed in the rMrgA receptor;

The Asn31 (TM1) makes hydrogen bonds with the side chain of Asp58 (TM2) and the backbone carbonyl of Cys256 (TM7) at the same time and contributes to the interhelical stability among TM1, TM2 and TM7. This Asp-Asn pair is highly conserved across the family A of GPCRs, corresponding to Asp83 and Asn55 in bovine rhodopsin. There is a similar pattern in rhodopsin structure[18] where a carbonyl group of A299 in the backbone of TM7 is as the common hydrogen bond acceptor for Asn55.

The Tyr95 (TM3) is conserved throughout the Mrg receptor family (although 5 of 36 have a Phe conservative replacement at this position). Here the hydroxyl group of Tyr forms an interhelical hydrogen bond with a backbone carbonyl group of C218 in TM6.

The highly conserved Asn53 (TM2) and Trp136 (TM4) form a hydrogen bond as observed in rhodopsin (Asn78 and Trp161).

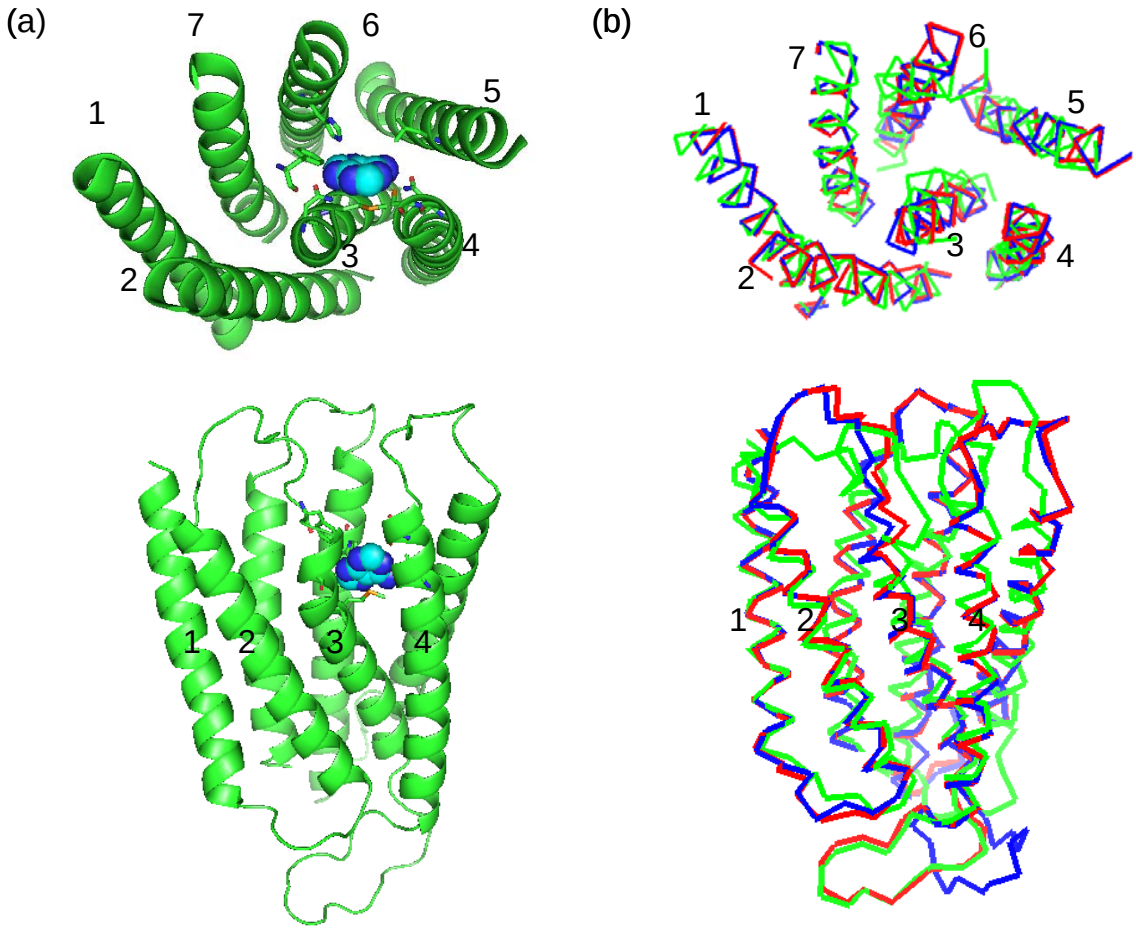

Figure 5.4 Predicted 3D structure of rMrgA receptor.

(a) Adenine (in spheres) is docked in rMrgA receptor. The residues within 5 Å of adenine are shown as sticks. (b) The rMrgA receptor (red) is overlapped with mMrgA1 (blue) and mMrgC11 (green). The top part shows the view from the extracellular side, while the bottom part shows the side view (with the extracellular part on top).

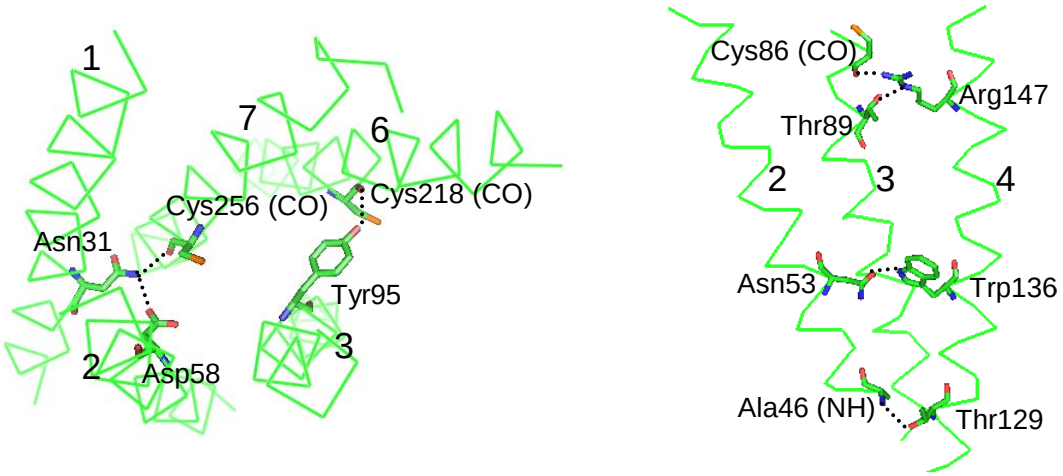

Figure 5.5 Interhelical hydrogen bonds (dashed lines) in rMrgA receptor, as identified using HBPLUS[19] (maximum D-A distance = 3.9 Å, minimum D-H-A angle = 90.0°).

One more hydrogen bond pair exists between Ala46 (TM2) and Thr129 (TM4) near the intracellular region.

In addition, the positively charged residue Arg147 (TM4) is oriented slightly towards the lipids and might contact with the negatively charged head group of the lipid molecule. We find that it forms the hydrogen bonds with Cys86 and Thr89 in TM3 that are one helical turn apart.

The highly conserved proline residues in TM6 and TM7 across the family A of GPCRs correspond to Pro221 (TM6) and Pro258 (TM7) in rMrgA receptor. They lead to bends of 15° and 18° in the α-helix structure.

The Pro94 (TM3) in rMrgA receptor corresponds to the double Gly in the middle of rhodopin. In both cases this leads to bending (19° for rMrgA and 13° for rhodopsin), making the overall backbone conformation of TM3 in these two receptors similar.

A major difference between rMrgA and most other family A GPCRs is that there is no Cys in the extracellular loop (EC) 2 or at the top of TM3. In rhodopsin and other amine receptors there are highly conserved cysteine residues in TM3 and in the EC2 that form a disulfide linkage

that constrains the structure of EC2. Thus for rMrgA receptor we find that EC2 has an open random coil conformation. (In rhodopsin this loop has a closed beta sheet structure).

5.3.2 QM results of ligand tautomers

The QM results of the free energies for the different tautomeric forms of 1MA and 6BAP are shown in Table S5.1. We find that in solution the free energy of 1MA1 is 1.87 kcal/mol lower. The relative abundance with respect to the tautomer with the lowest free energy was calculated from the free energy using the equation;

$$\frac{[tautomer]}{[tautomer]_{lowest}} = \exp\left(-\Delta G_{sol} \diagup RT\right),$$

where R is the gas constant (1.986 cal/mol·K) and T is the temperature (298.15 K). Thus we predict that the relative abundance of 1MA2 is only ~4 % of 1MA1. (In contrast 1MA1 is less stable than 1MA2 by 3.5 kcal/mol in the gas phase.)

There are three tautomers for 6BAP, but 6BAP1 is the most stable both in gas phase and in aqueous solution. Here the others forms have negligible abundance.

These calculations suggest that the majority species for 1MA or 6BAP have direct substitutions at the N1 or N6 of adenine. Therefore these forms were chosen for the docking studies.

5.3.3 Binding modes of adenine and other ligands

Location of the binding site

MSC-Dock predicts the adenine binding site lie between TM3, TM4, TM5 and TM6 as shown in Figure 5.4. This TM3-4-5-6 pocket (corresponding to region 2 in Fig. 5.3) is predicted to provide the binding site for the agonists to a number of other GPCRs (including dopamine, adrenergic, histamine). In addition the adenine is in a region similar to the β-ionone ring of 11-cis retinal in bovine rhodopsin (but the adenine leans more towards TM4 instead of TM6).

The scanning step also found a second binding site, denoted as region 1 in Figure 5.3. This other site is located in the interhelical hydrogen bond network between TM1, TM2 and TM7. In this site both adenine and guanine make a hydrogen bond with the highly conserved Asp58 in TM2, but the binding pocket is mostly hydrophobic except for this Asp residue. We found that the calculated binding energy of adenine in region 1 is only 66 % of that in region 2. The binding energy of guanine in region 1 was 73 % of that for adenine in region 2. Thus we conclude that this site is not the site for agonist binding (it could play a role for antagonists).

As discussed in section 3.1, Asp58 (TM2) plays a key role in stabilizing the TM1, 2, 7 triad, and it may be the site at which Na^+ binds for the allosteric regulation observed in human adenosine A1 receptor and α_{2A} adrenergic receptor[20, 21], making it unlikely to serve as the agonist binding site.

Based on these results we ruled out region 1 as a possible binding site.

Predicted Binding site of Adenine

Adenine is reported as the potential endogenous ligand for rMrgA receptor by Bender *et al.*[1]. The binding mode is detailed in Figure 5.6(a). The most critical residues for binding are Asn88 TM3 and Asn146 TM4. They each form bidentate hydrogen bonds with adenine, locking it tightly inside the pocket. The hydrogen bond partners of Asn146 are the same nitrogen atoms of adenine that participate in the DNA base pair. In addition Phe83 in TM3 and His225 in TM6 have good π stacking interactions with the purine ring. These features characterizing adenine binding site agree well with the empirical observations by Nobeli et al. to explain the molecular discrimination of adenine and guanine ligand moiety in complexes with proteins[22]. They observed that the protein aromatic residues stabilize an environment in which the ligand would have π stacking interaction with the side chain of these residues and that His is much more favorable for adenine. They found that amino acids with side chains like Asn that can form

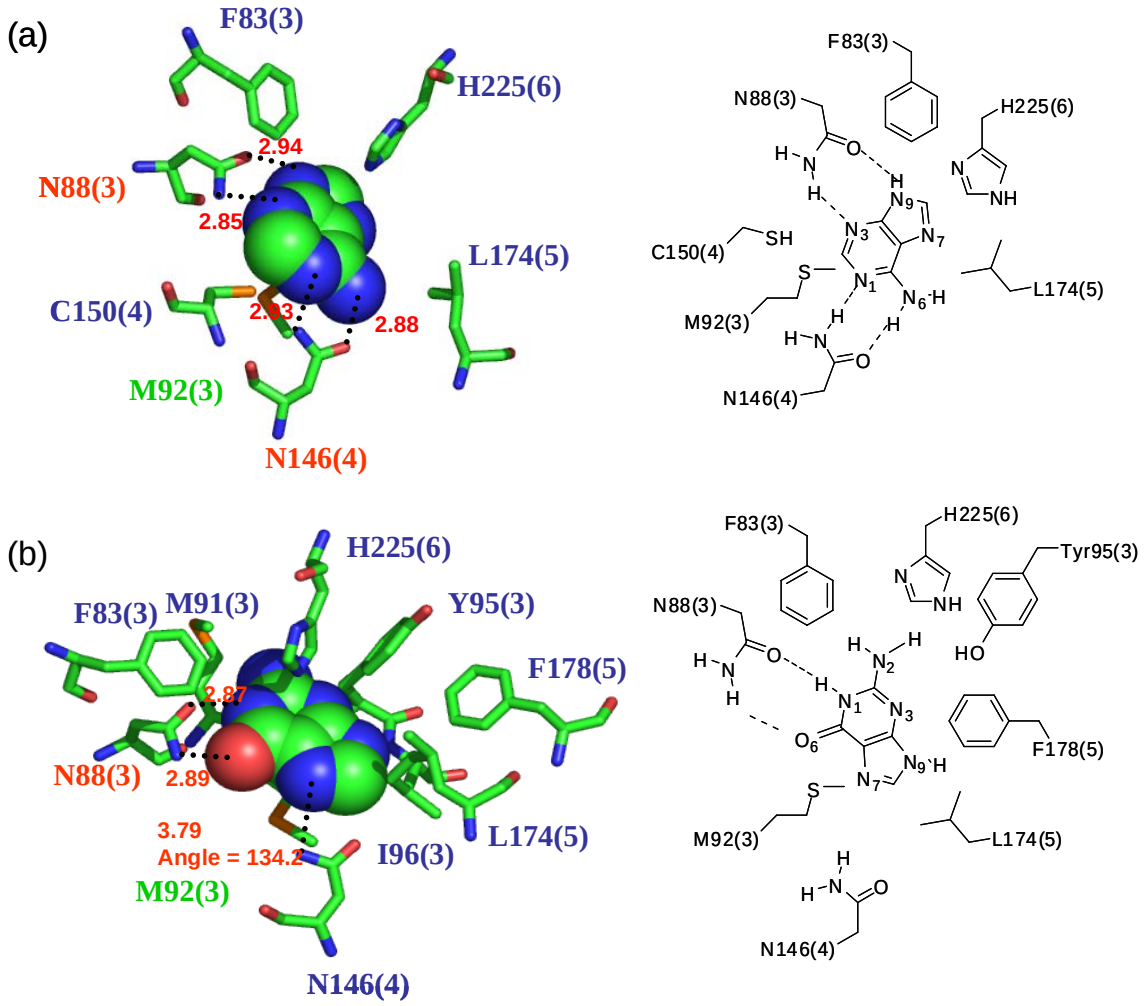

Figure 5.6 Predicted 5 Å binding pockets of adenine (top) and guanine (bottom) in the rMrgA receptor. The residue labels are colored according to the binding energy contributions from non bond interaction with the ligand:

red: greater than 10 kcal/mol contribution (best),

green: between 10 and 4 kcal/mol,

blue: worse than 4 kcal/mol (worst).

The hydrogen bonds are indicated by dotted lines with the distance between the donor and acceptor atoms. The number in parenthesis indicates the TM containing the residue.

simultaneously a donor hydrogen bonds and an acceptor hydrogen bond are favored for binding adenine.

The residues within the binding pocket in Figure 5.6 are grouped by color according to the intermolecular interaction energy with the ligand (red is strongest, blue is weakest). Here the intermolecular interaction energy includes Coulomb, van der Waals, and hydrogen bond terms. The most important are Asn88 and Asn146, which comes from strong hydrogen bond interactions. Met92 has moderate van der Waals interaction with adenine.

Predicted binding site of guanine

Changing the docked adenine structure to guanine, we find that the hydrogen bond donor and acceptor in the side chain of Asn146 does not match with the counterparts in guanine, resulting in a dramatic decrease in the predicted binding affinity (by 16 %) for guanine in this configuration. However N2 of guanine forms a new weak hydrogen bond with sulfur of Cys150.

Independently docking guanine, leads to a structure in which the guanine has the different orientation shown in Figure 5.6(b). Here its hydrogen bond interactions with Asn146 are not optimal. The carbonyl group of the Asn146 side chain loses a hydrogen bond partner and the Asn146 amine group does not make a good hydrogen bond. However the guanine retains similar interaction with the other residues.

Thus the predicted structure of rMrgA, explains the dramatic difference in bonding between adenine and guanine. Adenine can bind to both Asn in the active site leading to good hydrogen bonds for N1, N3, N6, and N9. In contrast guanine in the same configuration could make only half of these. As a result guanine binds in an alternate site where the sidechain of Asn88 form hydrogen bonds with the N1 and O6 atoms of guanine and Asn146 form a weak hydrogen bond with N7, but with binding that is 78 % weaker than for adenine. However if Tyr95 that is found nearby N2 and N3 of guanine is mutated to Gln, formation of two more

Table 5.1 Decomposition of total intermolecular interaction (kcal/mol) between ligand and rMrgA receptor, calculated for the residues within 5 Å of the ligand; the numbers in parentheses are the values relative to adenine

Ligand	Coulomb	VDW	Hbonds	TOTAL
Adenine	-2.37 (100)	-11.63 (100)	-28.17 (100)	-42.17 (100)
1MA	-1.20 (50)	-16.16 (138)	-23.87 (84)	-41.23 (97)
6BAP	-0.35 (14)	-29.90 (257)	-12.58 (44)	-42.82 (101)
HPX	-3.06 (129)	-12.64 (108)	-13.95 (49)	-29.65 (70)
Guanine	-3.27 (137)	-14.26 (122)	-16.59 (58)	-34.12 (80)
Adenosine	-1.23 (51)	-22.84 (196)	-12.95 (45)	-37.02 (87)

hydrogen bonds would be expected and might enhance the binding affinity in spite of the loss in van der Waals interactions. Indeed the predicted binding energy of guanine in the Tyr95Gln mutant is comparable to that of adenine in the wild type (99.9 % of adenine binding).

The total intermolecular interaction energy and its each component in the 5 Å binding pocket are tabulated in Table 5.1.

Predicted binding site of medium binders

For 1MA (Ki=4.4 μM) we also calculated two binding modes, one by perturbing adenine to 1MA, the other with independent docking. The binding modes of 1MA are described in Figure 5.7. The perturbed structure built by direct substitution at N1 in the docked adenine leads to a big clash between the bulky methyl group and Asn146. The independently docked 1MA is locked between Asn88 and Asn146 through hydrogen bonds with these two residues. However, this leads to slightly weakened bonding with Asn146 due to the loss of one of hydrogen bonds. This leads to a predicted binding affinity 83% of that to adenine. The methyl substituent of 1MA resides in the good hydrophobic environment.

For 6BAP, another mild binder (Ki = 58 μM), we find a docking orientation similar to that of 1MA. Here the large benzyl substituent has a close contact with Tyr95 with good π stacking interactions making the van der Waals term the dominant non bond interaction. 6BAP also forms

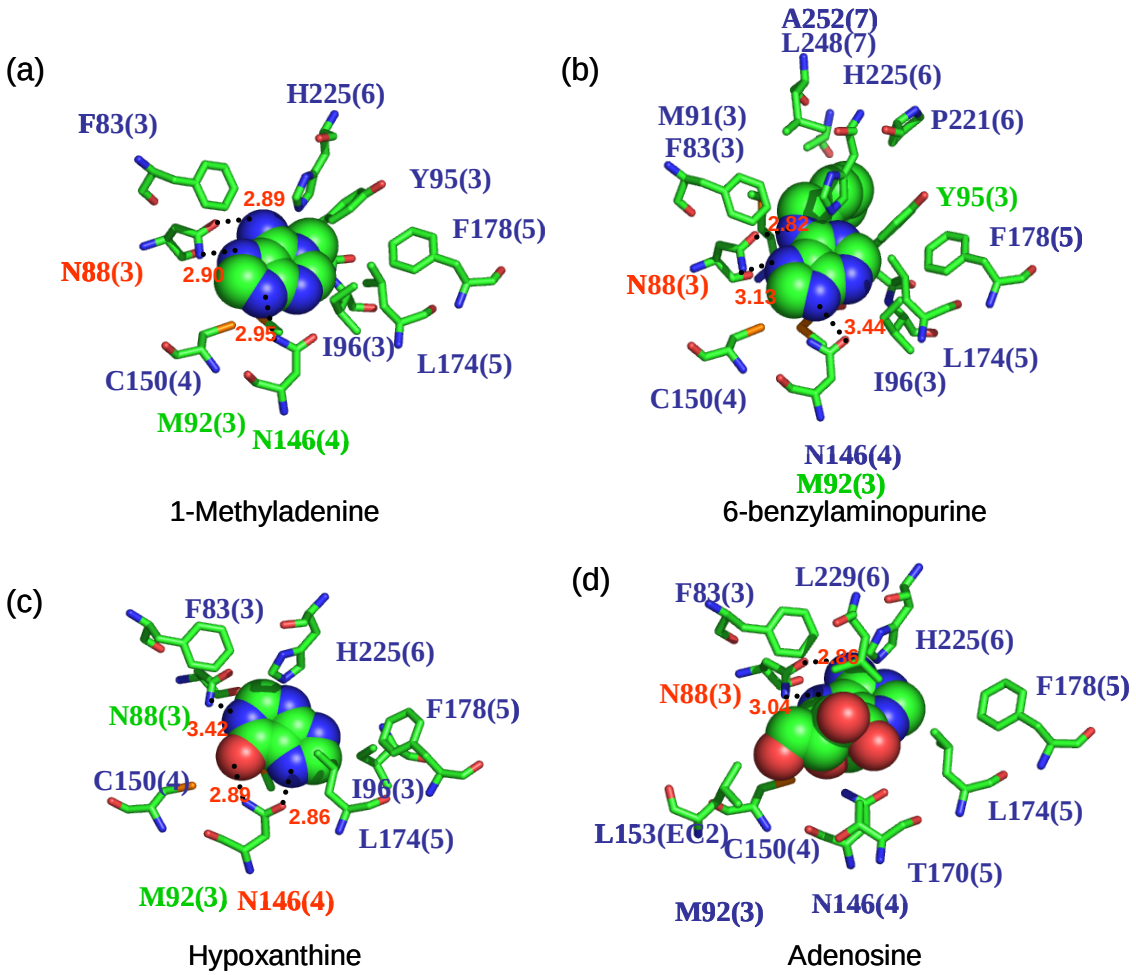

Figure 5.7 The 5 Å binding pockets for various ligands in the rMrgA receptor. The same color scheme is used as for Figure 5.6. (a) 1-Methyladenine, (b) 6-Benzylaminopurine, (c) Hypoxanthine, (d) Adenosine.

hydrogen bonds with Asn88 and Asn146, but the interaction with Asn146 is weaker than for adenine or 1MA. The loss of this interaction is partly compensated by the increased van der Waals interactions as shown in Table 5.1. The result is a binding affinity of 92 % of that of adenine.

Predicted binding site of poor binders

Hypoxanthine, one of the bad binders, makes nice contacts with Asn146 but has weak interactions with Asn88. Its hydrogen bond energy is comparable to 6BAP in Table 5.1, but the

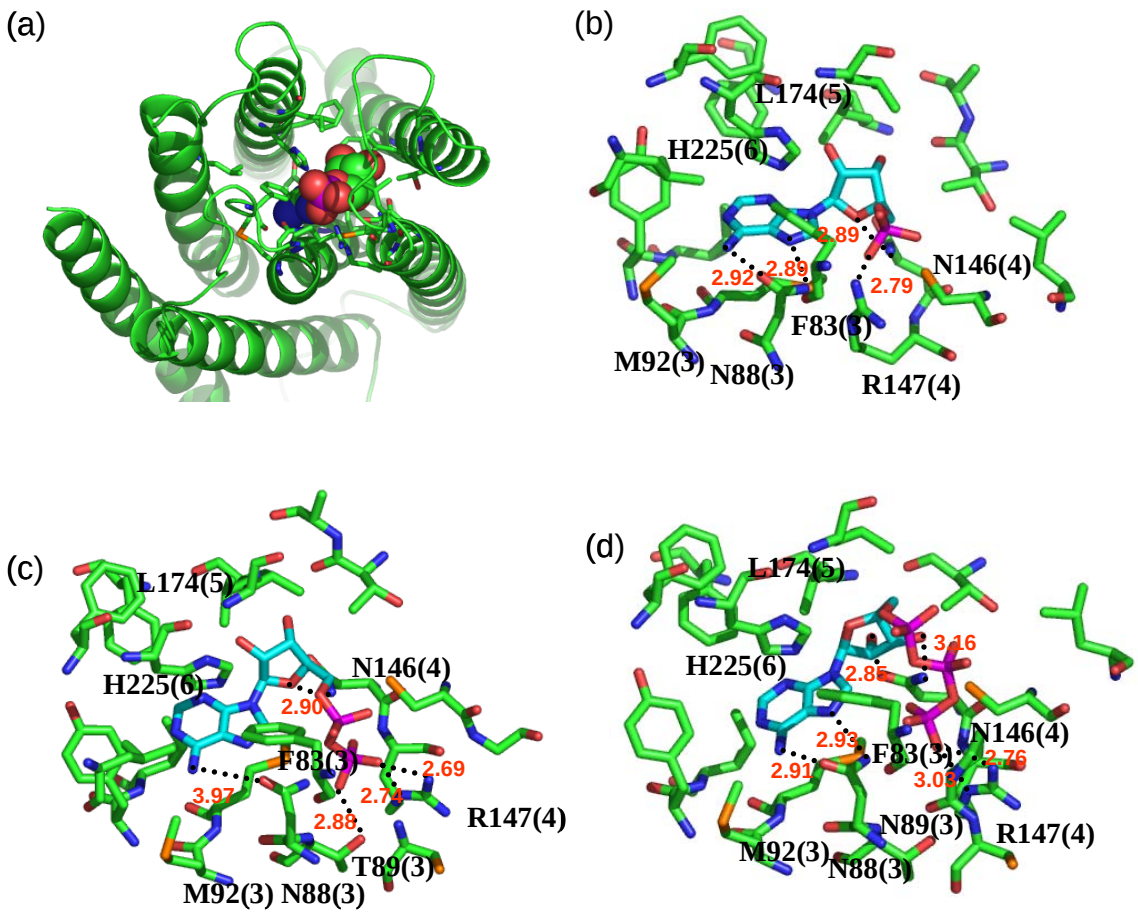

Figure 5.8 The 5 Å binding pockets of adenosine phosphates in the rMrgA receptor. (a) AMP, (b) AMP, (c) ADP, (d) ATP.

van der Waals interaction energy is insufficient to overcome the decreased hydrogen bond energy. The result is a binding affinity of 71 % of that of adenine.

For adenosine, we find that only Asn88 makes good hydrogen bond contacts with the ligand, with no other residues having good specific interactions. The result is a binding affinity of 71 % of that of adenine.

Predicted binding site of adenosine phosphates

Adenosine mono- and tri-phosphates (AMP and ADP) are observed to have binding constants to rMrgA in the range of 20-60 μM concentration. Our predicted structure is in Figure

5.8. We find that the adenine moiety forms good two hydrogen bonds with Asn88, but they have different glycosyl torsion angles. In both cases the sugar ring has a contact with Asn146. We find that the phosphate group points toward extracellular region and is stabilized by Arg147 in TM4 (on the boundary between the inside-bundle region and the membrane). This is only the positively charged residue located on the upper half of TM regions (excluding a Lys233 at the end of TM6). This further validates our prediction of binding site.

For neutral ligands such as adenine, the side chain of Arg147 leans more toward the membrane regions which might allow it to contact the head group of lipid as seen in the apo protein in Figure 5.5. However when the phosphate comes into the binding pocket, the Arg147 would move toward the pocket.

For adenosine diphosphate (ADP), the sugar ring interacts with Asn146 in the similar way to AMP but the adenine base does not interact strongly with Asn88 (see Fig. 5.8(c)). The phosphate group shows strong interaction with Arg147 and Thr89.

Comparison of calculated binding energy to ln K_i

The predicted binding energies for the various ligands are compared in Figure 5.9 with the experimental competition binding constant (inhibition constant) reported by Bender *et al.*[1] Of the nine compounds whose binding constants have been measured, we examined only the six neutral ligand with the fewest torsional degrees of freedom for docking (since the adenosine phosphates are highly negative-charged, the entropic effect in binding is no longer negligible and the uncertainty in calculated solvation energy increases). Figure 5.9 shows the good correlation between our calculated binding energy and the experimental inhibition constant, $ln\ K_i$. The calculating binding energy is for the minimized structure at 0 K, which ignores entropic effects. Except for adenosine all ligands are rigid with similar shapes so that the entropic contributions should be similar. This good correlation strongly validates our predicted structures and binding configurations.

Ligand	K_i, nM[a]	ln K_i	B.E.	B.E.[pert]
Adenine	18	2.89	100	
1-Methyladenine	4391	8.39	83	65
6-Benzylaminopurine	58328	10.97	92	74
Guanine	n.d.	-	78	83
Hypoxanthine	n.d.	-	71	52
Adenosine	n.d.	-	71	60

[a][1]
n.d.: not detectable up to the maximum concentration tried (~100 μM)
B.E.: relative binding energy (%) w.r.t adenine (52.02 kcal/mol)
B.E.[pert]: after being perturbed from docked adenine and optimized

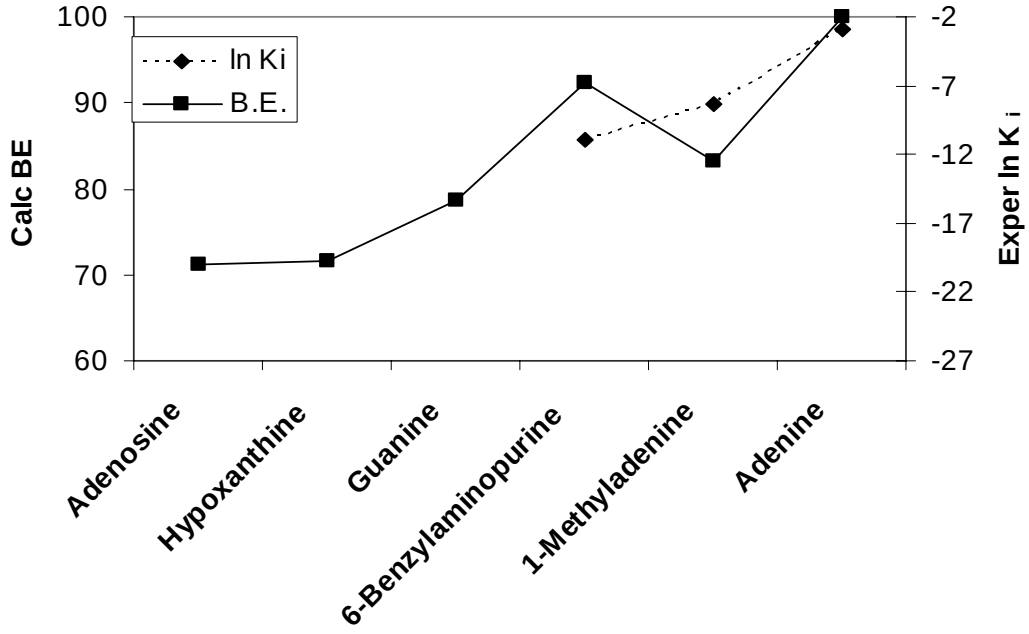

Figure 5.9 Comparison of calculated binding energies (left legend) with the experimental inhibition constants (right legend) for rMrgA ligands as described in the method section, the calculated energies are for the minimized structure (0K) without entropic contributions.

Table 5.2 Computational alanine-scanning results (SCAM) for adenine/rMrgA (energies in kcal/mol)[a]

	I.E.(WT)	ΔI.E.(Ala)	
Asn88	-17.787	17.161	41**%**
Asn146	-12.819	12.275	29**%**
Met92	-4.741	2.844	7%
Phe83	-1.545	1.433	3**%**
His225	-1.505	1.349	3**%**
Leu174	-0.665	0.432	1%
Tyr95	-0.422	0.344	0.8**%**
Ile96	-0.450	0.333	0.8%
Phe178	-0.298	0.265	0.6**%**
Cys150	-0.494	0.207	0.5%
Thr170	-0.273	0.168	0.4%
Met91	-0.321	0.161	0.4%
Arg147	-0.360	0.091	0.2%
Pro85	-0.393	0.086	0.2%
Leu177	-0.102	0.066	0.2%

[a] The intermolecular interaction energy (IE) for the wild type (WT, no mutation) is shown for all residues within 5 Å of the ligand. After mutating the residue to Ala and minimizing, we recalculated the IE of the ligand to this Ala, IE(Ala). The percentage change in binding of the mutant relative to the calculated total binding of WT is shown in the last column. These results show that the Ala mutations track well the calculated ligand-residue IE and confirm the important role of Asn88 (3), Asn146 (TM4), Met92 (TM3), Phe83 (TM3), and His225 (TM6) to the binding of adenine.

Effect of computational alanine-scanning mutations (SCAM) in the binding pocket

For the best binder, adenine, we carried out alanine scanning to assess the importance of various residues to binding. The residues within 5 Å of the ligand were each independently mutated to Ala and the energy for the ligand-protein complex was reoptimized (conjugate gradient minimization). Prior to the minimization we used SCREAM to reselect the side chain conformations of the other residues within 5 Å of the ligand. The results are summarized in Table 5.2.

As expected, the Asn88Ala and Asn146Ala mutations significantly reduce the binding affinity due to the loss of the hydrogen bonds. Mutation of either Phe83 or His225 abolishes the favorable van der Waals contacts.

The close correspondence between the contributions predicted for the wild type and the change in bonding calculated with the mutation to Ala, indicates that good estimates can be made without optimization of the coordinates.

5.3.4 Comparison of the adenine binding site in rMrgA to the nucleotide binding sites in adenosine receptors and purinergic receptors

We can compare the binding site of adenine to rat MrgA with the binding site of human A_1 and A_{2A} to adenosine (hA_1A and $hA_{2A}A$) receptors and human $P2Y_1$ to purinergic ($hP2Y_1$) receptor. These receptors all bind adenosine or ATP, with the adenine moiety in common, and all have been studied both experimentally and with modeling. The sequences of the adenosine receptors and the purinergic receptor were aligned separately with that of rMrgA receptor. The whole sequences were aligned first with Clustal-W while ensuring that specific highly conserved residues are matched to each other in the alignment: Asn at position 20 in TM1, Asp at position 13 in TM2, Arg in DRY sequence of TM3, Trp at position 12 in TM4, Pro at position 19 in TM6, Pro in NPXXY of TM7 (the number is counted from the starting residue of each TM in Figure. 5.10). Using the TM prediction of rMrgA receptor, the sequences for each TM were aligned independently. The averaged sequence identity of rMrgA receptor is ~22 % for hA_1A receptor and ~20 % for $hA_{2A}A$ receptor (considering only TM regions). For $hP2Y_1$ receptor, the TM sequence identity to rMrgA is ~24 %. The resulting TM sequence alignment is shown in Figure 5.10 where the key residues in adenosine receptors and $P2Y_1$ receptor identified from the binding or functional assay experiments are bolded and underlined[23, 24].

Recall that for rMrgA the adenine binding site mostly contacts with Asn88 (TM3), Asn146 (TM4) and Leu174 (TM5), with His225 (TM6) interacting closely with adenine.

TM1	MRGA_RAT	RTLIPNLLIIISGLVGLTGNAMVFWLLG	28
	AA1R_HUMAN	FQAAYIGI**E**VLIALVSVPGNVLVIWAVK	28
	AA2A_HUMAN	GSSVYITV**E**LAIAVLAILGNVLVCWAVW	28
	P2YR_HUMAN	QFYYLPAVYILVFIIGFLGNSVAIWMFV	28
TM2	MRGA_RAT	AFSVYILNLALADFLFLLCHIIDST	25
	AA1R_HUMAN	ATFCFIVSLAVADVAVGALVIPLAI	25
	AA2A_HUMAN	VTNYFVVSLAAADIAVGVLAIPFAI	25
	P2YR_HUMAN	GISVYMFNLALADFLYVLTLPALIF	25
TM3	MRGA_RAT	**F**LPCF**N**TV**MM**VP**YI**AGLSMLSAISTERC	28
	AA1R_HUMAN	TCLMVAC**P**VLI**LTQ**SSILALLAIAVDRY	28
	AA2A_HUMAN	GCLFIACF**V**LVL**T**QSSIFSLLAIAIDRY	28
	P2YR_HUMAN	MCKLQ**R**FI**FH**VNL**Y**GSILFLTCISAHRY	28
TM4	MRGA_RAT	KHTSTVMCSAIWVLSLLICIL**NR**YF**C**GF	28
	AA1R_HUMAN	PRRAAVAIAGCWILSFVVGLTPMFGWNN	28
	AA2A_HUMAN	GTRAKGIIAICWVLSFAIGLTPMLGWNN	28
	P2YR_HUMAN	KKNAICISVLVWLIVVVAISPILFYSGT	28
TM5	MRGA_RAT	LASNFFTAAC**L**IFL**F**VVLCLSSLALLVR	28
	AA1R_HUMAN	EFEKVISMEYMVYFNFFVWVLPPLLLMV	28
	AA2A_HUMAN	LFEDVVPMNYMVYFN**F**FACVLVPLLLML	28
	P2YR_HUMAN	FIYSMC**TT**VAM**F**CVPLVLILGCYGLIVR	28
TM6	MRGA_RAT	RLYATIMLTVLVFLLCGLPFGI**H**WFLLIWIK	31
	AA1R_HUMAN	KIAKSLALILFLFALS**W**LPL**H**IL**N**CITLFCP	31
	AA2A_HUMAN	HAAKSLAIIVGLFALCWLPL**H**II**N**CFT**F**FCP	31
	P2YR_HUMAN	KSIYLVIIVLTVFAVSYIPF**H**VM**K**TMNLRAR	31
TM7	MRGA_RAT	AYGLYLAA**L**VLT**A**VNSCANPIIYFFVG	27
	AA1R_HUMAN	PSILTYIAIFL**TH**GNSAMNPIVYAFRI	27
	AA2A_HUMAN	PLWLMYLA**I**VL**SH**TN**S**VVNPFIYAVRI	27
	P2YR_HUMAN	VYATYQVT**R**GLA**S**LNSCVDPILYFLAG	27

Figure 5.10 Sequence alignment of rat MrgA receptor with other receptors known to bind adenine components of ligands: human A_1 and A_{2A} adenosine receptors and human $P2Y_1$ purinergic receptor. The residues predicted to play an important role in ligand binding are in boldface and underlined.

In the putative A_{2A} binding site, the adenine moiety is recognized by TM3, TM5 and TM6[23]. The binding regions in TM3 overlap significantly throughout four receptors but we could not find any residue from adenosine or purinergic receptor that directly matches with Asn88 in rMrgA receptor. However, Gln92 in TM3 of hA1AR has the same functional group as Asn (shorter by one methylene) which was found to interact with the adenosine adenine moiety[25]. Asn146 in TM4 is a key residue in the adenine binding in rMrgA, but no similar residue is identified as a key residue in TM4 of adenosine or purinergic receptor. Arg157 in TM4 interacts with phosphate group of adenosine phosphates in rMrgA while Lys (TM6) and Arg (TM7) are involved in $P2Y_1$ receptor.

In conclusion, although similar residues recognize adenine, there is very little similarity in the location of the binding site of adenine in rMrgA receptor compared to adenosine and purinergic receptors. This suggests that rMrgA belongs to non-adenosine or non-purinergic receptor families even though adenine binds well and activates the receptor.

5.3.5 Comparison to other MrgA orthologs

We examined the sequences of the 8 mouse orthologs of rMrgA receptor to determine whether some might be good candidates for possible adenine binding receptors. These are collected together and compared to rMrgA in Figure S5.1. Among the eight mouse MrgA (mMrgA) receptors, we find that the mMrgA2 receptor has Asn residues at the same two positions in TM3 and TM4 as in rMrgA receptor. However, Bender *et al*. tested activation of the mMrgA2 receptor with adenine and found no activation[1]. Perhaps this is because mMrgA2 receptor does not have a proline in the middle of TM3 analogous to the Pro94 of for rMrgA receptor that we found to induce the bend in TM3. The change in the conformation of TM3 might put the Asn in TM3 of mMrgA2 receptor in the wrong orientation to bind sufficiently tightly with adenine to cause activation, explaining the lack of binding or activation by adenine mMrgA2 even though it has the same pair of Asn as rMrgA, This could be tested by mutating the Pro94 of

rMrgA to Val as in mMrgA2 to see if this causes a loss in activity or by mutating the Val94 of mMrgA2 to Pro to see if this leads to activity for adenine.

On the other hand, mMrgA5 receptor contains Pro in TM3 at the same position as in rMrgA and the Asn146 of rMrgA is also conserved. However, the Asn88 in TM3 of rMrgA is replaced with Tyr in mMrgA5 receptor. Here we suggest that mutation of Tyr87 to Asn in mMrgA5 might lead to adenine binding.

5.4 Summary and conclusions

We predicted the 3D structure of rMrgA receptor using homology to our MembStruk predicted mMrgA1 and MrgC11 structures and we predicted the binding sites for adenine and its derivatives using HierDock. The putative binding site is within TM3, 4, 5 and 6 with Asn88 in TM3 and Asn146 in TM4 serving as key residues in binding adenine. This Asn146 is homologous to Asp161 in mMrgC11 receptor that we previously identified as a key residue which was then validated experimentally. The side chain of Asn146 plays the role of the thymine in the same way as in the Watson-Crick hydrogen bond geometry of the A-T DNA base pair. It forms a bidentate hydrogen bond with both the N1 and N6 atom of adenine. The availability of the hydrogen bonds with these two Asn residues correlates with the binding affinity of the ligand.

These studies of the rMrgA receptor provide targets for mutagenesis experiments to further identify or validate important features in the binding site. This predicted binding site could be used to identify other small molecule ligands. Experimental tests of such ligands might help identify the endogenous ligand.

References

1. Bender, E., et al., *Characterization of an orphan G protein-coupled receptor localized in the dorsal root ganglia reveals adenine as a signaling molecule.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(13): p. 8573-8578.
2. Vaidehi, N., et al., *Prediction of structure and function of G protein-coupled receptors.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(20): p. 12622-12627.
3. Trabanino, R.J., et al., *First principles predictions of the structure and function of G-protein-coupled receptors: Validation for bovine rhodopsin.* Biophysical Journal, 2004. **86**(4): p. 1904-1921.
4. Bockaert, J. and J.P. Pin, *Molecular tinkering of G protein-coupled receptors: an evolutionary success.* Embo Journal, 1999. **18**(7): p. 1723-1729.
5. *MODELLER6v2*, University of California San Francisco: San Francisco.
6. 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 Research, 1994. **22**(22): p. 4673-4680.
7. Canutescu, A.A., A.A. Shelenkov, and R.L. Dunbrack, *A graph-theory algorithm for rapid protein side-chain prediction.* Protein Science, 2003. **12**(9): p. 2001-2014.
8. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - a Generic Force-Field for Molecular Simulations.* Journal of Physical Chemistry, 1990. **94**(26): p. 8897-8909.
9. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.

10. Lim, K.T., et al., *Molecular dynamics for very large systems on massively parallel computers: The MPSim program.* Journal of Computational Chemistry, 1997. **18**(4): p. 501-521.

11. 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.* Journal of Chemical Physics, 1992. **97**(6): p. 4309-4315.

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

13. Gasteiger, J. and M. Marsili, *Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges.* Tetrahedron, 1980. **36**(22): p. 3219-3228.

14. *Jaguar 5.5*. 2003, Schrodinger: Portland, Oregon.

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

16. 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.

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

18. Palczewski, K., et al., *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science, 2000. **289**(5480): p. 739-745.

19. McDonald, I.K. and J.M. Thornton, *Satisfying Hydrogen-Bonding Potential in Proteins.* Journal of Molecular Biology, 1994. **238**(5): p. 777-793.

20. Barbhaiya, H., et al., *Site-directed mutagenesis of the human A(1) adenosine receptor: Influences of acidic and hydroxy residues in the first four transmembrane domains on ligand binding.* Molecular Pharmacology, 1996. **50**(6): p. 1635-1642.

21. Ceresa, B.P. and L.E. Limbird, *Mutation of an Aspartate Residue Highly Conserved among G-Protein-Coupled Receptors Results in Nonreciprocal Disruption of Alpha(2)-Adrenergic Receptor G-Protein Interactions - a Negative Charge at Amino-Acid Residue-79 Forecasts Alpha(2a)-Adrenergic Receptor Sensitivity to Allosteric Modulation by Monovalent Cations and Fully Effective Receptor G-Protein Coupling.* Journal of Biological Chemistry, 1994. **269**(47): p. 29557-29564.

22. Nobeli, I., et al., *On the molecular discrimination between adenine and guanine by proteins.* Nucleic Acids Research, 2001. **29**(21): p. 4294-4309.

23. Kim, S.K., et al., *Modeling the adenosine receptors: Comparison of the binding domains of A(2A) agonists and antagonists.* Journal of Medicinal Chemistry, 2003. **46**(23): p. 4847-4859.

24. Jiang, Q.L., et al., *A mutational analysis of residues essential for ligand recognition at the human P2Y(1) receptor.* Molecular Pharmacology, 1997. **52**(3): p. 499-507.

25. Rivkees, S.A., H. Barbhaiya, and A.P. Ijzerman, *Identification of the adenine binding site of the human A(1) adenosine receptor.* Journal of Biological Chemistry, 1999. **274**(6): p. 3617-3621.

Supporting figures and tables

Figure S5.1 Multiple sequence alignment of rat MrgA with mouse MrgAs using Clustal-W

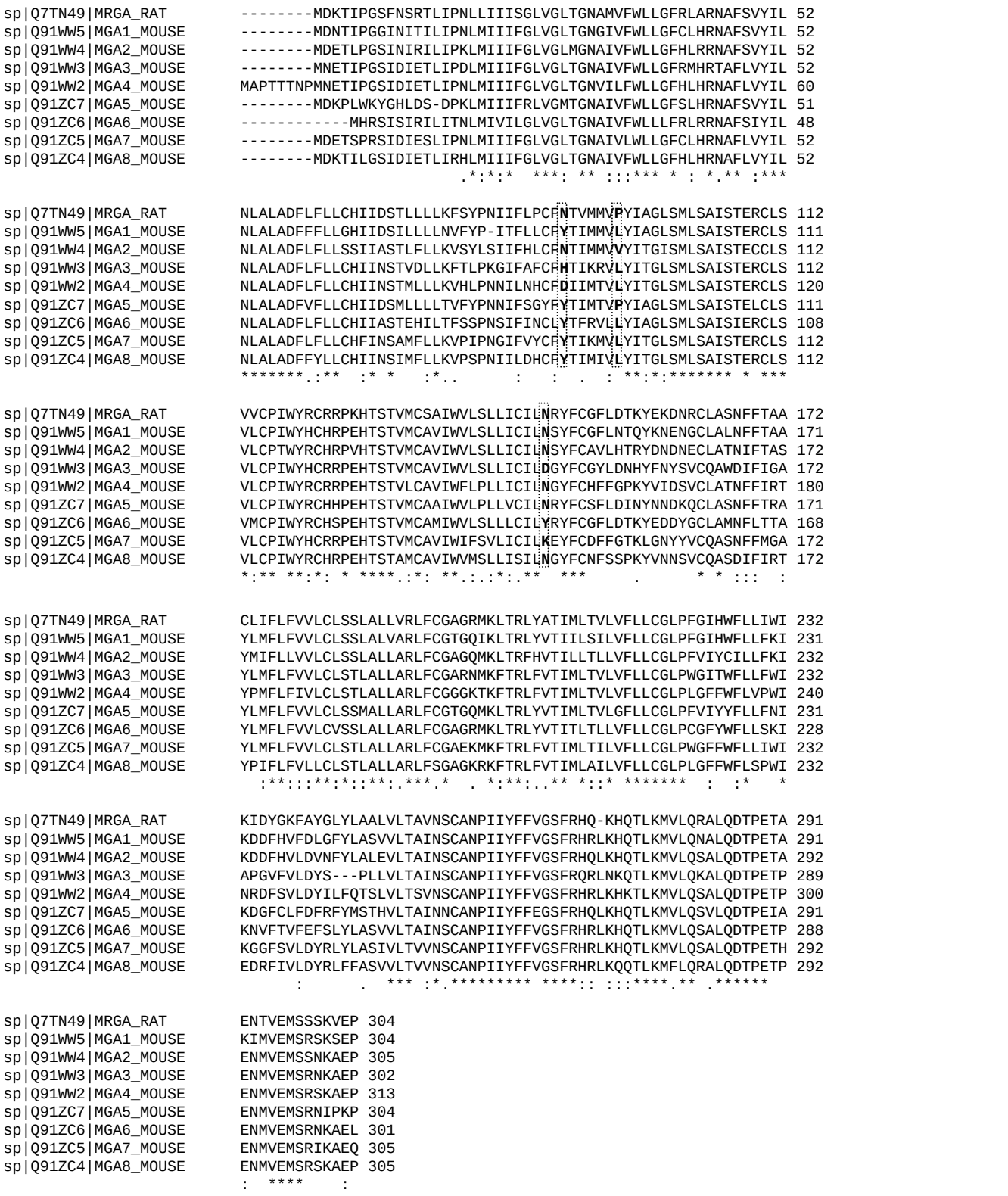

Table S5.1 The Gibbs free energies (kcal/mol) calculated from QM for various tautomeric forms of 1MA and 6BAP (numbered as shown in Figure 5.2)

Ligand	Ggas [a]	Gsol [b]	ΔGsol [c]	Relative abundance[d]
1MA1	-317838.40	-317864.54	0.00	1
1MA2	-317841.98	-317862.67	1.87	0.043
6BAP1	-462806.01	-462822.66	0.00	1
6BAP2	-462797.65	-462818.90	3.76	0.0017
6BAP3	-462788.10	-462812.30	10.36	2.5E-08

[a] Calculated using QM energy and vibrational frequencies for gas phase

[b] Calculated using Poisson-Boltzmann solvation in water

[c] relative to the most stable state

[d] abundance at 300K relative to the most stable