

# The Predicted 3D Structures of the Human M1 Muscarinic Acetylcholine Receptor with Agonist or Antagonist Bound

Joyce Yao-chun Peng,<sup>[a]</sup> Nagarajan Vaidehi,<sup>\*,[a, b]</sup> Spencer E. Hall,<sup>[a, b]</sup> and William A. Goddard, III<sup>\*,[a]</sup>

*The muscarinic acetylcholine G-protein-coupled receptors are implicated in diseases ranging from cognitive dysfunctions to smooth-muscle disorders. To provide a structural basis for drug design, we used the MembStruk computational method to predict the 3D structure of the human M1 muscarinic receptor. We validated this structure by using the HierDock method to predict the binding sites for three agonists and four antagonists. The intermolecular ligand–receptor contacts at the predicted binding sites agree well with deductions from available mutagenesis experiments, and the calculated relative binding energies correlate with measured binding affinities. The predicted binding site of all four antagonists is located between transmembrane (TM) helices 3, 4, 5, 6, and 7, whereas the three agonists prefer a site involving residues from TM3, TM6, and TM7. We find that Trp 157(4) contributes directly to antagonist binding, whereas Pro 159(4) provides an indirect conformational switch to position Trp 157(4) in the binding site (the number in parentheses indi-*

*cates the TM helix). This explains the large decrease in ligand binding affinity and signaling efficacy by mutations of Trp 157(4) and Pro 159(4) not previously explained by homology models. We also found that Asp 105(3) and aromatic residues Tyr 381(6), Tyr 404(7), and Tyr 408(7) are critical for binding the quaternary ammonium head group of the ligand through cation– $\pi$  interactions. For ligands with a charged tertiary amine head group, we suggest that proton transfer from the ligand to Asp 105(3) occurs upon binding. Furthermore, we found that an extensive aromatic network involving Tyr 106(3), Trp 157(4), Phe 197(5), Trp 378(6), and Tyr 381(6) is important in stabilizing antagonist binding. For antagonists with two terminal phenyl rings, this aromatic network extends to Trp 164(4), Tyr 179(extracellular loop 2), and Phe 390(6) located at the extracellular end of the TMs. We find that Asn 382(6) forms hydrogen bonds with selected antagonists. Tyr 381(6) and Ser 109(3) form hydrogen bonds with the ester moiety of acetylcholine, which binds in the gauche conformation.*

## Introduction

The five muscarinic acetylcholine receptors (M1–M5) belong to the family of G-protein-coupled receptors (GPCRs) and play a critical role in a wide range of diseases, including cognitive dysfunctions such as Alzheimer's disease,<sup>[1]</sup> schizophrenia, and Parkinson's disease,<sup>[2]</sup> as well as smooth-muscle disorders such as overactive bladder, irritable bowel syndrome, and chronic obstructive pulmonary disease.<sup>[3]</sup> Significant effort has been focused on identifying M1 agonists and M2 antagonists as a potential treatment of Alzheimer's disease by stimulating the cholinergic system.<sup>[1,4]</sup> The rationale for developing M1 agonists is based on the role of acetylcholine in learning and memory function and the finding that cholinergic neurons degenerate in patients of Alzheimer's disease. The development of muscarinic M1 receptor agonists for the treatment of cognitive dysfunctions resulted in a series of drug candidates, but many were discontinued because of undesirable side effects from cross-reactivity to other receptor subtypes.<sup>[1]</sup> Such relatively poor selectivity of compounds displaying muscarinic agonist activity might be expected in view of the high sequence identity (63%) within the transmembrane (TM) regions among the five human muscarinic receptor subtypes. To decrease these side effects, it would be useful to apply structure-based

drug-design methods to develop drugs with improved subtype specificity. Unfortunately there is currently no crystal structure available for any muscarinic receptor subtype of any species. The only GPCR for which there is an experimental 3D structure is bovine rhodopsin,<sup>[5]</sup> but its sequence identity to the receptors M1–M5 (16% overall and 21% in the TM regions for the M1 receptor) is too low for homology methods to predict sufficiently accurate structures to aid in structure-based drug design. Consequently, the use of M1 homology models in virtual ligand screening has led to poor hit rates for agonists and antagonists.<sup>[6]</sup>

[a] J. Y.-c. Peng, N. Vaidehi, S. E. Hall, W. A. Goddard, III  
Materials and Process Simulation Center (74-139)  
California Institute of Technology, Pasadena, CA 91125 (USA)  
E-mail: wag@wag.caltech.edu

[b] N. Vaidehi, S. E. Hall  
Present address:  
Division of Immunology, Beckman Research Institute of City of Hope  
1500, Duarte Road, Duarte, CA 91010 (USA)  
E-mail: NVaidehi@coh.org

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

We report herein the 3D structure of human M1 muscarinic receptor predicted with the MembStruk computational method. MembStruk starts with the primary sequence for the target and uses alignments with related GPCRs to predict the seven transmembrane regions, including the hydrophobic center for each TM region (used to determine the  $z$  position for each TM in the same plane) and the amphoteric moment used to orient the axis of each TM so that the hydrophobic parts project into the lipid environment. These TM translations and orientations are combined with the  $x,y$  positions and helical tilts from the low-resolution structure (7 Å) of frog rhodopsin<sup>[7]</sup> to obtain a starting structure. This structure is optimized by using a hierarchical strategy involving Monte Carlo and molecular dynamics methods to optimize the TM regions within a lipid environment, including optimization of all internal coordinates. This procedure (MembStruk 4.0) generates an ensemble of low-energy favorable packing conformations of the TM domains. Selection of these low-energy conformations is based on total energy calculated with an all-atom force field, the number of interhelical hydrogen bonds, and the number of salt bridges to determine the rotational orientation of the TM regions.

We then used the HierDock method to predict the ligand binding sites and energies for seven structurally distinct agonists and antagonists of the M1 receptor. The HierDock method makes no assumptions on the binding site and performs an unbiased search of the receptor space to predict the binding site and calculate the binding energies. The MembStruk and HierDock methods have been validated for bovine rhodopsin,<sup>[8,9]</sup> human  $\beta_2$  adrenergic receptor,<sup>[10]</sup> human D<sub>2L</sub> dopamine receptor,<sup>[11]</sup> and mouse and human olfactory receptors,<sup>[12–15]</sup> for which the predicted binding sites are in good agreement with the experimental results.

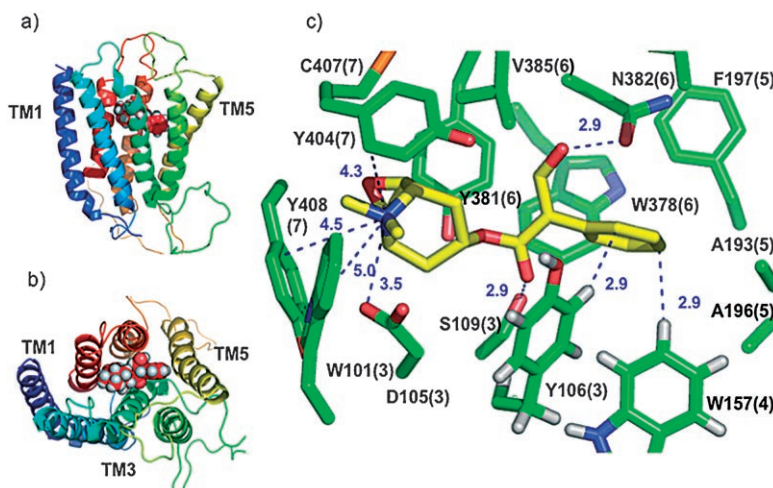
The predicted ligand–protein structures were then validated by comparison with extensive mutagenesis and ligand binding data available for the muscarinic receptors, especially the M1 receptor,<sup>[16–21]</sup> for which there is excellent agreement for both agonists and antagonists. The calculated relative binding energies also correlate well with measured ligand binding affinities. Herein, we show the binding sites for each agonist and antagonist studied, including extensive comparisons with results from known mutagenesis and binding experiments. In addition, we discuss possible mechanisms for receptor activation by comparing the agonist and antagonist binding. We also suggest new mutation candidates that we predict would affect antagonist binding.

This study provides the structural insight for improved understanding of the M1 receptor and its binding sites. Similar studies are underway for the other four subtypes of muscarinic receptors. We anticipate that such results could enable the development of subtype-specific agonists and antagonists with diminished side effects.

## Results

### Predicted structure of the human M1 receptor

We predicted the 3D structure of M1 with the MembStruk 4.0 method described in the Computational Methods section. This structure is shown in Figure 1 along with the binding site for a



**Figure 1.** a) Side view of antagonist NMS binding to M1. b) Top view of NMS binding to M1. c) Residues within 4 Å of the binding site of NMS. The numbers in parentheses refer to the TM helix to which the residues belong. Distances (Å) of the important interactions between NMS and the M1 binding site residues are indicated.

classic antagonist, the *S*-(–) enantiomer of *N*-methylscopolamine (NMS), predicted with the HierDock method (discussed below). The helical bending angles and helical tilt of the predicted M1 structure are listed in Table 1 of the Supporting Information.

The hydrophobicity profile of the multiple sequence alignment used for predicting the boundaries of the TM regions of the M1 receptor is shown in Figure 1 of the Supporting Information. The position of maximum hydrophobicity in the hydrophobicity profile of each TM helix is called the TM hydrophobic center. Table 1 lists the TM regions and the hydrophobic center of each helix predicted by MembStruk hydrophobicity profiling and capping procedures.<sup>[9]</sup> To validate the predicted structure, we used the HierDock 2.0 method<sup>[8]</sup> to predict the binding sites and binding energies for three agonists and four antagonists (Table 2).

### Predicted binding sites for agonists and antagonists in the human M1 receptor

#### Scanning the entire M1 receptor for binding sites


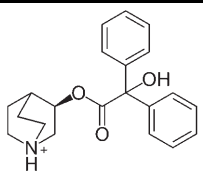
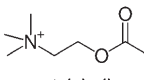
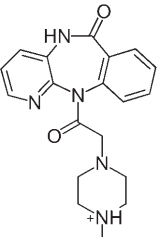
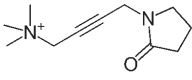
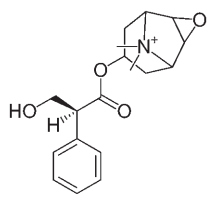
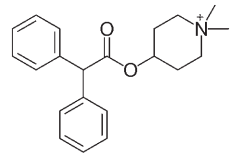
To locate the binding region of the agonists and antagonists, the void space in the entire receptor structure was partitioned into 27 regions (each  $10 \times 10 \times 10$  Å<sup>3</sup>), and the method for scanning the entire receptor for binding sites (Computational Methods, below) was used to determine the preferred binding

**Table 1.** The predicted TM regions and hydrophobic centers (bold) of each helix.

Helix	Number of Residues	Sequence
N term.	22	MNTSAPPAVSPNITVLAPGKGP
TM1	29	WQVAFIGITGLSLATVTGNLLVLISFK
IC1	6	VNTELEK
TM2	31	TVNNYFLLSLACADLIIGTFSMNLYTTYLLM
EC1	6	GHWALG
TM3	28	TLACDLWLALDYVASNASVMNLLISFD
IC2	21	RYFSVTRPLSYRAKRTPRRAA
TM4	23	LMIGLAWLVSFVLWAPAILFWQY
EC2	16	LVGERTVLAGQCYIQF
TM5 <sup>[a]</sup>	28	LSQPIITFGTAMAAFYLPVTVMCTLYWR
IC3	25	IYRETNRRARELAATFSLVKEKAA
TM6 <sup>[a]</sup>	28	RTLSAILLAFILT <b>W</b> TPYNIMVLVSTFCK
EC3	5	DCVPE
TM7	21	TLWELGYWLCYVNSTINPMCY
C term.	43	ALCNKAFRDTRLLLLCWDKRRWRKIPKRPGSVHRTPSRQC

[a] The third intracellular loop between TM5 and TM6 was truncated to 25 residues with residues 225–353 deleted; it corresponds to the dLoop mutation structure that was shown experimentally to have minimal effect on ligand binding.<sup>[41]</sup>

**Table 2.** The seven ligands used in this study.

Agonists	Antagonists
 tetramethylammonium	 (-)-quinuclidinyl benzilate (QNB)
 acetylcholine	 pirenzepine
 oxotremorine-M	 (S)-N-methylscopolamine (NMS)
	 4-diphenylacetoxy-N-methylpiperidine methiodide (4-DAMP)

region of agonists and antagonists. The ligands used for determining the binding region included NMS, QNB, acetylcholine, tetramethylammonium, and 4-DAMP. For each ligand, we considered the five most stable conformations; among those, at least 80 % of the ligand surface area was buried. This was used to select the best binding region for subsequent docking of all ligands. Next, we docked all agonists and antagonists into this best-binding region, leading to a detailed binding site with associated binding energies that were calculated with HierDock 2.0.

### Predicted binding site for the antagonist

#### N-methylscopolamine (NMS)

Figure 1 shows the HierDock-predicted binding site of the biologically active (S)-(–) enantiomer of the NMS antagonist in the predicted M1 receptor structure. Figure 1a and 1b show that NMS binds between TM helices 3, 4, 5, 6, and 7. Figure 1c shows the residues within 4 Å of the NMS ligand. The charged quaternary ammonium head group forms an electrostatic interaction with Asp 105(3) at a distance of 3.5 Å. NMS interacts with surrounding aromatic residues, including Tyr404(7), Tyr408(7), Trp 101(3), and Tyr381(6) through cation– $\pi$  interactions.<sup>[22]</sup> The hydroxymethyl group on the tropic acid side chain of NMS is hydrogen bonded to Asn382(6) at a distance of 2.9 Å. The carbonyl group of the ester moiety is hydrogen bonded to Ser109(3) at a distance of 2.8 Å. The phenyl group at the tropic acid side chain of NMS interacts with the surrounding aromatic residues Tyr106(3), Trp157(4), Trp378(6), and Phe197(5). (Figure 3c highlights the favorable aromatic–aromatic interactions between the negatively charged carbon atoms at the NMS phenyl group and the positively charged hydrogen atoms of Tyr106(3) and Trp157(4) aromatic rings.)

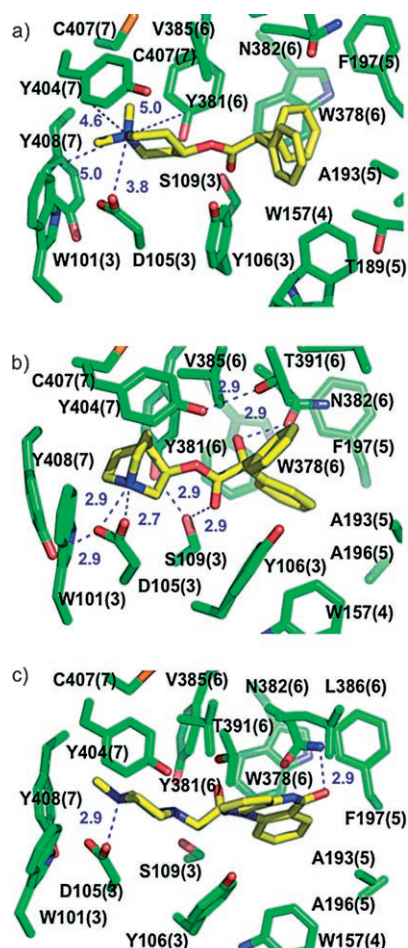
#### Predicted binding site for the antagonist 4-DAMP

The binding site of the antagonist 4-diphenylacetoxy-N-methylpiperidine methiodide (4-DAMP) is located between the TM helices 3, 4, 5, 6, and 7, as indicated in Figure 2a, which shows the 4-Å binding site of 4-DAMP. The quaternary ammonium head group forms a weak electrostatic interaction (distance = 3.8 Å) with Asp 105(3) as expected for a quaternary ammonium group. The surrounding aromatic residues Tyr404(7), Tyr408(7), and Tyr381(6) interact with the ligand head group through cation– $\pi$  interactions (Figure 2a). The carbonyl group of the ester moiety forms a hydrogen bond to Ser109(3) ( $d = 2.9$  Å). Ser109(3) also forms an interhelical hydrogen bond with Tyr381(6) ( $d = 2.9$  Å). The two phenyl groups of 4-DAMP and its surrounding aromatic residues Tyr106(3), Trp157(4), Trp378(6), Phe197(5), and Tyr381(6) form an aromatic network that stabilizes the 4-DAMP binding site.

#### Predicted binding site for the antagonist quinuclidinyl benzilate (QNB)

Figure 2b shows the residues within the 4-Å binding site of (–)-QNB, which is located between helices 3, 4, 5, 6, and 7.





**Figure 2.** Residues within 4 Å of the binding site of a) antagonist 4-DAMP; b) antagonist QNB; and c) antagonist pirenzepine. The numbers in parentheses refer to the TM helix to which the residues belong.

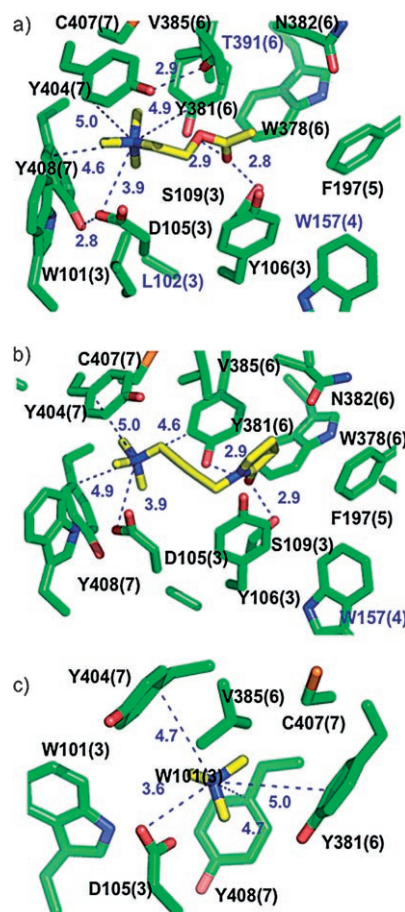
The charged protonated tertiary amine group forms a salt bridge ( $d=2.7$  Å) with Asp105(3). The hydroxy group of QNB is hydrogen bonded ( $d=2.9$  Å) to Asn382(6). The ester moiety of QNB is hydrogen bonded ( $d=2.9$  Å) to Ser109(3). As in 4-DAMP, the QNB phenyl ring is caged by the surrounding aromatic residues Tyr106(3), Trp157(4), Trp378(6), and Phe197(5), thus stabilizing the binding site.

#### Predicted binding site for the antagonist pirenzepine

Figure 2c shows the 4-Å binding site of pirenzepine, which involves residues from TM helices 3, 4, 5, 6, and 7. The charged protonated tertiary amine head group of pirenzepine forms a salt bridge ( $d=2.9$  Å) with Asp105(3). The lactam group of pirenzepine forms a hydrogen bond ( $d=2.9$  Å) with Asn382(6). As in other ligands for M1, the pirenzepine benzene ring and the surrounding aromatic residues Tyr106(3), Trp157(4), Phe197(5), Trp378(6), and Tyr381(6) form an aromatic cage that stabilizes the binding site.

#### Predicted binding site for acetylcholine: the endogenous ligand

The predicted binding site of acetylcholine is located between TM helices 3, 6, and 7, as shown in Figure 3a. The quaternary ammonium head group of acetylcholine forms an electrostatic



**Figure 3.** Residues within 4 Å of the binding site of agonists a) acetylcholine; b) oxotremorine-M; and c) tetramethylammonium. The numbers in parentheses refer to the TM helix to which the residues belong. The residues that are in close proximity but not within 4 Å from the ligands are shown in blue.

interaction with Asp105(3) at a distance of 3.9 Å. The head group is also stabilized by the surrounding aromatic side chains Tyr381(6), Tyr404(7), and Tyr408(7) through cation- $\pi$  interactions (Figure 3A). Trp101(3) is also in close proximity to the acetylcholine head group. The carbonyl group of the ester moiety of acetylcholine is hydrogen bonded to both Tyr381 ( $d=2.9$  Å) and Ser109 ( $d=2.8$  Å).

#### Predicted binding site for the agonist oxotremorine-M

Figure 3b shows the 4-Å binding site of the agonist oxotremorine-M. This agonist binds between helices 3, 6, and 7 in a location similar to that of acetylcholine. There is an electrostatic interaction ( $d=3.9$  Å) between Asp105(3) and the quaternary

ry ammonium head group of oxotremorine-M. The head group is also stabilized by the surrounding aromatic side chains including Tyr381(6), Tyr404(7), and Tyr408(7) through cation- $\pi$  interactions (Figure 3B). The carbonyl group of the ester moiety of oxotremorine-M is hydrogen bonded to both Tyr381(6) and Ser109(3) with a distance of 2.9 Å.

#### **Predicted binding site for the agonist tetramethylammonium**

The tetramethylammonium ion,  $N(CH_3)_4^+$ , is also an agonist ( $pIC_{50}=2.9^{[17]}$ ) and binds similarly to the acetylcholine head group (Figure 3c). The tetramethylammonium ion is stabilized by both the electrostatic interaction with Asp105(3) ( $d=3.6$  Å) and by cation- $\pi$  interactions with the surrounding aromatic residues Tyr381(6), Tyr404(7), and Tyr408(7). The distances of the cation- $\pi$  interactions are shown in Figure 3c.

## **Discussion**

### **Antagonist binding: comparison of the NMS binding site with mutagenesis data**

The results of alanine scanning mutagenesis<sup>[16,23]</sup> experiments, in which a single residue in TM2, TM3, TM4, TM5, TM6, or TM7 was mutated to Ala, revealed eight residues for which mutation to Ala caused a greater than 25-fold decrease in NMS affinity ( $K_d=125$   $\mu$ M).<sup>[17]</sup> Of these, seven are in the predicted 4-Å NMS binding site (Figure 1c). These residues include Asp105(3), Tyr106(3), Trp157(4), Tyr381(6), Asn382(6), Tyr404(7), and Tyr408(7), for which Ala substitution decreased the NMS binding affinity by a factor of 100, 63, 123, 800, 1000, 25, and 48, respectively.

In addition, the Ala substitution experiment for Pro159(4) caused a 63-fold decrease in NMS affinity,<sup>[19]</sup> but our calculations found that the mutation Pro159(4)→Ala causes only a very small reduction (less than 3%) in calculated binding energy to NMS. However, our predictions found that Pro159(4) induces a bend in TM4 that helps to position Trp157(4) at the NMS binding site. Our calculations started with the predicted native structure; after mutation of Pro159 to Ala, the resulting structure was energy-minimized. We did not do a full MembStruk prediction of the mutated structure. We believe that in a full re-prediction of the structure, such a mutation would likely lead to a dramatic change in the bend of the TM4 helix, which would likely change the packing of the helices. In particular, it would change the interaction with Trp157(4), which we found to be directly involved in binding to antagonists in M1. Mutation of Trp157(4)→Ala caused a 123-fold decrease in NMS binding,<sup>[19]</sup> a 132-fold decrease in 4-DAMP binding,<sup>[21]</sup> and a 330-fold decrease in QNB binding affinity.<sup>[19]</sup> As we have not re-predicted the structure for the Pro159(4) mutation, we consider that our structure explains seven of the eight mutation experiments.

Placement of TM4 in the predicted M1 structure is critical to explaining the experimental observations on Trp157. In the MembStruk-predicted M1 structure, Trp157(4) is located next to Tyr106(3) at the predicted binding site, whereas in homolo-

gy models, Trp157(4) was placed near the N-terminal end of TM3, close to Asp99(3). As the homology model did not explain the effect of Trp157 mutation experiments, it was postulated that Trp157 might form a putative secondary docking site before ligands enter the main binding site.<sup>[16,19]</sup> However, the MembStruk-predicted structure suggests that Trp157(4) interacts directly with the antagonists at the central binding site through aromatic-aromatic interactions. Our results provide a structural basis for explaining the strong effects of these mutations on ligand binding. The step of the MembStruk procedure that optimizes the relative translation of the helices by aligning the hydrophobic centers (the position of the maximum hydrophobicity as shown in Table 1) for each helix into a plane is essential in determining the ligand binding sites as shown by Trp157(4)

### **Common features of antagonist binding sites**

#### **Antagonists with a charged quaternary ammonium head group**

Many M1 antagonists, including NMS and 4-DAMP ( $K_i=900$   $\mu$ M),<sup>[21]</sup> have a charged quaternary ammonium head group. We find that at the binding sites for both NMS and 4-DAMP, the ligand head group was stabilized by the negatively charged Asp105 of TM3 and the surrounding tyrosine residues (Tyr381, Tyr404, and Tyr408) of TM6 and TM7 primarily through cation- $\pi$  interactions.

#### **Binding site of the charged tertiary amine group in antagonists**

Antagonists QNB ( $K_d=16$   $\mu$ M)<sup>[17]</sup> and pirenzepine ( $K_i=4.4$  nM)<sup>[24]</sup> have a tertiary amine rather than the quaternary ammonium head group characteristic of other ligands. The calculated energy of interaction for each residue in the binding site suggests that Asp105(3) plays a much more significant role in binding antagonists with a charged tertiary amine head group than those with a quaternary ammonium head group. (In these calculations, we used the charges from Hartree-Fock calculations (6-31G\*\* basis set).) For quaternary ammonium, we found extensive delocalization of the positive charges onto the four surrounding methyl groups (leading to a net charge of -0.45 on the N atom and 0.17 on each of the 12 H atoms), whereas the positive charges in a tertiary amine are more concentrated on the protonated hydrogen atom (0.33 on H, -0.46 on N). This leads to predicted structures with a salt bridge between the terminal oxygen atoms of the carboxylate group of Asp105(3) and the amine group of QNB ( $d=2.7$  Å) and pirenzepine ( $d=2.9$  Å) as shown in Figure 2b and 2c. This explains the dramatic experimental results, in which mutation of Asp105(3) to Ala causes a 6000-fold decrease in QNB binding, but only a 100-fold decrease in affinity for NMS.<sup>[20]</sup> In the NMS binding site, the M1 aromatic residues interact with the quaternary ammonium head group of NMS through cation- $\pi$  interactions. We predict that the Asp105(3)→Ala mutation will also have a significant effect on pirenzepine binding. As the QNB

head group is mainly stabilized by the electrostatic interaction with Asp105(3), whereas the NMS head group is primarily stabilized by the surrounding aromatic residues (Tyr381(6), Tyr404(7), and Tyr408(7)) through cation- $\pi$  interactions, our prediction is consistent with the mutagenesis studies in which it was found that the Ala substitutions of Tyr381(6), Tyr404(7), and Tyr408(7) have a strong effect on the affinity for NMS but not QNB.<sup>[17,19]</sup>

Owing to the hydrophobic nature of the antagonist binding sites, we used quantum mechanics (B3LYP with the 6-31G\*\* basis set) to calculate the optimum position of the hydrogen atom upon interaction of QNB with Asp105(3). Starting with the docked geometry, we obtained the results listed in Table 3

**Table 3.** Results from quantum mechanics (B3LYP/6-31G\*\*) geometry optimization for a complex with the QNB head group and the Asp side-chain fragment.<sup>[a]</sup>

Environment	Dielectric Constant	Proton Transferred
vacuum	1.0	yes
cyclohexane	2.0	yes
benzene	2.3	yes
methanol	33.6	no
water	80.4	no

[a] The Poisson-Boltzmann continuum solvent method was used to include solvent effects; the result indicates that in hydrophobic environments, the proton from the QNB head group transfers to Asp105 of M1.

which suggest that in a hydrophobic environment (cyclohexane with a dielectric constant of 2.0 and benzene with a dielectric constant of 2.3), the proton transfers from the QNB tertiary amine to the carboxylate group of Asp105(3), leading to two neutral species hydrogen bonding. In contrast, hydrophilic environments (methanol with a dielectric constant of 33.6 and water with a dielectric constant of 80.4) favor a protonated QNB, which forms a salt bridge with Asp105(3). As the QNB head group is surrounded by aromatic residues, we consider that the environment is similar to benzene, suggesting that proton transfer occurs upon QNB binding.

A second piece of evidence to support the proposal that the proton transfers to form neutral Asp105 is that the Asp105(3)→Asn mutation diminishes QNB binding to an undetectable level, whereas it decreases NMS affinity by only 100-fold.<sup>[25]</sup> This result is consistent because the Asp105→Asn mutation eliminates the possibility of proton transfer that contributes to the stabilization of the QNB binding site. Antagonists with a quaternary ammonium head group such as NMS lack the proton that can transfer to Asp105(3) upon binding, and this explains why Asp105(3) is more important in binding tertiary antagonists than quaternary antagonists.

#### The aromatic cage that stabilizes the ligand-receptor complex

It has been shown experimentally that the phenyl group at the end of the NMS tropic acid side chain is essential for high-affin-

ity NMS binding.<sup>[17]</sup> Our predicted structure shows that this phenyl group is stabilized by the surrounding aromatic residues including Tyr106(3), Trp157(4), and Trp378(6) (Figure 1 c). As described above, the Ala substitution mutations of both Tyr106(3) and Trp157(4) significantly decrease NMS binding, as expected. Another study<sup>[21]</sup> found that the NMS binding was decreased 115-fold for the Trp378(6)→Ala mutation, whereas the Trp→Phe mutation had negligible effect (1.8-fold). This confirms the importance of aromatic character in this residue. In the predicted NMS binding site, an aromatic cage is formed by the NMS phenyl group, Tyr106(3), Trp157(4), Trp378(6), and additional aromatic residues including Phe197(5) and Tyr381(6) that contribute to the stability of the receptor-NMS complex. The aromatic-aromatic interactions present in our predicted binding site are similar to those commonly observed in many protein crystal structures for which the positively charged hydrogen atoms at one aromatic ring lie close to the negatively charged carbon atoms at another aromatic ring, forming favorable edge-face or parallel-displaced interactions.<sup>[26,27]</sup>

The predicted binding sites of QNB and 4-DAMP show that similar aromatic-aromatic interactions are also critical for the binding of QNB and 4-DAMP, both of which contain two phenyl groups that interact with the M1 aromatic residues. The available mutagenesis results and the effect on 4-DAMP and QNB binding confirm the importance of these aromatic residues in the binding site. The known mutations include Trp378(6)→Ala which reduced 4-DAMP binding 288.2-fold,<sup>[21]</sup> Trp157(4)→Phe mutation which reduced 4-DAMP binding 131.9-fold,<sup>[21]</sup> Trp157(4)→Ala mutation which reduced QNB binding 330-fold,<sup>[19]</sup> and Tyr106(3)→Ala mutation which reduced QNB binding 31-fold.<sup>[20]</sup> All these residues are within 4 Å of the binding sites of 4-DAMP and QNB. We predict that the mutations of these aromatic residues should affect the binding affinity of pirenzepine by similar amounts.

#### Role of aromatic residues on the EC2 loop and the extracellular end of TM helices

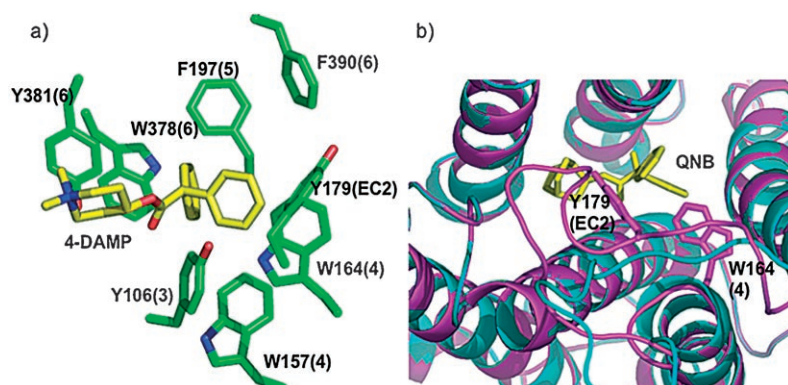
The aromatic residues that participate in the aromatic network in addition to those mentioned above include:

- Trp164(4) near the extracellular end of TM4;
- Tyr179(EC2) on the second extracellular (EC2) loop; and
- Phe390(6) near the extracellular end of TM6.

In particular, Trp164(4) and Phe390(6) are within 5 Å of 4-DAMP and interact with the upper phenyl group of 4-DAMP (Figure 4a). This extended aromatic network explains why the addition of an extra phenyl group at the end of the antagonist increases the ligand binding affinities significantly.<sup>[17]</sup>

In contrast, the above aromatic residues do not lie immediately next to QNB, which also has two terminal phenyl groups. In the crystal structure of bovine rhodopsin, it was observed that the extracellular loop 2 (the EC2 loop) is closed to form a roof over the binding site of 11 *cis*-retinal. To further optimize the EC2 loop and to examine the effect of closing the EC2





**Figure 4.** a) The extensive aromatic network at the 4-DAMP binding site. b) The structure in purple has a closed EC2 loop obtained after extensive annealing dynamics that allow the EC2 residues, the antagonist QNB, and several residues at the extracellular end of TM4 and TM5 to move, while keeping all the other residues fixed. The original structure (open loop) is shown in cyan. The closure of the EC2 loop brings Tyr 179 of the EC2 loop and Trp 164 at the extracellular end of TM4 to within 5 Å of QNB. These aromatic residues extend the aromatic cage at the QNB binding site described earlier.

loop, we conducted extensive annealing dynamics, allowing the EC2 residues, the antagonist QNB, and the terminal four residues at the extracellular end of TM4 and TM5 to move (while keeping all other residues fixed). We observed that the EC2 loop moves closer to the binding site, adopting a curved conformation similar to the EC2 loop in the bovine rhodopsin crystal structure. Figure 4b compares this EC2 closed structure with the original structure with an open EC2 loop (built using the MODELLER homology modeling program<sup>[28]</sup>). Closing the EC2 loop brings Tyr 179(EC2) and Trp 164(4) to within 5 Å of QNB, allowing these aromatic residues to participate in the aromatic network that stabilizes the QNB binding site. The predicted structure suggests that the alanine mutations of Trp 164(4) (conserved across M1–M5), Tyr 179(EC2), or Phe 390(6) would decrease the binding affinities of QNB and 4-DAMP. Such studies would help to validate the proposed role of these aromatic residues in antagonist binding.

#### The role of Asn 382 of TM6 in antagonist binding

In the predicted NMS binding site, Asn 382(6) forms a hydrogen bond with the hydroxymethyl group in the tropic acid side chain of NMS (Figure 1c). Mutation of Asn 382(6) to Ala decreased NMS affinity by greater than 1000-fold, indicating its importance to NMS binding.

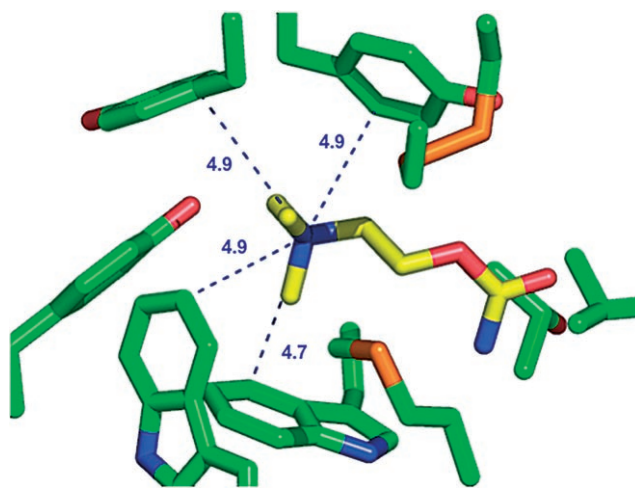
Pirenzepine binds to M1 with its lactam group forming a hydrogen bond to Asn 382(6) (Figure 2c). This hydrogen bond and the salt bridge with Asp 105(3) anchor pirenzepine to M1 from both ends. This is consistent with the strong effect of the Asn 382(6)→Ala mutation on pirenzepine binding, reducing the binding affinity by 7700-fold in M1.<sup>[24]</sup> QNB binds to M1 with its hydroxy group forming a hydrogen bond with Asn 382(6) (Figure 2b). Indeed, mutation of Asn 382(6) to Ala caused a 90-fold decrease in QNB affinity. QNB has two terminal aromatic rings that can interact with the aromatic residues in the binding site, whereas NMS has only one aromatic ring.

This explains why mutation of Asn 382(6) to Ala had a stronger effect on NMS binding affinity (decreasing it over 1000-fold) than on QNB affinity (decreasing it by 90-fold).

#### The binding site of the ester moiety of antagonists

For all the antagonists except pirenzepine, our predicted binding sites have the carbonyl oxygen atom of the ester moiety hydrogen bonded to the hydroxy group of Ser 109(3). However, the calculated interaction energy of these ligands with the residues in the binding site (within 5 Å) suggests this to be a relatively weak hydrogen bond. For example, the calculated non-bond interaction energy is  $-1.3 \text{ kcal mol}^{-1}$  between NMS and Ser 109(3), and  $-2.0 \text{ kcal mol}^{-1}$  between QNB and Ser 109(3). In comparison, the interaction energy between NMS and Asn 382(6) is calculated to be  $-6.8 \text{ kcal mol}^{-1}$ , which explains why the Ala mutation of Ser 109(3) did not have a significant effect on NMS and QNB binding.<sup>[20]</sup> Because the ester

acts as a relatively weak hydrogen bond acceptor,<sup>[29]</sup> we suspect that anchoring the two ends of an antagonist is sufficient for its high-affinity binding. This view is consistent with the crystal structure complex of carbamylcholine bound to acetylcholine-binding protein,<sup>[30]</sup> a water-soluble homolog of the ligand-binding domain of the nicotinic acetylcholine receptor. In this crystal structure complex, there is no hydrogen bond between the receptor and the carbamylcholine ester moiety (Figure 5).



**Figure 5.** Acetylcholine binding to nicotinic acetylcholine receptor; the cation- $\pi$  interaction distances (Å) which stabilize the acetylcholine head group are highlighted.

#### The agonist binding site in the M1 receptor

##### The binding site of the charged quaternary ammonium head group: comparison with the nicotinic acetylcholine binding site

Although acetylcholine is the endogenous agonist for the muscarinic acetylcholine receptors, its binding affinity to M1 is only

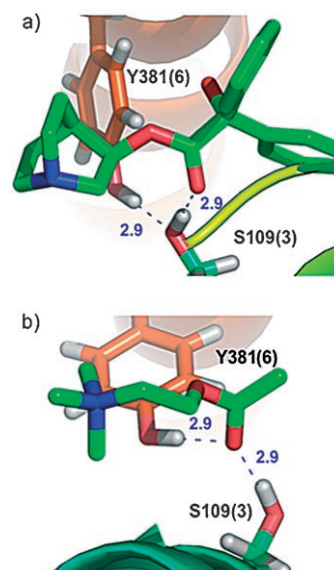
17  $\mu\text{M}$ .<sup>[17]</sup> Oxotremorine-M is an agonist with a binding affinity of 9.9  $\mu\text{M}$ .<sup>[21]</sup> In the predicted binding sites (shown in Figure 3a and 3b), the quaternary ammonium head group of both acetylcholine and oxotremorine-M are stabilized through electrostatic interaction by the negatively charged Asp 105(3) as well as through cation- $\pi$  interactions by the surrounding aromatic residues Tyr 381(6), Tyr 404(7), and Tyr 408(7). These residues were shown experimentally to be important in acetylcholine binding. For example, acetylcholine mustard, an analogue of acetylcholine in which the head group is replaced by a chemically reactive aziridinium moiety, selectively labeled Asp 105.<sup>[31]</sup> In addition, the Ala substitution mutations of Tyr 381(6), Tyr 404(7), and Tyr 408(7) each decreased acetylcholine binding affinity by about 30-fold.<sup>[17,19]</sup> The distances between the acetylcholine head group and the surrounding aromatic rings (Figure 3a) are similar to those observed in the crystal structure<sup>[30]</sup> of the nicotinic acetylcholine receptor binding site (Figure 5). Similar distances between the quaternary ammonium group and the phenyl rings were also observed in the binding sites of antagonists (NMS and 4-DAMP) that have a quaternary ammonium head group (Figure 1c and Figure 2a).

#### Binding site of the ester moiety of acetylcholine leads to breaking the interhelical hydrogen bond between TM3 and TM6

In the predicted binding site in M1, the carbonyl group of the acetylcholine ester moiety and the analogous carbonyl group at the oxotremorine cyclic amine are hydrogen bonded to both Tyr 381(6) and Ser 109(3) (Figure 3a and 3b). This is consistent with the observation that mutation of Tyr 381(6) and Ser 109(3) to Ala decreased acetylcholine affinity by 30- and 10-fold, respectively.<sup>[17,20]</sup> In addition, it has been shown<sup>[17]</sup> that the hydroxy group of Tyr 381(6) is important in acetylcholine binding in M1 because the mutation of this residue to Phe decreased acetylcholine affinity by a magnitude similar to that of an Ala mutation (while having negligible effects on NMS binding). In M2, the analogous mutation (Tyr 403  $\rightarrow$  Phe) also has a differential effect on agonist and antagonist binding. This mutation reduced the NMS and QNB affinities 2.5- and 1.7-fold, respectively, but decreased acetylcholine affinity 280-fold.<sup>[32]</sup>

To test this differential effect on agonist and antagonist binding in the predicted M1 structure, we mutated Tyr 381(6) to Phe at the NMS, QNB, and acetylcholine binding sites and recalculated the binding energy. We observed a small decrease (6%) in binding energy for NMS and no change in binding energy for QNB. However, there is a significant decrease (44%) in binding energy for acetylcholine, which parallels the experimental data.

Our predicted structure has an interhelical hydrogen bond (2.9 Å) between Tyr 381(6) and Ser 109(3). Binding of the agonists (acetylcholine and oxotremorine-M) to this site breaks this interhelical hydrogen bond; the molecules wedge between the two residues and compensate by making a hydrogen bond with both the ester moiety of TM3 and Tyr 381(6) of TM6 (Figure 6b). The interhelical hydrogen bond between Tyr 381(6) and Ser 106(3) is not broken upon antagonist bind-



**Figure 6.** The interhelical hydrogen bond a) between Tyr 381(6) and Ser 109(3) is not disrupted by antagonist binding (QNB, for example); it breaks b) upon agonist (such as acetylcholine) binding.

ing to this site (Figure 6a). The breakage of this interhelical hydrogen bond upon agonist binding might contribute to the receptor activation by weakening the intramolecular contacts between TM3 and TM6.

#### The internal torsional conformation of docked acetylcholine

Our predicted binding site has acetylcholine in the *gauche* conformation with an N-C-O torsion angle of 59°. This is the lowest-energy conformation determined from the extensive internal conformational sampling of the HierDock simulated annealing procedures used for predicting the binding site. This predicted torsion angle corresponds extremely well to the experimentally observed value (an N-C-O torsion angle of 60°) in a transferred NOE NMR study of the acetylcholine analogue (*S*)-(+)-acetyl- $\beta$ -methylcholine bound to the ground state of the M2 receptor.<sup>[33]</sup> The same experiment also suggested that the conformation of bound acetylcholine in the activated state of the receptor is *trans* with an N-C-O torsion angle of 132°. This result suggests that our predicted structure might correspond to the inactive state of the receptor.

#### The role of TM5 in agonist and antagonist binding

In the rotational optimization of the helical orientations (Computational Methods and Figure 11 of the Supporting Information), two possible low-energy orientations were found for TM5. We refer to them as 0° and 90° structures. Above, we discussed the 0° rotation structure because it is lowest in energy in the rotational energy scans for TM5. This structure has TM5 at the distal end of the binding site with Ala 193(5) pointing toward the binding site. This rotational orientation of TM5 corresponds well to the experimental results reported by Allman et al.,<sup>[18]</sup> who mutated each of the residues Ile 188(5)-Ala 196(5)



to Cys and found that only the Ala 193→Cys mutation affected the binding affinity (a decrease of 3.6-fold). Additionally, the sulfhydryl reagent *N*-trimethyl-2-aminoethyl methanethiosulfonate reacted selectively with Ala193(5) and the residues on both sides of Ala193 (Phe190(5) and Thr192(5)) when they were mutated to Cys.<sup>[18]</sup>

Allman et al. also proposed that an agonist-induced rotation of TM5, in which Thr192(5) points toward the binding site, might accompany the receptor activation.<sup>[18]</sup> This is consistent with our observed second-energy minimum in TM5 at the orientation in which TM5 is rotated 90° clockwise (as observed from the extracellular side). This puts Thr192 toward the binding site, which is exactly the TM5 orientation of the active state suggested by Allman and co-workers.<sup>[18]</sup>

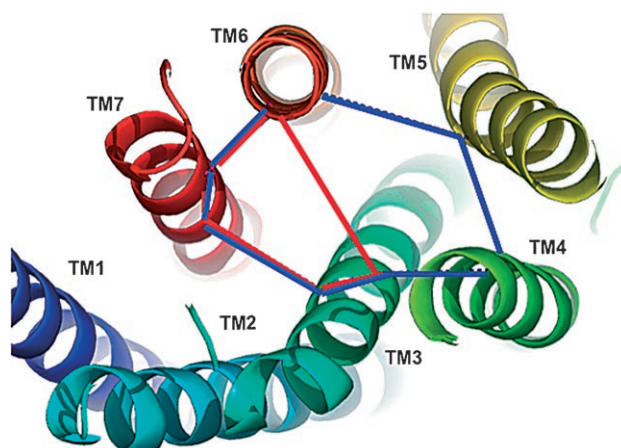
### The role of Trp157(4) in agonist binding

The Trp157(4)→Ala mutation that most significantly affected antagonist binding also decreased acetylcholine binding affinity 90-fold. In the predicted structure, Trp157(4) is within 5.5 Å of the bound acetylcholine (Figure 3a) and is a part of the aromatic network. Thus we expect that mutation of Trp157(4) to Ala would affect the packing of the binding site residues including its immediate neighbor, Tyr106(3), which interacts with acetylcholine. With TM5 rotated 90°, which we believe might correspond to the orientation in the activated state, the residues Thr189(5), Thr192(5), and Ala196(5), of which the Ala mutation selectively decreased acetylcholine binding by 5–10-fold, are located immediately next to Trp157(4). This suggests a possible role of Trp157(4) in acetylcholine binding in the active state of the receptor.

### Comparison of agonist and antagonist binding sites

A comparison of the structures of M1 agonists and antagonists shows that acetylcholine, oxotremorine-M, and many other agonists can be superimposed on the structure of the antagonists NMS and 4-DAMP. They share the quaternary ammonium head group and the ester moiety; however, acetylcholine lacks the two phenyl groups of 4-DAMP and the hydroxymethyl and phenyl group at the end of the NMS tropic acid side chain that can form aromatic–aromatic interactions or hydrogen bonds with the binding site residues of TM3, TM4, TM5, and TM6. We suspect that it is through this stronger interaction with the TM helices that antagonists keep the receptors from activation, whereas agonists allow the conformational flexibility required for activation. A similar view was suggested earlier.<sup>[34]</sup>

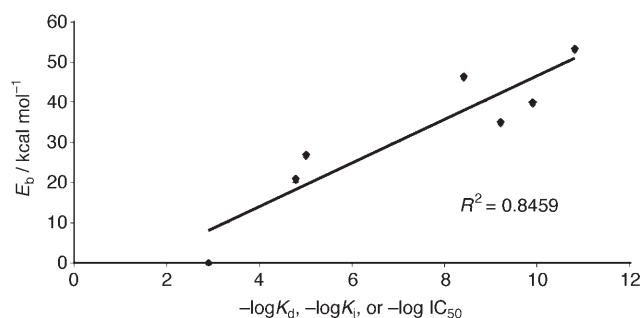
The antagonist nature of benzilylcholine<sup>[17]</sup> also supports the above view on activation. In benzilylcholine, two phenyl groups and a hydroxy group (as observed in QNB) replace the terminal methyl group (next to the ester) of acetylcholine, which switches it from an agonist to an antagonist. Figure 7 summarizes the binding regions of antagonists and agonists reported herein. The binding regions for agonists and antagonists overlap between TM3, TM6, and TM7, whereas the antagonist binding region extends to TM4 and TM5.



**Figure 7.** Comparison of the binding regions for antagonists (blue) and agonists (red). There is a significant overlap of the two binding regions in TM3, TM6, and TM7, but the binding region for antagonists extends to TM4 and TM5.

### Comparison of the calculated ligand binding energies and experimental dissociation constants

The calculated binding energies of the ligands used in this study are compared with the experimental dissociation constants (either  $-\log K_d$ ,  $-\log K_i$ , or  $-\log IC_{50}$ ) in Figure 8 ( $R^2 = 0.85$ ). We consider this to be a good correlation, as these mole-



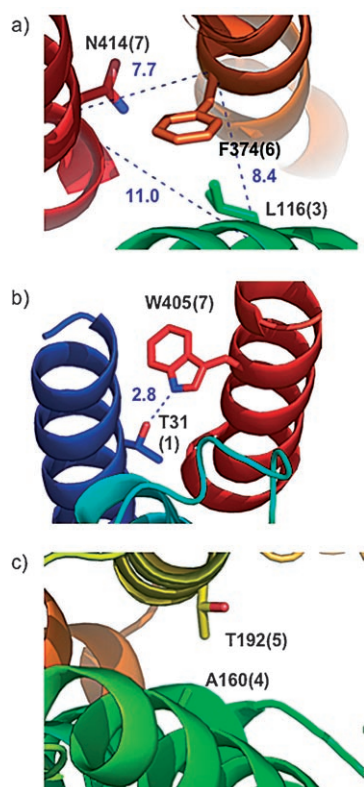
**Figure 8.** Comparison of calculated relative binding energies ( $E_b$ ) for the seven ligands studied with the experimental dissociation constants for the M1 receptor. The strongest binders are in the upper right-hand region. No corrections to the binding energies were made for zero-point energy, dynamics, or entropy.

cules represent a diverse set of ligands, including both agonists and antagonists. Our calculations ignore many important effects likely to be important to total binding energy, including contributions the changes in entropy upon binding and the effect of thermal vibrations on the binding site. However, we believe that the nature of the best binding site is accurately predicted.

### Additional validations

It has been observed experimentally that substitutions of Leu116(3), Phe374(6), and Asn414(7) with one or more His res-

idues lead to the formation of a  $\text{Zn}^{2+}$  binding site in the ground state of the M1 receptor.<sup>[35]</sup> Our predicted structure has Leu116(3), Phe374(6), and Asn414(7) in close proximity. The predicted  $\alpha$ -carbon distances among these three residues are 7.7, 8.4, and 11.0 Å (Figure 9a). For the imidazole side chains to coordinate a metal ion, the  $\alpha$ -carbons of these three residues must be separated by less than 13 Å,<sup>[35]</sup> which is consistent with the observed formation of the  $\text{Zn}^{2+}$  binding site (Figure 9a)



**Figure 9.** Interhelical interactions in the predicted structure of the M1 receptor. a) In the predicted structure, Leu116(3), Phe374(6) and Asn414(7) are located in close proximity. This is confirmed by experiments in which these three residues, upon mutation to His, were observed to form a  $\text{Zn}^{2+}$  binding site in the ground state of the M1 receptor.<sup>[35]</sup> b) The predicted interhelical hydrogen bond between Trp405(7) and Thr32(1) is suggested to contribute to the stabilization of the inactive state of the receptor. c) Ala160(4) is predicted to interact with Thr192(5) through hydrophobic interactions. The disruption of this interaction might contribute to the stabilization of the active state of the receptor.

Mutation of Leu116(3) to Ala decreased the receptor expression level, increased agonist affinity, and raised constitutive activity, a phenotype suggested for residues that form intramolecular contacts important for stabilizing the inactive state but not the activated state of the receptor.<sup>[16]</sup> The Ala substitution of Phe374(6) and Asn414(7) also caused a 10-fold decrease in the receptor expression level of M1. For the M5 receptor, the Ala substitution of the residue analogous to Phe374(6) in M1 was also reported to induce constitutive activity.<sup>[16]</sup> These experimental results are consistent with our predicted structure,

in which these three residues contact each other at the core of the receptor.

Among other residues for which Ala substitution increased acetylcholine affinity and basal signaling activity (but to a lesser extent) are the following:

- Trp405(7), which forms an interhelical hydrogen bond with Thr32(1) in the predicted structure (Figure 9b);
- Ala160(4), which forms hydrophobic interactions with Thr192(5) (Figure 9c), a residue shown to affect agonist but not antagonist binding.<sup>[24]</sup> As discussed earlier, TM5 might adopt different rotational orientations in the active and inactive states. The mutation of Ala160(4) would disrupt its hydrophobic interaction with Thr192(5) which might contribute to the stabilization of the other energy minimum corresponding to the active state in our scanning result;
- Ser120(3), which points toward the core of the receptor and contacts both Leu116(3) and Phe374(6).

In addition, several residues that are highly conserved among GPCRs are predicted to be located at positions similar to the respective positions in bovine rhodopsin. For example, the conserved Trp150(4) is located near both TM2 and TM3. Asp71(2), the most conserved aspartic acid residue across many GPCRs on TM2, is hydrogen bonded to Asn43(1). The compatibility of our predicted structure with the roles for the various considered residues and with the Ala mutagenesis results gives further credence to the accuracy of the predicted structure.

## Conclusion

The predicted 3D structure of the human M1 receptor and the predicted binding sites of various agonists and antagonists are consistent with the effects on ligand binding in a large variety of mutagenesis experiments. In particular, the predicted binding sites provide a structural basis to explain:

- the large decrease in ligand binding affinity and signaling efficacy by mutations of Trp157 and Pro159 on TM4 which was previously unexplained. (Trp157(4) contributes directly to ligand binding at the binding site. Pro159(4) provides an indirect conformational switch to position Trp157(4).);
- the differential effect of the Tyr381(6)→Phe mutation on agonist and antagonist binding;
- the stronger decrease in binding affinity by the Asp105(3)→Ala mutation for antagonists with a tertiary amine head group; and
- the larger effect of the Asn382(6)→Ala mutation on NMS than on QNB binding.

We suggest several further mutation studies that would test the predictions, including:

- Trp164 at the extracellular end of TM4 (for its effect on the binding affinities for antagonists with two terminal phenyl groups) and

- additional aromatic residues (Phe197(5), Tyr179(EC2), and Phe390(6)) that are part of the aromatic network stabilizing the antagonist binding sites.

The predicted binding site for all four antagonists studied is located between TM helices 3, 4, 5, 6, and 7, whereas the predicted binding site for the three agonists involves residues from helices 3, 6, and 7. In addition to the critical Asp105(3), aromatic residues Tyr381(6), Tyr404(7), and Tyr408(7) are important for binding the quaternary ammonium head group of the ligand through cation- $\pi$  interactions. For ligands with a charged tertiary amine head group, we suggest that proton transfer from the ligand to Asp105(3) occurs upon binding. Moreover, we find that an extensive aromatic network involving Tyr106(3), Trp157(4), Phe197(5), Trp378(6), and Tyr381(6) is important in stabilizing the binding of antagonists. For antagonists with two terminal phenyl rings, this aromatic network extends to Trp164(4) and possibly Tyr179(EC2) and Phe390(6) located at the extracellular end of the TMs. Asn382(6) forms hydrogen bonds with selected antagonists. Tyr381(6) and Ser109(3) form hydrogen bonds with the ester moiety of acetylcholine, which binds in the *gauche* conformation.

The validated detailed structure of M1 and its binding sites provides the basis for initiating structure-based drug design of these systems. It would be most useful at this point, to carry out similar studies to predict the structures and ligand binding affinities of the other four muscarinic receptor subtypes. Together these five structures would serve as targets to design subtype-selective agonists and antagonists.

## Computational Methods

**Structure prediction of GPCR:** *The MembStruk method:* We used MembStruk method (version 4.0) to predict the M1 structure. A detailed description of the MembStruk method (version 3.5) has been published.<sup>[9]</sup> This procedure (MembStruk 4.0) generates an ensemble of low-energy favorable packing conformations of the TM domains. Selection of these low-energy conformations is based on total energy calculated by using an all-atom force field, the number of interhelical hydrogen bonds, and the number salt bridges to determine the rotational orientation of the transmembrane (TM) regions. This resulting ensemble of conformations involving various rotational orientations of the helices is used to determine the ligand binding sites. Details of the MembStruk 4.0 method are given in reference [36].

The accuracy of structure prediction using MembStruk has been validated by comparison with the crystal structure of bovine rhodopsin.<sup>[8,9]</sup> Furthermore, for several other GPCRs, the structure has been validated by predicting the binding sites of strongly bound agonists and antagonists, and comparing these against the available mutation data. This has been done for the human dopamine D<sub>2L</sub> receptor<sup>[11]</sup> and for the human  $\beta$ 2 adrenergic receptor.<sup>[10]</sup> Additionally, predictions of the structure and odorant binding sites in olfactory receptors are supported by experimental data on odorants that elicit response from olfactory receptors.<sup>[12–15]</sup> These validations show that the structures of GPCRs predicted with MembStruk are sufficiently accurate to predict binding sites for ligands to GPCRs. These methods serve as an alternative to the earlier ap-

proach of constructing GPCR models from homology to the crystal structure of bovine rhodopsin; the earlier approach requires substantial experimental mutagenesis data as constraints in constructing the GPCR model.

All energy and force evaluations were conducted with the DREIDING force field<sup>[37]</sup> and CHARMM22 charges<sup>[38]</sup> for the protein. The following section describes the details of the various steps as applied to prediction of the human M1 structure.

**TM prediction (TM2ndS):** We predicted each TM region by using a BLAST search, which found 51 sequences with sequence similarities to M1 in the range between 24 and 99%. This set includes 45 M1–M5 sequences from various species, four sequences from histamine receptors, and two from octopamine receptors. The M1–M5 receptors share over 40% sequence identity overall and 60% sequence identity in the TM regions. The histamine and octopamine receptors identified by BLAST have the highest sequence identity to M1 from outside of the muscarinic receptor family. Histamine and M1 receptors have 29% sequence identity overall and 37% sequence identity in the TM regions, whereas octopamine and M1 receptors have 24% sequence identity overall and 35% sequence identity in the TM regions. The inclusion of sequences with lower homology to M1 improved the resolution in identifying the boundary between TM2 and TM3 (the EC1 loop), which is highly hydrophobic in all receptors M1–M5. We then calculated the average hydrophobicity for every residue position over all the sequences in the multiple sequence alignment (performed using ClustalW<sup>[39]</sup>) and averaged over window sizes of 12–20 residues. The baseline for this profile served as the threshold value for determining the TM regions (Figure 1). As a helix can extend past the membrane surface, we carried out an additional helix-capping step as described by Trabaino et al.<sup>[9]</sup> based on properties of known helix breaker residues to further refine where each predicted TM region ends.

**Optimization of the translational orientation of the helices:** TM2ndS also predicts the position of maximum hydrophobicity known as the “hydrophobic center” for each helix. The seven canonical  $\alpha$  helices were built with the predicted TM sequences, and the helical axes were positioned based on the 7.5-Å density map of frog rhodopsin.<sup>[7]</sup> Relative translational orientation of the helices was optimized by fitting the hydrophobic centers of the TM helices to a common plane.

**Optimization of helical bends and kinks:** We constructed canonical helices for the predicted TM segments and optimized the individual helices by using energy minimization followed by Cartesian molecular dynamics (MD) at 300 K for 500 ps. The best energy structure was chosen for each helix from the MD trajectory. This procedure optimizes the bends and kinks in each helix.

**Optimization of rotational orientation of the helices:**

**Step 1, hydrophobic moment:** The rotational orientation of the helices was determined according to the hydrophobic moment of the middle portion of each helix to point the moment of each helix towards the lipids.

**Step 2, Monte Carlo optimization (RotMin):** In this procedure, each of the seven TMs was optimized through a range of rotations one at a time while the other six helices were re-optimized in response. After each 5° rotation, the side-chain rotamers were reassigned for all the residues of all seven helices in the TM region using SCWRL.<sup>[40]</sup> The potential energy of the rotated helix was then minimized in the field of all the other helices. This procedure was carried out for a grid of rotation angles (every 5° for a range of  $\pm 25^\circ$ ) to determine the optimum rotation of each helix. Once the helix was rotated to its optimum rotation, the helix was kept fixed while



allowing each of the other six helices to be rotated and optimized. The optimization of these six helices was done iteratively until the entire grid of rotation angles was searched. This method optimized the rotations locally.

**Step 3, Energy scan:** This step is a recent refinement over the MembStruk 3.5 procedure. It is included only in MembStruk 4.0 and is described in detail in reference [36]. We took the structure generated from Step 2 as the initial structure and rotated each individual helix while fixing the other six helices. We rotated the helix in 5° intervals for 360° and calculated the energy of the rotated helix after reassigning the side-chain conformation (with SCRWL) and minimizing the energy of the entire protein.<sup>[40]</sup> We also calculated the number of interhelical hydrogen bonds and salt bridges to identify favorable interhelical interactions. Step 3 allows the identification of additional low-energy rotation orientations for each helix, assuming that Steps 1 and 2 predicted good rotational orientations for other helices.

We conducted energy scans for each of the seven helices. Subsequently, the conformation corresponding to the minimum energy found from the scans for each helix was taken, and a TM bundle was assembled. We repeated the energy scan procedure until we reached a structure for which the minimum for each of the seven helices is within 15° of its lowest-energy minimum from the previous scan, indicating convergence to a minimum. For the case of M1 receptor, we had to perform two iterations for convergence. The energy scanning results are shown in Figures 2–15 of the Supporting Information; the 0° rotational orientation represents the final conformation of each helix. As Steps 2 and 3 are more important for optimizing helix 3, which is less amphipathic owing to its location near the center of the bundle, we docked ligands to both the 0° conformation and the conformation of its second-lowest-energy minimum, where TM3 was rotated 20° clockwise from its lowest minimum (at 0°). Whereas most ligands dock very well to the 0° structure, the antagonist NMS, which has a larger head group than acetylcholine and other ligands, does not dock as well in the 0° structure as it did in the 20° structure. Therefore, the 20° structure was selected and reported herein.

**Loop building and optimization of the final model:** The loops were added to the helices by using the loop builder in the MODELER homology modeling program.<sup>[28]</sup> We omitted the amino terminal region before TM1 and carboxyl terminal region after TM7. In addition, the third intracellular loop (154 residues) between TM5 and TM6 was truncated to 25 residues by deleting residues 225–353 (the dLoop mutation) because such large loops are so flexible that they lead to a multiplicity of conformations with similar energies. It has been shown experimentally that the dLoop mutation had little effect on ligand binding affinities<sup>[41]</sup> and should have little effect on the structure of the TM region.

Disulfide linkages were created between Cys 98(TM3)–Cys 178(EC2). We then performed side-chain replacement with SCWRL and optimized the final structure with conjugate gradient minimization of all atoms in the structure.

**Functional prediction of GPCRs: The HierDock protocol:** The ligands used in this study were described with the DREIDING force field and Gasteiger charges.<sup>[42]</sup> To locate their binding sites, we used the HierDock 2.0 procedure,<sup>[8]</sup> a ligand-screening protocol that follows a hierarchical strategy to identify the most favorable ligand-binding conformations. This method has been tested for various GPCRs,<sup>[8–15]</sup> outer membrane protein A,<sup>[43]</sup> and globular proteins.<sup>[44–47]</sup> The HierDock protocol is described in detail in these references. In brief, the parameters used for ligand binding site predictions in M1 are as follows:

**Scanning the entire receptor for binding sites:** We first scanned and docked the five representative ligands (NMS, QNB, 4-DAMP, acetylcholine, and tetramethylammonium) to the TM3 0° structure (see Step 3 energy scan above) to search for binding sites. The void space in the entire receptor was mapped with the SPHGEN DOCK 4.0 utility.<sup>[48]</sup> The entire set of spheres was divided into 27 regions for scanning and locating the binding site of ligands without any prior bias. In each of these regions, we carried out DOCK 4.0 to generate and score 500 conformations, of which 50 were selected by using a buried surface area cutoff of 85% and energy scoring from DOCK 4.0.

In Step 2, these 50 best conformations were subjected to all-atom minimization using the DREIDING force field, keeping the protein fixed, but the ligand movable. The five best-scoring conformations based on the potential energy of the ligand in the protein were selected from each of the 27 regions. We then examined which boxes contained these top five hits and selected the best boxes based on results from all the ligands. We merged the spheres in the top two scoring boxes as the putative binding region for the next step (Figure 7). The same steps were repeated using the same set of ligands for the TM3 20° structure. The scanning procedure identified the same binding site for both structures.

**Optimizing the binding sites:** Once the plausible binding regions were determined, we docked these five representative ligands shown in Figure 6 into the putative binding region using the HierDock 2.0 protocol. The best five bound structures for each ligand–M1 receptor complex were chosen. All five ligands except for NMS docked well with the TM3 0° structure. Therefore, we scanned and docked the antagonist NMS to a second low-energy structure: the TM3 20° structure. The result suggests that NMS fits well in the binding pocket of this structure. Owing to the similarity of the two structures, the coordinates of the other four ligands originally docked in the 0° structure were used as a starting point for further optimization in the 20° structure, as described below. We optimized the binding site using the SCREAM side chain replacement program (V. W.-t. Kam, N. Vaidehi, W. A. Goddard, unpublished data) to reassign side chains for the residues within 4 Å of the ligand. Finally, we performed simulated annealing dynamics to optimize the binding sites, allowing both the receptor side chains and the ligands to move. Specifically, we ran annealed dynamics for a total of 120 ps, cycling between 300 and 700 K for 300 cycles for all docked ligands, allowing the ligands and the side-chain atoms within 5 Å from the ligands to move freely while keeping the main-chain atoms fixed.

The remaining two ligands, pirenzepine and oxotremorine-M, were matched to the optimized QNB and acetylcholine docked structures at the binding site, respectively, because of their structural similarity with these ligands. The matched binding site was further optimized for pirenzepine and oxotremorine-M with the side-chain replacement and simulated annealing dynamics procedure described above.

**Binding energy calculation:** We calculated the binding energy ( $E_b$ ) of each ligand using the equation  $E_b = (E_{p,Protein}) - (E_{p,Solvent})$  as the difference between the potential energy ( $E_p$ ) of the ligand in the protein and the potential energy of the ligand in water. The energy of the ligand in water was calculated by using DREIDING force field and the surface generalized Born continuum solvation method.<sup>[49]</sup>

## Acknowledgements

We thank Rene Trabanino, Huazhang Shen, Victor Wai-tak Kam, Peter Keken-Huskey, John Wendel, Wely Floriano, Peter Freddolino, Julius Su, Yashar Kalani, Youyong Li, Fangqiang Zhu, Jiyoung Heo, Shantanu Sharma, and Chris McClendon for many useful discussions and for use of methods they developed. General support for method development was provided by the National Institutes of Health (R21-MH073910, RO1-GM625523, R29AI40567, and HD365385), and from Sanofi-Aventis and Berlex Biosciences. The facilities of the Materials and Process Simulation Center used in these studies were supported by the Defense University Research Instrumentation Program, the Army Research Office, and the Office of Naval Research. The computational facilities were also supported by the Major Research Instrumentation grant from the National Science Foundation, a Shared University Research Grant from IBM, and grants from Chevron, Intel, and Nissan Corporations.

**Keywords:** cation- $\pi$  interactions • computer chemistry • drug design • GPCRs • muscarinic receptors

- [1] W. S. Messer, Jr., *Curr. Top. Med. Chem.* **2002**, 2, 353–358.
- [2] F. P. Bymaster, D. L. McKinzie, C. C. Felder, J. Wess, *Neurochem. Res.* **2003**, 28, 437–442.
- [3] R. M. Eglén, A. Choppin, N. Watson, *Trends Pharmacol. Sci.* **2001**, 22, 409–414.
- [4] W. Greenlee, J. Clader, T. Asberom, S. McCombie, J. Ford, H. Guzik, J. Kozłowski, S. Li, C. Liu, D. Lowe, S. Vice, H. Zhao, G. Zhou, W. Billard, H. Binch, R. Crosby, R. Duffy, J. Lachowicz, V. Coffin, R. Watkins, V. Ruperto, C. Strader, L. Taylor, K. Cox, *Farmacology* **2001**, 56, 247–250.
- [5] 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.
- [6] C. Bissantz, P. Bernard, M. Hibert, D. Rognan, *Proteins Struct. Funct. Genet.* **2003**, 50, 5–25.
- [7] G. F. X. Schertler, *Eye* **1998**, 12, 504–510.
- [8] N. Vaidehi, W. B. Floriano, R. J. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, G. Zamanakos, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 12622–12627.
- [9] R. J. Trabanino, S. E. Hall, N. Vaidehi, W. B. Floriano, V. W.-t. Kam, W. A. Goddard III, *Biophys. J.* **2004**, 86, 1904–1921.
- [10] 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 III, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 2736–2741.
- [11] 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 III, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 3815–3820.
- [12] W. B. Floriano, N. Vaidehi, W. A. Goddard III, M. S. Singer, G. M. Shepherd, *Proc. Natl. Acad. Sci. USA* **2000**, 97, 10712–10716.
- [13] W. B. Floriano, N. Vaidehi, W. A. Goddard III, *Chem. Senses* **2004**, 29, 269–290.
- [14] S. E. Hall, W. B. Floriano, N. Vaidehi, W. A. Goddard III, *Chem. Senses* **2004**, 29, 595–616.
- [15] P. Hummel, N. Vaidehi, W. B. Floriano, S. E. Hall, W. A. Goddard III, *Protein Sci.* **2005**, 14, 703–710.
- [16] E. C. Hulme, Z. L. Lu, M. S. Bee, *Recept. Channels* **2003**, 9, 215–228.
- [17] S. D. Ward, C. A. Curtis, E. C. Hulme, *Mol. Pharmacol.* **1999**, 56, 1031–1041.
- [18] K. Allman, K. M. Page, C. A. Curtis, E. C. Hulme, *Mol. Pharmacol.* **2000**, 58, 175–184.
- [19] Z. L. Lu, J. M. Saldanha, E. C. Hulme, *J. Biol. Chem.* **2001**, 276, 34098–34104.
- [20] Z. L. Lu, E. C. Hulme, *J. Biol. Chem.* **1999**, 274, 7309–7315.
- [21] H. Bourdon, S. Trumpp-Kallmeyer, H. Schreuder, J. Hoflack, M. Hibert, C. G. Wermuth, *J. Comput.-Aided Mol. Des.* **1997**, 11, 317–332.
- [22] S. Mecozzi, A. P. West, Jr., D. A. Dougherty, *Proc. Natl. Acad. Sci. USA* **1996**, 93, 10566–10571.
- [23] E. C. Hulme, Z. L. Lu, J. W. Saldanha, M. S. Bee, *Biochem. Soc. Trans.* **2003**, 31, 29–34.
- [24] X. P. Huang, P. I. Nagy, F. E. Williams, S. M. Peseckis, W. S. Messer, Jr., *Br. J. Pharmacol.* **1999**, 126, 735–745.
- [25] C. M. Fraser, C. D. Wang, D. A. Robinson, D. G. Jeannine, J. C. Venter, *Mol. Pharmacol.* **1989**, 36, 840–847.
- [26] S. K. Bruley, G. A. Petsko, *Science* **1985**, 229, 23–28.
- [27] O. Guvench, C. L. Brooks III, *J. Am. Chem. Soc.* **2005**, 127, 4668–4674.
- [28] M. A. Marti-Renom, A. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, 29, 291–325.
- [29] S. Deechongkit, P. E. Dawson, J. W. Kelly, *J. Am. Chem. Soc.* **2004**, 126, 16762–16771.
- [30] P. H. N. Celie, S. E. V. Rossum-Fikkert, W. J. V. Dijk, K. Brejc, A. B. Smit, T. K. Sixma, *Neuron* **2004**, 41, 907–914.
- [31] T. A. Spalding, N. J. M. Birdsall, C. A. M. Curtis, E. C. Hulme, *J. Biol. Chem.* **1994**, 269, 4092–4097.
- [32] W. K. Vogel, D. M. Sheehan, M. I. Schimerlik, *Mol. Pharmacol.* **1997**, 52, 1087–1094.
- [33] H. Furukawa, T. Hamada, M. K. Hayashi, T. Haga, Y. Muto, H. Hiroa, S. Yokoyama, K. Nagasawa, M. Ishiguro, *Mol. Pharmacol.* **2002**, 62, 778–787.
- [34] A. C. Tanczos, R. A. Palmer, B. S. Potter, J. W. Saldanha, B. J. Howlin, *Comput. Biol. Chem.* **2004**, 28, 376–385.
- [35] Z. L. Lu, E. C. Hulme, *J. Biol. Chem.* **2000**, 275, 5682–5686.
- [36] S. E. Hall, PhD thesis, California Institute of Technology, Pasadena, CA (USA), **2005**.
- [37] S. L. Mayo, B. D. Olafson, W. A. Goddard III, *J. Phys. Chem.* **1990**, 94, 8897–8909.
- [38] A. D. MacKerell, Jr., D. Bashford, R. L. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kucsera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkerich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wierkiewicz-Kucsera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, 102, 3586–3616.
- [39] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, J. D. Thompson, *Nucleic Acids Res.* **2003**, 31, 3497–3500.
- [40] M. Bower, F. E. Cohen, R. L. Dunbrack, Jr., *J. Mol. Biol.* **1997**, 267, 1268–1282.
- [41] Z. L. Lu, C. A. Curtis, P. G. Jones, J. Pavia, E. C. Hulme, *Mol. Pharmacol.* **1997**, 51, 234–241.
- [42] J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, 36, 3219–3228.
- [43] D. Datta, N. Vaidehi, W. B. Floriano, K. S. Kim, N. V. Prasadara, W. A. Goddard III, *Proteins Struct. Funct. Genet.* **2003**, 50, 213–221.
- [44] D. Datta, N. Vaidehi, X. Xu, W. A. Goddard III, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 2636–2641.
- [45] D. Zhang, N. Vaidehi, W. A. Goddard III, J. F. Danzer, D. Debe, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 6579–6584.
- [46] P. Wang, N. Vaidehi, D. A. Tirrell, W. A. Goddard III, *J. Am. Chem. Soc.* **2002**, 124, 14442–14449.
- [47] P. M. Keken-Huskey, N. Vaidehi, W. B. Floriano, W. A. Goddard III, *J. Phys. Chem. B* **2003**, 107, 11549–11557.
- [48] T. A. Ewing, I. D. Kuntz, *J. Comput. Chem.* **1997**, 18, 1175–1189.
- [49] D. Bashford, D. A. Case, *Annu. Rev. Phys. Chem.* **2000**, 51, 129–152.

Received: February 22, 2006

Published online on August 31, 2006