ISOLATION AND SEQUENCING OF MUSCARINIC ACETYLCHOLINE RECEPTOR GENES FROM FISHES

THESIS

Presented to the Graduate Council

of Texas State University-San Marcos

in Partial Fulfillment

of the Requirements

for the Degree

Master of SCIENCE

By

Prasad V. Phatarpekar, B. Sc., M. Sc.

San Marcos, Texas

August 2004

ACKNOWLEDGEMENTS

First and foremost I would like to thank my mentor Dr. Dana García for her support and guidance. Her encouragement, patience and the intellectual freedom available in her lab made my transition from a non-molecular research background to a molecular one quite enjoyable. I would also like to thank my other committee members, Drs. Joseph Koke and Michael Forstner for their advice and suggestions at every stage of the project. I thank Dr. Forstner for encouraging me to do my best. I would like to thank Dr. Chris Nice for his help in primer design.

Thanks are also due to Ms. Shazia Saleem, a former graduate student in our lab for helping me learn different techniques and for patiently answering my unending questions. I thank Mr. Aaron Marquardt for his help in sequence analysis and his cheerful availability throughout the project. I thank Mr. Chad Copeland and Mr. James Neece for their readiness to help whenever I asked for it. Thanks are also due to my other lab mates for putting up with me. I thank Mr. Zachariah Gompert for his help in employing phylogenetic techniques.

Finally, I would like to thank my family, especially my wife Chhaya, for her support through the ups and downs of this project.

This work was supported by National Science Foundation grants (NSF IBN-0077666 and IBN-0235523) to Dr. D. M. García.

This manuscript was submitted on April 29th 2004.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

ISOLATION AND SEQUENCING OF MUSCARINIC ACETYLCHOLINE RECEPTOR GENES FROM FISHES

By

Prasad V. Phatarpekar, B.Sc., M. Sc.

Texas State University-San Marcos

August 2004

SUPERVISING PROFESSOR: DANA M. GARCÍA

In fish, adaptation to changes in light intensities is achieved in part by pigment granule movements within the retinal pigment epithelium. Based on studies using the cholinergic agonist carbachol it has been suggested that acetylcholine may play a role in light adaptive pigment granule dispersion. The effects of acetylcholine are mediated through the activation of either nicotinic or muscarinic receptors. Muscarinic receptors belong to the superfamily of receptors which initiate intracellular responses by interacting with G-proteins and are characterized by seven transmembrane domains. There are five known subtypes of muscarinic receptors (M_1-M_5). Pharmacological studies performed previously have hinted at the role of odd numbered muscarinic receptors (M_1, M_3 and M_5) in carbachol-induced pigment granule dispersion in retinal pigment epithelium in bluegill (*Lepomis macrochirus*), a model used in our lab to understand the signaling pathways involved in light adaptive pigment granule migration in retinal pigment epithelium. Due to the atypical affinity profiles exhibited by pharmacological agents for non-mammalian

muscarinic receptors, it is necessary to molecularly characterize muscarinic receptors in bluegill in order to establish unambiguously the subtype identities of muscarinic receptors involved in light adaptive pigment granule movement. As a first step in that direction, a fragment of putative M_5 gene from bluegill genomic DNA was isolated and amplified, using polymerase chain reaction employing primers based on the homologous regions among known putative M_5 receptor genes from fish. The fragment is 1385 nucleotides in length and has an open reading frame encoding 461 amino acids. The deduced amino acid sequence showed higher identity to known M_5 receptor proteins than to other subtypes of muscarinic receptors. A putative M_2 receptor coding strand from fugu genomic database was also identified using search tools. This putative gene is 1500 nucleotides in length and encodes 500 amino acids. Comparison of the amino acid sequence with protein databases showed high identity with other M_2 receptor proteins. Both the sequences, the bluegill putative M_5 receptor fragment and fugu putative M_2 receptor, have the critical amino acid motifs conserved across G-protein coupled receptors and the critical amino acids conserved among muscarinic receptors. These motifs have been shown by others to be required for ligand binding and G-protein coupling. Phylogenetic analyses of these putative receptors using both nucleotide and protein alignments and employing different methods grouped these receptors with their respective subtypes, thus confirming their subtype identity.

INTRODUCTION

The vertebrate visual system adjusts to a wide range of light intensities. In mammals, flexible pupil size controls the amount of light that reaches the retina. Fish pupils have fixed diameter and adaptation to changes in light intensities is achieved in part by pigment granule movements within the retinal pigment epithelium (RPE) and photoreceptor movement (see Burnside and Nagle, 1983). RPE is a monolayer of tissue located between the neural retina and the choroid (see Zinn and Marmor, 1979). In light, cone photoreceptors contract, rod photoreceptors elongate and RPE pigment granules disperse into the cells' long apical processes which interdigitate with the photoreceptors. RPE pigment granules protect rod outer segments from bleaching in bright light. In the dark, opposite photoreceptor movements occur and RPE pigment granules aggregate into the cell bodies. Collectively these movements orient the appropriate photoreceptors to their optimum light conditions (see García and Burnside, 1994). From studies using the cholinergic agonist carbachol, García (1998) suggested that acetylcholine might play a role in light adaptive pigment granule dispersion in green sunfish (*Lepomis cyanellus*). The effects of acetylcholine, a neurotransmitter, are mediated through the activation of either ionotropic nicotinic receptors or the metabotropic muscarinic receptors (Caulfield and Birdsall, 1998). There are five different subtypes of muscarinic acetylcholine receptors (M_1-M_5). Gonzalez et al. (2004) reported that carbachol-induced pigment granule dispersion in RPE isolated from bluegill (*Lepomis macrochirus*) is mediated by

odd numbered muscarinic acetylcholine receptors. This conclusion was based on pharmacological evidence wherein the antagonists specific for M_1 and M_3 muscarinic receptors blocked pigment granule dispersion while an agonist specific for M_1 receptor activated dispersion. The agonists and antagonists specific for even numbered muscarinic receptors (M_2 and M_4) failed to induce or inhibit pigment granule dispersion, respectively. However, because subtype specific pharmacological agents, which have been characterized predominantly for mammalian muscarinic receptors, are known to exhibit different affinity profiles for non-mammalian muscarinic receptors (Tietje and Nathanson, 1991; Hsieh and Liao, 2002), Gonzalez's conclusion can still be regarded itself as a hypothesis in need of further testing. Furthermore, the lack of agonists and antagonists having very high selectivity for any particular subtype leaves the pharamacological demonstration of a functional receptor subtype rather incomplete (Caulfield and Birdsall, 1998). Molecular characterization of muscarinic receptors in fish and studies of their expression and function in RPE might help resolve the problem of identification of subtypes involved in pigment granule movement in fish RPE. The focus of my thesis research was to molecularly characterize muscarinic receptors in bluegill (*Lepomis macrochirus*), which we use to study the signaling pathway involved in light adaptive pigment granule migration in RPE. Therefore, although my thesis research does not specifically test any hypotheses (other than the hypothesis that bluegill fish have genes for muscarinic receptors), isolating and sequencing muscarinic receptor genes is an essential first step towards detecting muscarinic receptor expression in fish RPE.

Muscarinic acetylcholine receptors belong to a superfamily of receptors, which initiate intracellular responses by interacting with G-proteins and are broadly

characterized by seven transmembrane segments. The G-protein coupled, seven transmembrane segment receptors comprise the largest superfamily of proteins (Gether, 2000). To date more than 1000 different G-protein coupled receptors have been identified. The G-protein coupled receptor superfamily is divided into three major families. The receptors related to the light receptor rhodopsin and β_2-adrenergic receptor form family A, the receptors related to the glucagon receptor form family B, and the receptors related to the metabotropic neurotransmitter receptors make up family C (Gether, 2000). G-protein coupled receptors do not share any overall sequence homology. The only structural feature shared by all G-protein coupled receptors is the presence of seven, membrane-spanning, α-helical domains connected by alternating intracellular and extracellular loops, with the amino terminus located on the extracellular side and the carboxy terminus on the intracellular side (Figure 1) (Gether, 2000). However, some degree of sequence homology is found within the families. The muscarinic receptors belong to family A. Within this family muscarinic receptors are grouped into a subfamily composed of biogenic amine receptors (Gether, 2000). This subfamily includes α– and ß-adrenergic receptors, dopamine receptors and serotonin receptors (Gether, 2000).

The muscarinic receptors are the predominant cholinergic receptors in the central and peripheral nervous systems. They are found in cardiac and smooth muscle and in various exocrine glands (Peralta et al., 1987a). In the heart, muscarinic receptors regulate the rate and force of contraction (see Hsieh and Liao, 2002). In the central nervous system they are involved in motor control, temperature regulation, cardiovascular regulation, learning and memory, circadian rhythms, formation of ocular dominance columns and seizure activity (Caulfield and Birdsall, 1998; Hsieh and Liao, 2002).

Dysfunction of muscarinic receptor signaling has been implicated in brain disorders such as Alzheimer's disease, Parkinson's and schizophrenia (Birdsall et al., 2001; Flynn et al.,

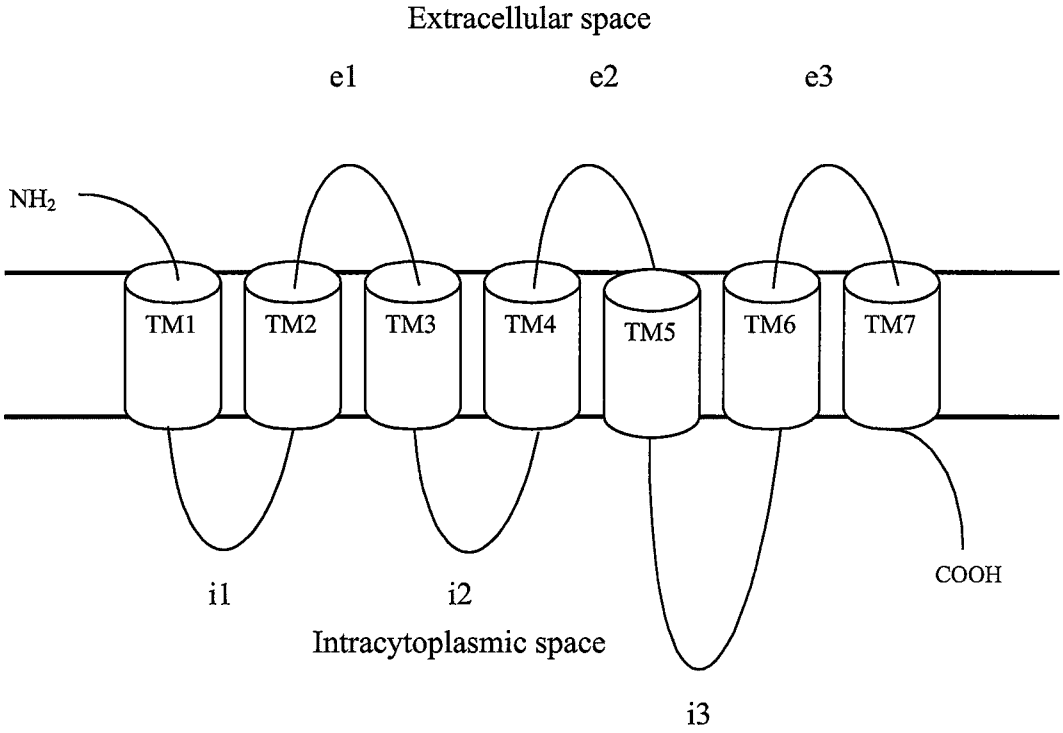


Figure 1: Schematic diagram of the general structure of G-protein coupled receptors. All receptors of this type contain seven transmembrane α-helical domains (TM) linked by alternative intra- (i) and extracellular (e) loops. The third intracytoplasmic loop (i3) is the largest. The amino (NH_2) terminus is extracellular whereas the carboxy (COOH) terminus is intracytoplasmic.

1995; Growdon, 1997). Molecular cloning has identified five different subtypes of muscarinic receptors in mammals, each encoded by a distinct gene lacking introns in the coding region (Bonner et al. 1988). The genes of muscarinic receptor subtypes cloned in other vertebrates have also been found to be intronless (Tietje et al. 1990; Tietje and

Nathanson, 1991; Gadbut and Galper 1994; Creason et al., 2000; Herrera et al., 1994; Hsieh and Liao, 2002). The receptor sequences are highly conserved across mammalian species, at least at the amino acid level (Caulfield and Birdsall, 1998). This sequence conservation is observed in the seven transmembrane domains with much less homology in the amino and carboxy terminal domains. The large third intracytoplasmic loop (i3) between transmembrane domain 5 and transmembrane domain 6 is the most divergent region with length varying among different subtypes and even among same subtype in different species. Even so, the first and last 20 amino acids are partially conserved among some of the subtypes in a manner that correlates with their second messenger preference (Bonner et al., 1987; Bonner, 1989).

The muscarinic receptors are generally divided into two groups, M_{odd} and M_{even}, according to their functional coupling. M_{odd} receptors preferentially couple to pertussis toxin insensitive $G_q/_{11}$ proteins to mediate stimulation of phospholipase C. Upon agonist activation of these subtypes, phospholipase C hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP_2) leading to the formation of inositol 1,4,5-trisphosphate and diacylglycerol. These products act as second messengers by mobilizing Ca^{2+} from intracellular stores and activating protein kinase C respectively (Wess, 1996). M_{even} receptors preferentially couple to G-inhibitory (G_i) protein to mediate inhibition of adenylate cyclase, thereby decreasing cyclic AMP levels. However, the relationship between receptor subtype, G-protein coupling and physiological response is not absolute, further underscoring the necessity of molecular characterization. For example, cloned M_{even} receptors have been shown to be coupled to stimulation of phospholipase C through pertussis toxin-sensitive G-protein (Ashkenazi et al. 1987), and transfected M_{odd} receptors

(M_1) have been shown to inhibit adenylyl cyclase through a pertussis toxin-dependent mechanism (Felder, 1995). In addition to their usual functional coupling, both M_{even} and M_{odd} receptors are known to mediate an increase in intracellular cyclic AMP levels in cells expressing both endogenous and transfected receptors (Baumgold and Fishman, 1988; Shapiro et al., 1988; Peralta et al., 1988; Olianas and Onali, 1991; Jones et al., 1991). M_{even} and M_{odd} receptors also couple to multiple G-proteins (Migeon and Nathanson, 1994; Offermanns et al. 1994). The multiple interactions of a single receptor with various G-proteins depend on the cell surface density of the receptor and on the variation in the amount and nature of G-proteins available to interact with the receptor and the affinity of the receptor for those G-proteins (Eglen and Nahorski, 2000; Felder, 1995).

The promiscuity at the level of G-protein coupling, transduction of multiple signals and cell- and receptor density-dependent mechanisms make it necessary to study signal transduction by the muscarinic receptor subtype under focus by first molecularly characterizing the receptor. In addition, studying its endogenous expression in tissues of interest as well as by transfecting it into cells which lack endogenous expression and employing molecular, immunological and pharmacological techniques enhances our understanding of muscarinic receptor cell biology.

Of the five different subtypes of muscarinic receptors that have been cloned in mammals, all have been found to be intronless. The first muscarinic genes to be cloned were porcine M_1 and M_2 receptors (Kubo et al., 1986 a,b) followed by human M_1-M_5 receptors (Allard et al., 1987; Peralta et al., 1987b; Bonner et al., 1988;) and rat M_1-M_5 receptors (Bonner et al., 1987, 1988; Gocayne et al., 1987). The M_3 muscarinic receptor

has also been cloned from pig (Akiba et al., 1998). Five muscarinic receptor subtypes have also been cloned from guinea pig and mouse (Shapiro et al., 1988; van Koppen et al., 1993). Relatively few muscarinic receptor genes have been cloned from non-mammalian vertebrates and invertebrates. Four subtypes M_2, M_3, M_4 and M_5 have been cloned from chick (Tietje and Nathanson 1991; Gadbut and Galper 1994; Tietje et al., 1990; Creason et al., 2000). Subtype M_4 and subtypes M_2 and M_5 have been cloned in *Xenopus laevis* and zebrafish, respectively (Herrera et al., 1994; Hsieh and Liao, 2002). No muscarinic receptor genes have been cloned from bluegill.

The strategy used for cloning the first two muscarinic receptor genes from pig was screening of complementary DNA libraries with oligonucleotide probes based on the partial amino acid sequence obtained from the receptor purified from cerebral cortex of pig (Kubo et al., 1986a, 1986b). Later rat M_3 and M_4 were cloned by screening a cerebral cortex complementary DNA library using an oligonucleotide probe based on the homologous region between rat M_1 and hamster α_2-adrenergic receptor. Comparison of M_1-M_5 complementary DNAs in human and rat and pig M_2 complementary DNAs with their respective genomic clones indicated that, while there are no introns in the coding region or in the 3' untranslated sequence, there is at least one intron located in the 5' untranslated region (Bonner et al., 1987; Bonner, 1989; Peralta et al., 1987b). Since the initial cloning of these muscarinic receptor genes, the main strategy for cloning muscarinic receptor genes has been to screen genomic libraries. As genes for all the five subtypes were found to be intronless in the coding sequence; the screening of genomic libraries became an appealing strategy as no assumptions had to be made about the tissue distribution, and the problem of obtaining a full length clone was obviated (Hulme et al.,

1990). The probes used for screening genomic or complementary DNA libraries were either full length muscarinic receptor genomic clones (Tietje et al., 1990; Tietje and Nathanson, 1991) or full length muscarinic receptor complementary DNA (Gadbut and Galper, 1994). Probes based on highly conserved amino acid regions present in all five subtypes in mammalian muscarinic receptors were used to clone M_4 receptor from *Xenopus laevis* (Herrera et al., 1994), while Creason et al. (2000) and Hsieh and Liao (2002) used probes which were amplified PCR fragments to screen genomic libraries.

In this thesis I report the results of my efforts to identify bluegill muscarinic receptor genes using a PCR approach. This approach required primers, which were designed based on the sequences of putative zebrafish and fugu muscarinic receptor genes. Fugu putative muscarinic genes were identified from the Fugu Whole Genome Database (http://fugu.hgmp.mrc.ac.uk). During this process a new putative fugu muscarinic receptor gene was identified. The subtype identity of these newly discovered muscarinic receptor genes from fishes was confirmed using phylogenetic analyses. The identification of these genes will permit further studies to detect the expression of muscarinic receptors in fish RPE.

MATERIALS AND METHODS

Isolation of Genomic DNA

Bluegill (*Lepomis macrochirus*) were obtained from Johnson Lake Management (San Marcos, TX). The following procedure was approved by the Texas State Institutional Animal Care and Use Committee (approval # 5JdP9V_01). Prior to killing, fish were anesthetized using MS-222 (Argent Chemical Laboratories Inc., Redmond, WA). Fish were killed by severing the spine. The skin was removed and the muscle tissue along the spine was excised. The tissue was placed in a weigh boat that had been cleaned with ethanol and dried. Approximately 0.5 cubic centimeter of tissue sample was used for genomic DNA extraction by phenol-chloroform-isoamyl alcohol method (Hillis et al., 1996). Alternatively genomic DNA was extracted using DNeasy tissue kit (Qiagen Inc., Valencia, CA) following manufacturer's instructions.

Primer Design

A single contiguous sequence on a single clone, available on the fugu genomics project website (http://fugu.hgmp.mrc.ac.uk; Release 3) was isolated by using a combination of BLAST (Basic Local Alignment Search Tool; Altschul et al., 1990) search algorithms. Zebrafish M_5 coding strand posted at NCBI website (www.ncbi.nlm.nih.gov) by Liao et al. (submitted to NCBI in 2001, but unpublished) was used as a query sequence in a BLASTN search, which compares a nucleotide sequence to a nucleotide sequence database. Several fugu clones were identified, and the one (fugu

clone M000234) showing the highest alignment score was further analyzed. The alignment was examined, and the regions corresponding to the 5' region and the 3' region of the zebrafish coding strand were identified and compared visually against the whole fugu clone. A putative start codon was identified as was an in-frame putative stop codon. The resulting contiguous sequence from the fugu clone was used as a query sequence in a BLASTN search of the NCBI database, and the sequence with the highest identity and similarity was the zebrafish M_5 coding strand. The fugu sequence was also used as a query sequence in a BLASTX search of the NCBI database. BLASTX compares a translated nucleotide query sequence with the protein database (Gish and States, 1993). The inferred protein sequence showed highest identity with the zebrafish M_5 receptor.

Zebrafish and fugu M_5 coding strands were used to search *Tetraodon nigroviridis* whole genome sequence traces available online at NCBI (www.ncbi.nlm.nih.gov/BLAST). Zebrafish and fugu m5 coding strands and homologous whole genome sequence traces of *Tetraodon nigroviridis* were aligned using the program CLUSTAL W available on computer program BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Conserved regions, at least 20 nucleotides long, were selected near 5' and 3' ends of the coding strands to design forward (M5F) and reverse (M5R) primers, respectively (Table 1).

Table 1: Primers used for the amplification of putative muscarinic receptor gene from bluegill genomic DNA

Primer	*5'-3' Sequence*
M5F	CAC AGC CTS TGG GAG GTG ATC
M5R	CAC ATG GGG TTG ACG GTG CTG TTG AC
M5FN1	TAT GCT GGC AAT ACT TCG TAG G
M5RN1	TCA GAG GAG GCA TAA CTG TTG AAG G
M5FN2	GTC AGC CTC ATC ACT ATT GTG G
M5RN2	AAT GTT ATC ACA GGC TCA GAG
M5FN3	GAC AGG AAT CAA GCC TCT TGG TCT TC
M5RN3	CAA GAT AGC ACT GAG AGT CTG AGC TG
M5FN4	GTG TGT GTC CTA CAA GTT CAA GC
M5RN4	AAT GCA GTC TGA GCA GAA GGT G
M5F3end	CAG CTC AGA CTC TCA GTG CTA TCT TG
M5R5end	CAC CAA TGA TGA GGT CAG CAG CTG

M5F and M5R were used initially to amplify a fragment. All other primers except M5F3end and M5R5end were designed to sequence the fragment in forward and reverse directions. M5F3end and M5R5end were used to sequence the 3' and 5' ends of the fragment, respectively.

Amplification

Genomic DNA (gDNA) was amplified by polymerase chain reaction (PCR). A 50 µl PCR reaction had 5 µl of 10X buffer (final concentration 10 mM Tris-HCl, pH 9.0 at 4°C, 50 mM KCl and 0.1% Triton® X-100), 1µl gDNA, 1µl dNTP (10mM each), 3-7µl 25mM $MgCl_2$, 1µl M5F primer (50pmol/ µl), 1µl M5R primers (50pmol/µl) synthesized by Biosynthesis (Lewisville, TX), 0.5µl (2.5 units) *Taq* polymerase (Promega, Madison, WI). The volume was raised to 50µl with distilled water; ingredients were combined in a sterile, 0.5 ml microcentrifuge tube. The reaction was carried out in a Progene thermal cycling machine (Techne, Burlington, NJ) according to following protocol: initial denaturation at 94°C for 1 minute 4 seconds, 40 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 1 minute 30 seconds followed by final extension at 72°C for 5 minutes and stored at 4°C. The PCR product was subjected to electrophoresis on a 1% agarose gel at 120 V for 30 minutes, stained with ethidium bromide and then viewed under UV light to verify the presence of amplified product. A picture was obtained using Nikon COOLPIX 990 digital camera with a UV filter and was processed with Adobe Photoshop version 5.5 software (Adobe Systems Inc., San Jose, CA). The PCR product was purified by removing free nucleotides and primers using Nucleospin PCR purification kit (Qiagen, Valencia, CA) following manufacturer's instructions, or the PCR product was sent to North Woods DNA, Inc. (Becida, MN) for gel purification when multiple bands were detected in the product.

Automated DNA Sequencing

Cycle sequencing reactions were performed with an Applied Biosystems Big Dye V 3.0 kit (ABI, Foster City, CA). The cycle sequencing reaction contained 3.5µl

template, 0.5 µl primer (either forward or reverse), 3µl Big Dye and 2µl distilled water; these ingredients were combined in a sterile, 0.5 ml microcentrifuge tube. The reaction was subjected to thermal cycling in a Progene thermal cycling machine (Techne, Burlington, NJ) according to the following protocol: 25 cycles of denaturation at 96 °C for 10 seconds, annealing at 50°C for 5 seconds and extension at 60°C for 4 minutes. Following cycle sequencing reactions, cycles sequencing products were separated from unused primers and dye using Sephadex G-50 column and centrifuging in an Eppendorf table top centrifuge. The products were dried using a Savant speed vac vacuum centrifuge (Thermo, Walttham, MA) and then were loaded onto a polyacrylamide gel and electrophoresed on an ABI model 377 sequencer. Alternatively some samples of the PCR product and primers were sent to North Woods DNA, Inc. (Becida, MN) for gel purification and sequencing.

Internal Primers

The sequences obtained from North Woods DNA, Inc. (Becida, MN) using forward and reverse primers were analyzed using Sequencher 4.1 software (Gene Codes Corp., Ann Arbor, MI). These two non-contiguous sequences were aligned with the fugu putative M_5 receptor gene using the same software. To obtain the sequence of the non-overlapping region, forward (M5FN1) and reverse (M5RN1) primers were designed based on the sequences of the forward and reverse strands, respectively. The PCR product and the primers were sent to North Woods DNA Inc. for sequencing. The sequences obtained using M5F, M5R, M5FN1 and M5RN1 were analyzed using Sequencher 4.1 to form a contiguous sequence, 1304 nucleotides in length. This sequence was used to design four more pairs of primers, namely M5FN2, M5RN2, M5FN3,

M5RN3, M5FN4, M5RN4, M5F3end and M5R5end (Table 1). The first three pairs of primers were designed to sequence the amplified fragment in forward and reverse directions while the last pair of primers was designed to sequence the 3' and 5' ends of the fragment, respectively. All primers were synthesized by Biosynthesis (Lewisville, TX). The PCR products and the primers were sent to North Woods DNA Inc. for sequencing. Final sequences from all the fragments were analyzed and assembled using Sequencher 4.1 to form a contiguous sequence.

Comparison with Known Sequences

The nucleotide sequence was compared with the nucleotide database using BLAST tools such as BLASTX (comparing translated nucleotide query sequence with protein database) and TBLASTX (comparing the six translation frames of a nucleotide query sequence against the six frame translations of nucleotide sequence database) available at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). The nucleotide sequence was translated to amino acids using Translate Tool software available on 123 Genomics website (http://us.expasy.org/tools/dna.html). The deduced amino acid sequence was compared with the protein database using BLAST tools such as TBLASTN (comparing a protein query against translated nucleotide sequence database) and BLASTP (comparing a protein query against protein database). The deduced amino acid sequence was compared with known muscarinic receptor sequences using the computer program BL2SEQ (BLAST 2 sequences) (Tatusova and Madden, 1999) available at NCBI (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html) to calculate percent identity and percent similarity. Percent identity of the deduced amino acid sequence was calculated by dividing the number of identical residues by the total length of the deduced amino acid

query sequence, while percent similarity was calculated by dividing number of identities and conservative substitutions by the total length of the deduced amino acid sequence of the query sequence. The putative transmembrane domains and intra- and extracellular loops were predicted using software available on 123 Genomics website (http://www.cbs.dtu.dk/services/TMHMM-2.0). The seven transmembrane domains and the intra- and extracellular loops were compared with the corresponding domains of known vertebrate muscarinic receptors, and percent identity and percent similarity were calculated as explained above.

Fugu Genomic Database Search

The Fugu Genomic Database, available on the fugu genomics project website, was searched using zebrafish M_2 coding strand (Hsieh and Liao, 2002) employing BLASTN. Based on the alignment produced by BLASTN, the fugu clone showing the highest alignment score was further analyzed. The results of the BLASTX of the region of this clone, which showed high identity with the query sequence, that are available on fugu genomics project website were examined to identify fugu putative M_2 coding strand. The nucleotide sequence was compared with the nucleotide database as mentioned before. The putative fugu M_2 coding strand was translated and compared with the protein database and with known muscarinic receptors and identity and similarity were calculated as explained previously. The domain prediction and comparison of domains with corresponding domains of known muscarinic receptors was performed as mentioned before.

Phylogenetic Analyses

Amino acid sequences of different muscarinic receptor subtypes from all available taxa were downloaded from the protein database available at the NCBI website (Table 2). Sequences of coding strands (nucleotide) of muscarinic receptors were downloaded in FASTA format from NCBI GenBank (Table 3). Amino acid sequence alignment was performed with Clustal X (Thompson et al., 1997). Many different alignments were performed with various settings for gap opening and gap extension penalties for pair-wise and multiple alignment parameters. Each resulting alignment was assessed visually. The criterion used in deciding the optimum alignment was the presence of identical and similar amino acids in the region starting from about the middle of the first transmembrane domain to the amino terminal region of the third intracytoplasmic loop (i3) and again from the carboxy terminal region of the i3 loop to the carboxy terminal region of the protein. These regions are conserved across subtypes (Bonner, 1989). The other criterion was the perfect alignment of motifs conserved across all muscarinic receptors. Using this method the optimum alignment was obtained with gap opening penalty of 52 and gap extension penalty of 1.25 for pair-wise alignment parameters and gap opening penalty of 22 and gap extension penalty of 0.45 for multiple alignment parameters.

Table 2: Name, species and the NCBI accession number of muscarinic receptor proteins used in phylogenetic analyses.

Name	Species	Accession No.	Name	Species	Accession No.
HsM1	*Homo sapiens*	NP_000729	MmM3	*Mus musculus*	NP_150372
Mmul1	*Macaca mulatta*	AAB95157	CpM3	*Cavia porcellus*	AAL67911
SsM1	*Sus scrofa*	CAA28003	GgalM3	*Gallus gallus*	AAA65961
RnM1	*Rattus norvegicus*	AAB20705	HsM4	*Homo sapiens*	NP_000732
MmM1	*Mus musculus*	NP_031724	MmM4	*Mus muscles*	NP_031725
CpM1	*Cavia porcellus*	AAL67909	CpM4	*Cavia porcellus*	AAL67912
HsM2	*Homo sapiens*	NP_000730	GgalM4	*Gallus gallus*	AAA48563
SsM2	*Sus scrofa*	CAA28413	XlM4	*Xenopus laevis*	CAA46694
RnM2	*Rattus norvegicus*	NP_112278	HsM5	*Homo sapiens*	NP_036257
MmM2	*Mus musculus*	AAG14343	MmulM5	*Macaca mulatta*	AAB95159
CpM2	*Cavia porcellus*	AAL67910	RnM5	*Rattus norvegicus*	AAA40658
GgalM2	*Gallus gallus*	AAB04106	MmM5	*Mus musculus*	AAL26028
DrM2	*Danio rerio*	AAK93793	CpM5	*Cavia porcellus*	AAL67913
TrM2	*Takifugu rubripes*	NA	GgalM5	*Gallus gallus*	AAF19027
HsM3	*Homo sapiens*	AAM18940	DrM5	*Danio rerio*	AAK93794
PpM3	*Pongo pygmaeus*	BAA94483	TrM5	*Takifugu rubripes*	NA
PtM3	*Pan troglodytes*	BAA94481	LmM5	*Lepomis macrochirus*	NA
GgM3	*Gorilla gorilla*	BAA94482	DmM	*Drosophila melanogaster*	NP_523844
SsM3	*Sus scrofa*	CAA31215	CeM	*Caenorhabditis elegans*	AAD48771
RnM3	*Rattus norvegicus*	NP_036659			

M1, M2, M3, M4 and M5 refer to the muscarinic receptor subtypes1, 2, 3, 4 and 5, respectively NA=not available in NCBI database. Sequences denoted as NA were isolated and identified in the present study

Table 3: Name, species and the Genbank accession number of muscarinic receptor genes (coding strands) used in phylogenetic analyses.

Name	Species	Accession No.	Name	Species	Accession No.
HsM1	*Homo sapiens*	NM_000738	MmM3	*Mus musculus*	NM_033269
Mmul1	*Macaca mulatta*	AF026262	CpM3	*Cavia porcellus*	AY072060
SsM1	*Sus scrofa*	X04413	GgalM3	*Gallus gallus*	L10617
RnM1	*Rattus norvegicus*	S73971	HsM4	*Homo sapiens*	NM_000741
MmM1	*Mus musculus*	NM_007698	MmM4	*Mus musculus*	NM_007699
CpM1	*Cavia porcellus*	AY072058	CpM4	*Cavia porcellus*	AY072061
HsM2	*Homo sapiens*	NM_000739	GgalM4	*Gallus gallus*	J05218
SsM2	*Sus scrofa*	X04708	XlM4	*Xenopus laevis*	X65865
RnM2	*Rattus norvegicus*	NM_031016	HsM5	*Homo sapiens*	NM_012125
MmM2	*Mus musculus*	AF264049	MmulM5	*Macaca mulatta*	AF026264
CpM2	*Cavia porcellus*	AY072059	RnM5	*Rattus norvegicus*	M22926
GgalM2	*Gallus gallus*	M73217	MmM5	*Mus musculus*	AF264051
DrM2	*Danio rerio*	AY039653	CpM5	*Cavia porcellus*	AY072062
TrM2	*Takifugu rubripes*	NA	GgalM5	*Gallus gallus*	AF201960
HsM3	*Homo sapiens*	AF498917	DrM5	*Danio rerio*	AY039654
PpM3	*Pongo pygmaeus*	AB041398	TrM5	*Takifugu rubripes*	NA
PtM3	*Pan troglodytes*	AB041396	LmM5	*Lepomis macrochirus*	NA
GgM3	*Gorilla gorilla*	AB041397	DmM	*Drosophila melanogaster*	NM_079120
SsM3	*Sus scrofa*	X12712	CeM	*Caenorhabditis elegans*	AF139093
RnM3	*Rattus norvegicus*	NM_012527			

M1, M2, M3, M4 and M5 refer to the muscarinic receptor subtypes1, 2, 3, 4 and 5, respectively. NA=not available in NCBI database. Sequences denoted as NA were isolated and identified in the present study.

Protein alignment was used to obtain nucleotide alignment using program CodonAlign (http://www.sinauer.com/hall/), which creates a DNA alignment based on alignment of the corresponding proteins. It introduces into each DNA sequence a triplet gap at the position of each gap in the aligned protein sequence (Hall, 2001). The protein and DNA alignment files were executed in PAUP* (Swofford, 2002). Under the distance criterion an unrooted phylogram was obtained using both the alignments. In the resulting trees all the muscarinic receptors formed a monophyletic network except *Drosophila melanogaster* and *Caenorhabditis elegans* muscarinic receptors, which formed a separate monophyletic group (Figure 2). For this reason these two muscarinic receptors were set as an outgroup.

Tree topologies were obtained under the parsimony criterion with heuristic search with 10,000 replications of stepwise additions using protein as well as DNA alignment. One hundred bootstrap replications with 2500 heuristic, random stepwise additions were performed to determine statistical support levels. Maximum likelihood analysis of protein alignment was set up using Puzzle 5.0 (Schmidt et al. 2002). JTT substitution frequency matrix (Jones et al. 1992) with amino acid usage estimated from the data and uniform rate for site-to-site variation were employed to infer the phylogenetic tree. Ten thousand quartet-puzzling steps were run to obtain support values for each node.

A tree was obtained using DNA alignment under the distance criterion using neighbor joining method (NJ). Tamura-Nei distance correction was used. Tamura-Nei was used as the sequences had different GC content and different transition/transversion ratios and Tamura-Nei corrects for both biases (Nei and Kumar, 2000). In neighbor-joining analyses, bootstrap replications were performed 2500 times.

Modeltest 3.06 (Posada and Crandall, 1998) was used to select the model of evolution for Bayesian and maximum likelihood analysis using DNA alignment. The general time reversible + I (invariant sites) + G (gamma distribution) model was selected for having the highest log likelihood. Maximum likelihood analysis was set up in PAUP* using parameter estimates obtained in Modeltest results. One hundred bootstrap replications with one heuristic, random stepwise addition were performed to determine statistical support levels. The Bayesian analysis was performed using MrBayes 3.0 (Huelsenbeck and Ronquist, 2001). *Caenorhabditis elegans* muscarinic receptor was selected as an outgroup for the analysis. The initial setting included Markov Chain Monte Carlo search which was set to run one thousand generations with a sample frequency of one hundred. Based on the time required to run one thousand generations, another run was set up to run for fifteen minutes. The runs were repeated with increasing number of generations till the sum of the log likelihoods of trees converged to a stable value. Based on the number of generations taken to stabilize the log likelihood value, a final run was set up in which the number of generations were twenty times the number of generations taken to stabilize the sum of log likelihood values of the trees. The final setting included Markov Chain Monte Carlo search set to run four hundred thousand generations with a sample frequency of one hundred, and burnin, the number of trees that would be ignored while the consensus tree was created, was set to 0.1 times the number of trees (400). The tree file produced in MrBayes was opened in PAUP* and a majority consensus tree was constructed.

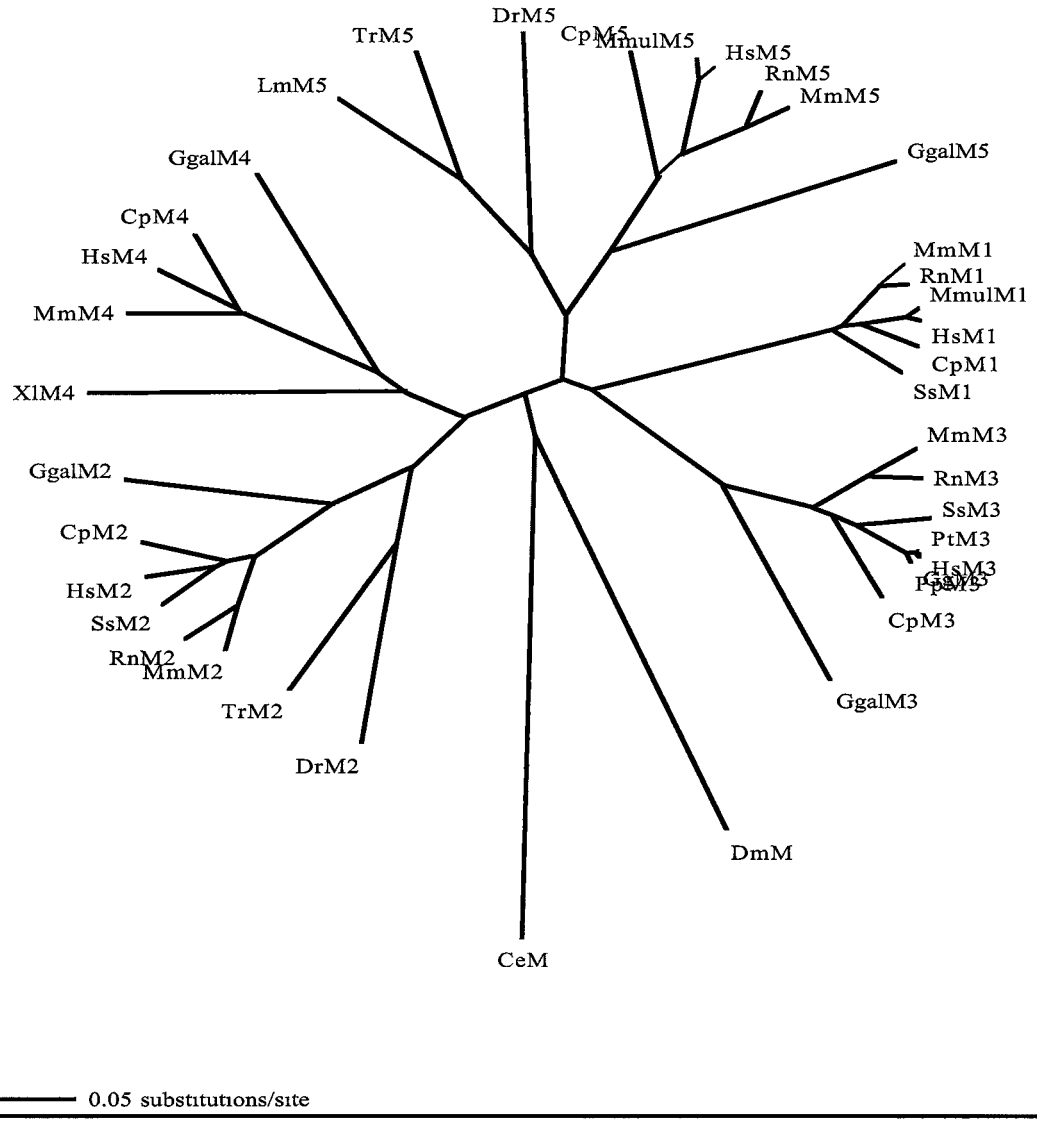


Figure 2: Unrooted phylogram obtained using PAUP* under distance criterion using nucleotide alignment of muscarinic receptors. All the muscarinic receptors formed a monophyletic network except *Drosophila melanogaster* (DmM) and *Caenorhabditis elegans* (CeM) muscarinic receptors, which formed a separate monophyletic group. For this reason these two muscarinic receptors were set as an outgroup in subsequent phylogenetic analyses.

RESULTS

Identification of Bluegill M_5 Gene Fragment

Using the primers M5F and M5R an ~1400 bp fragment from bluegill genomic DNA was amplified (Figure 3). Sequencing this fragment using M5F and M5R generated two sequences, which were non-contiguous and together 1100 nucleotides in length. Based on the sequence of forward and reverse strands, primers M5F1 and M5R1 were designed to sequence the non-overlapping region. The sequences obtained using all these primers yielded a 1304 bp long contiguous sequence. This sequence was used to design primers which enabled me to sequence the ~ 1400 bp product in forward and reverse directions as well as the 5' and 3' ends of the product. The sequences thus obtained generated a 1385 bp contiguous sequence. The positions of the primers and sequence fragments relative to the contiguous sequence obtained are shown in Figures 4 and 5 respectively. Results of BLASTX and TBLASTX showed that among all the available sequences with which it was compared, this sequence showed the highest homology with muscarinic receptor proteins and muscarinic receptor genes, respectively. Among muscarinic receptors, the highest homology was observed for the M_5 subtype.

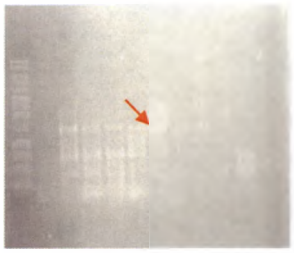


Figure 3: Gel image of ~1400 bp PCR product obtained with primers M5F and M5R. The product is indicated with an arrow.

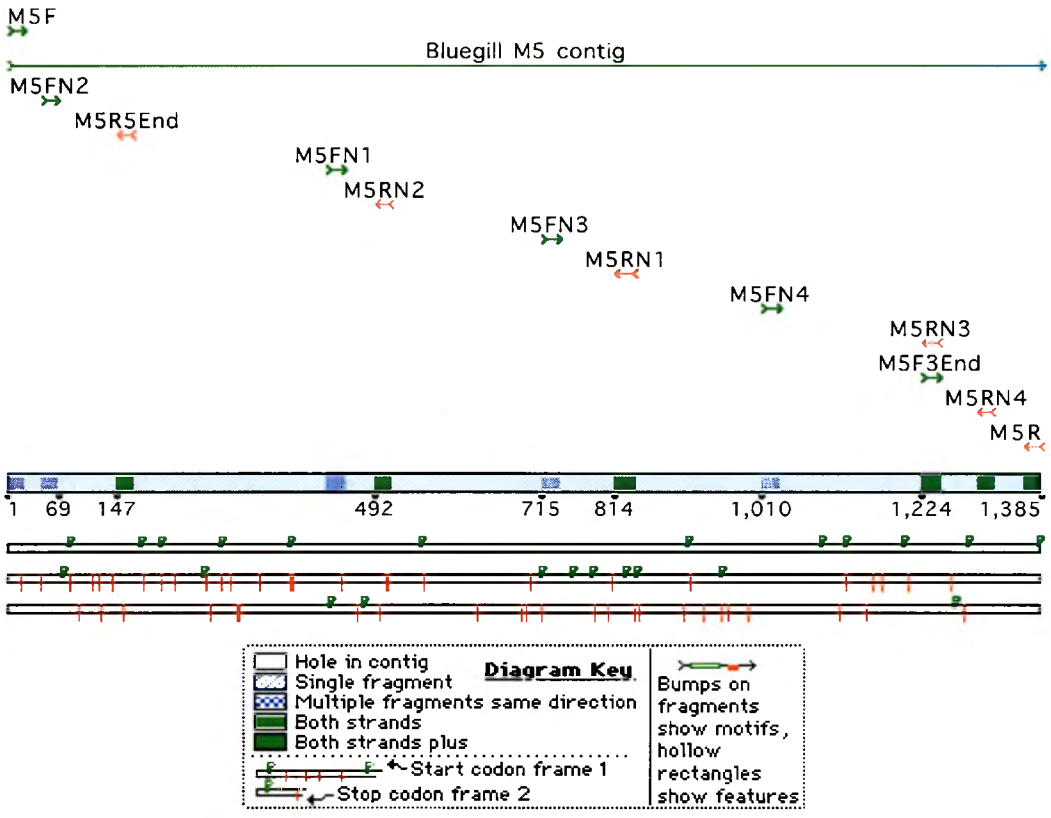


Figure 4: Positions of primers relative to contiguous sequence of bluegill M_5 gene fragment.

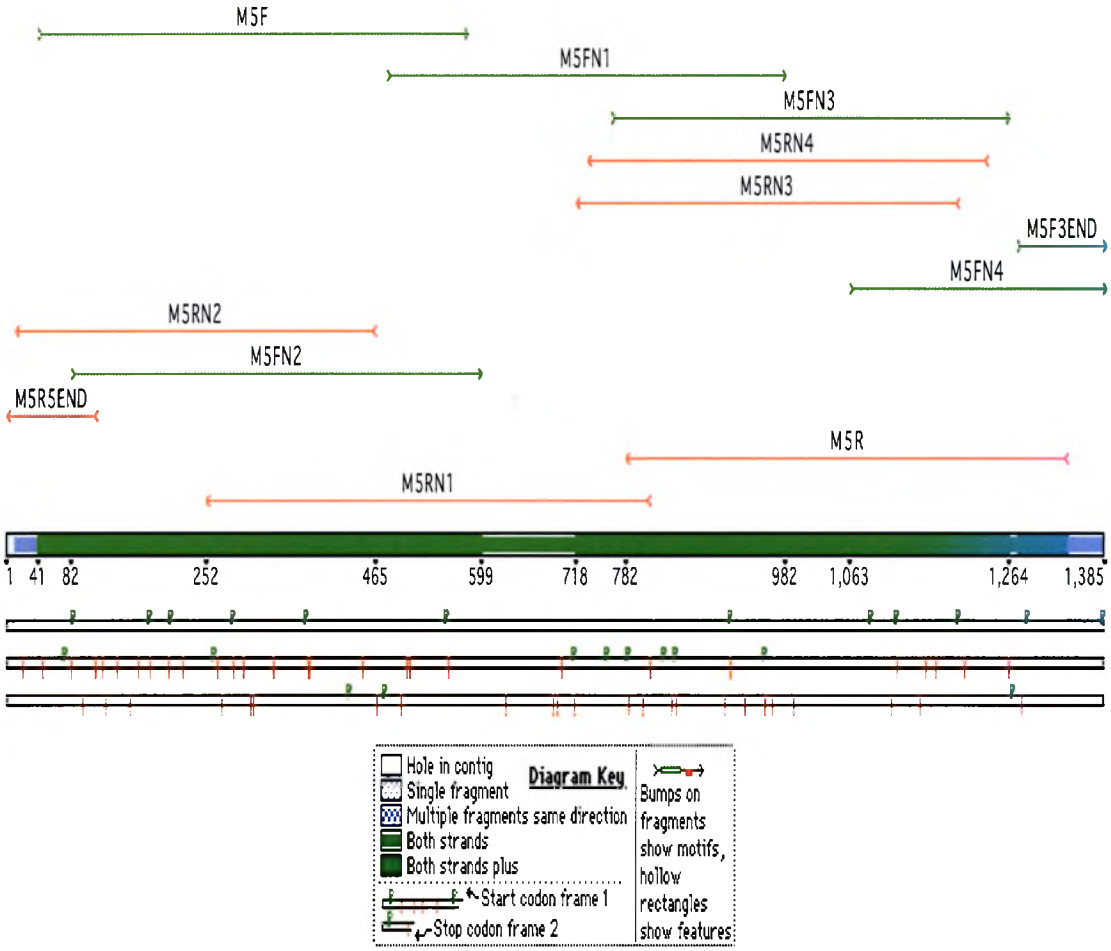


Figure 5: Position of PCR fragments obtained using primers listed in Table 1. These fragments were used to assemble the contiguous sequence of bluegill M_5 gene fragment (shaded bar).

Deduced Amino Acid Sequence

The 1385 bp long contiguous sequence was translated in six reading frames and was found to have a single, 5'-3' frame 1, open reading frame encoding 461 amino acids (Figure 6). When compared against protein databases, the deduced amino acid sequence showed the highest homology with muscarinic receptor proteins. Among muscarinic receptors, the highest homology was observed for M_5 subtype. The deduced amino acid sequence showed higher identity and similarity to the M_5 receptor proteins in fish than to other vertebrate M_5 receptors (Table 4). An alignment of bluegill M_5 receptor fragment with other known M_5 receptors is shown in Figure 7.

```
CACAGCCTGTGGGAGGTGATCACCATTGCGACTGTGTCAGCTATAGTCAGCCTCATCACTATT 63
 H  S  L  W  E  V  I  T  I  A  T  V  S  A  I  V  S  L  I  T  I  21
GTGGGGAATGTCCTGGTGATGCTCTCCTTTAAAGTCAACAGCCAGCTAAAGACAGTGAATAAT 126
 V  G  N  V  L  V  M  L  S  F  K  V  N  S  Q  L  K  T  V  N  N  42
TACTACCTGCTGAGTCTGGCAGCTGCTGACCTCATCATAGGTGTTTTCTCCATGAATCTGTAT 189
 Y  Y  L  L  S  L  A  A  A  D  L  I  I  G  V  F  S  M  N  L  Y  63
ACCTCTTACATACTGATGGGCTACTGGGCCTTAGGAAACCTCGCCTGCGATCTGTGGTTGGCG 252
 T  S  Y  I  L  M  G  Y  W  A  L  G  N  L  A  C  D  L  W  L  A  84
GTGGACTATGTAGCCAGTAACGCCTCAGTCATGAACCTGTTGGTAATCAGTTTTGATAGATAT 315
 V  D  Y  V  A  S  N  A  S  V  M  N  L  L  V  I  S  F  D  R  Y  105
TTTTCCATCACCAGACCTCTGACCTACAGGGCCAAACGGACTCCCAAACGAGCTGGGATCATG 378
 F  S  I  T  R  P  L  T  Y  R  A  K  R  T  P  K  R  A  G  I  M  126
ATAGGTTTGGCCTGGCTGGTTTCACTTATCCTTTGGGCGCCCCCTATTCTATGCTGGCAATAC 441
 I  G  L  A  W  L  V  S  L  I  L  W  A  P  P  I  L  C  W  Q  Y  147
TTCGTAGGAAAAAGGACTGTCCCTGAGAGGCAATGCCAGATCCAGTTTTTCTCTGAGCCTGTG 504
 F  V  G  K  R  T  V  P  E  R  Q  C  Q  I  Q  F  F  S  E  P  V  168
ATAACATTTGGGACAGCGATTGCTGCCTTTTATATCCCTGTATCTGTCATGACAATCCTATAC 567
 I  T  F  G  T  A  I  A  A  F  Y  I  P  V  S  V  M  T  I  L  Y  189
TGTCGAATCTACAAGGAGACAGAGAAGAGGACCAAAGATCTGGCGGAGCTGCAGGGGATTAAC 630
 C  R  I  Y  K  E  T  E  K  R  T  K  D  L  A  E  L  Q  G  I  N  210
TATCCCACAGAACCTGGGGTCACCCAGCCTCAGAAGACCATTATCAGATCTTGTTTTAGCTGT 693
 Y  P  T  E  P  G  V  T  Q  P  Q  K  T  I  I  R  S  C  F  S  C  231
AAGTTAAGGTCAGCTTCAAATGACAGGAATCAAGCCTCTTGGTCTTCCTCCAGCAGAAGCAAT 756
 K  L  R  S  A  S  N  D  R  N  Q  A  S  W  S  S  S  S  R  S  N  252
GCTGCCAAATCAGCAGCCACCACCAATGACGAGTGGTCCAAAGCTAGTCAGCTGACCACCTTC 819
 A  A  K  S  A  A  T  T  N  D  E  W  S  K  A  S  Q  L  T  T  F  273
AACAGTTATGCCTCCTCTGAGGATGAGGACAGGCCTGTGTCTCCAGGGGGATTCCAGGTGCCC 882
 N  S  Y  A  S  S  E  D  E  D  R  P  V  S  P  G  G  F  Q  V  P  294
TCTTTCAGGAACCAGGCTTGTGAGACCATGAAGACTGGGGTGGGCAGTGAGAACGAGCAGCTC 945
 S  F  R  N  Q  A  C  E  T  M  K  T  G  V  G  S  E  N  E  Q  L  315
AGCAGCTATGAAGAGGATAGCTTCTTCCAGACACCACCCAAAAGTAACTCTCAGAGGAGCAAC 1008
 S  S  Y  E  E  D  S  F  F  Q  T  P  P  K  S  N  S  Q  R  S  N  336
AAGTGTGTGTCCTACAAGTTCAAGCCTGTGGCCAAGGACACGCACGTGGAGCACCACAGCAAA 1071
 K  C  V  S  Y  K  F  K  P  V  A  K  D  T  H  V  E  H  H  S  K  357
AACGGAGACACCAAAATGGCTTCGTCCACGTTCTCCTCGGCTGAGTCCATGAGCGTTCCATCC 1134
 N  G  D  T  K  M  A  S  S  T  F  S  S  A  E  S  M  S  V  P  S  378
ACCTCGTCAACATCTAAGCCCATAGACGCCACGCTGAAGAACCAGATCACCAAGAGGAAGCGG 1197
 T  S  S  T  S  K  P  I  D  A  T  L  K  N  Q  I  T  K  R  K  R  399
ATGGTGCTGATCAAGGAGAGGAAGGCAGCTCAGACTCTCAGTGCTATCTTGCTGGCCTTCATC 1260
 M  V  L  I  K  E  R  K  A  A  Q  T  L  S  A  I  L  L  A  F  I  420
CTAACATGGACGCCTTATAACATCATGGTGCTTATTTCCACCTTCTGCTCAGACTGCATTCCC 1323
 L  T  W  T  P  Y  N  I  M  V  L  I  S  T  F  C  S  D  C  I  P  441
CTCTCGCTCTGGCATTTGGGCTACTGGCTGTGCTACGTCAACAGCACCGTCAACCCCATGTG  1385
 L  S  L  W  H  L  G  Y  W  L  C  Y  V  N  S  T  V  N  P  M     461
```

Figure 6: Nucleotide sequence of bluegill putative M_5 gene fragment and the deduced amino acid sequence. The nucleotide and amino acid sequences are 1385 and 461 in length respectively.

Table 4: Comparison of muscarinic receptor subtypes of other vertebrates with bluegill putative M_5 receptor fragment and fugu putative M_2 receptor at amino acid level.

Organism	Subtype	Bluegill M_5	Fugu M_2
Fugu	M_2	41(55)	-
	M_5	82(84)	42(55)
Zebrafish	M_2	41(56)	76(80)
	M_5	72(76)	42(56)
Chick	M_2	42(56)	65(72)
	M_3	50(64)	42(58)
	M_4	43(55)	57(66)
	M_5	63(72)	41(57)
Human	M_1	49(61)	40(54)
	M_2	42(57)	63(69)
	M_3	51(64	42(55)
	M_4	42(54)	56(65)
	M_5	61(71)	44(59)

Numbers represent percent identity and percent similarity (in parenthesis). Genbank accession numbers are NP_840086 (zebrafish M_2), AAK93794 (zebrafish M_5), P30372 (chick M_2), NP_990730 (chick M_3), A35546 (chick M_4), AAF19027 (chick M_5), NP_000729 (human M_1), P08172 (human M_2), NP_000731 (human M_3), NP_000732 (human M_4) and P08912 (human M_5). Fugu putative M_2 and M_5 were identified in the present study.

Figure 7: Alignment of deduced amino acid sequence of bluegill putative M5 receptor fragment with known vertebrate M5 receptors using the program Clustal X (Thompson et al. 1997). The names of the sequences are prefixed with Hs (*Homo sapiens*, human), Mmul (*Macaca mulatta*, rhesus monkey), Mm (*Mus musculus*, mouse), Rn (*Rattus norvegicus*, rat), Cp (*Cavia porcellus*, guinea pig), Ggal (*Gallus gallus*, chick), Dr (*Danio rerio*, zebrafish), Tr (Takifugu rubripes, fugu) and Lm (*Lepomis macrochirus*, bluegill). Genbank accession numbers for these sequences are NP_036257 (HsM5), AAB95159 (MmulM5), AAL26028 (MmM5), AAA40658 (RnM5), AAL67913 (CpM5), AAF19027 (GgalM5), AAK93794 (DrM5). TrM5 and LmM5 represent fugu putative M5 receptor and bluegill putative M5 receptor fragment respectively. Residues identical to that of human M5 are indicated by dots. Dashes in the alignment represent gaps inserted. Transmembrane domains (TM) of bluegill putative M5 receptor fragment are delineated by dashes below the sequences. Amino acid motif and residues which are either conserved across GPCR family or only in muscarinic and other closely related receptors are shown in bold, colored fonts. The conserved aspartic acid-arginine-tyrosine (DRY) motif is boxed. The conserved motif and residues are critical for muscarinic receptor function. For details see text.

```
HsM5       MEGDSYHNATTVNGTPVNHQPLERHRLWEVITIAAVTAVVSLITIVGNVLVMISFKVNSQ   [60]
MmulM5     ................Y.............S.............................   [60]
MmM5       ...E....E...........A....G.................M.................   [60]
RnM5       ...E..N-ES..........A....G........V.......M.................   [59]
CpM5       ...EA.P......P..G......Y.QV...V.....S.......................   [60]
GgalM5     ..VNLFS.S.VT.SSSI..KQ..G.S.................................   [60]
DrM5       -----MGVEKLTLRANITEVQ.VT.S.....A..T.S.I..F...I..I...L.......   [55]
TrM5       --MEEASAVNSTSN.SSADTH.VT.S........T.S.I.............L.......   [58]
LmM5       ------------------------.S........T.S.I.............L.......   [36]
                                    ----------TM1----------

HsM5       LKTVNNYYLLSLACADLIIGIFSMNLYTTYILMGRWALGSLACDLWLALDYVASNASVMN   [120]
MmulM5     ....................................................................   [120]
MmM5       ...........................................V........................   [120]
RnM5       ...........................................V........................   [119]
CpM5       .........F..........V.....................................................   [120]
GgalM5     ..............................S...I.H.S.....................   [120]
DrM5       .............F......V.......S.....Y.....................   [115]
TrM5       .............A......V.......S.....Y....N.S.....................   [118]
LmM5       .............A......V.......S.....Y....N........V.....   [96]
               ----------TM2----------                    ---------TM3--

HsM5       LLVISFDRYFSITRPLTYRAKRTPKRAGIMIGLAWLISFILWAPAILCWQYLVGKRTVPL   [180]
MmulM5     ..............................V.....................................   [180]
MmM5       ...........................................V......................P   [180]
RnM5       ...........................................V......................P   [179]
CpM5       ..............................L......V......................P   [180]
GgalM5     ..............................................V......F..E....P   [180]
DrM5       ........................I..........V.......P......F.......E   [175]
TrM5       ...........................................V.LV....P......F.......E   [178]
LmM5       ...........................................V.L.....P......F.......E   [156]
           ------                        ----------TM4----------

HsM5       DECQIQFLSEPTITFGTAIAAFYIPVSVMTILYCRIYRETEKRTKDLADLQGSDSVTKAE   [240]
MmulM5     ..........................................................E..   [240]
MmM5       ....................................................AEVK   [240]
RnM5       ....................................................AE.K   [239]
CpM5       ........................V.........................E.......AE.A   [240]
GgalM5     E.......Y..............................K..........E....E..AEF.   [240]
DrM5       RQ.....F...V...........F..........L..K...R......E...IN.S.NSS   [235]
TrM5       RQ.....F...V.............................K..........E...INYT.DSV   [238]
LmM5       RQ.....F...V.............................K..........E...INYP.EPG   [216]
                    ----------TM5----------

HsM5       KRKPAHRALFRSCLRCPRPTLAQRERNQASWSSSRRSTSTTGKPSQATGPSANWAKAEQL   [300]
MmulM5     ............................T........A..S.......D..T.Q......   [300]
MmM5       .......T.L..FFS....S.........................T...DL..D.E....V   [300]
RnM5       ..E..Q.T.L..FFS....S.........................TT...DL..D.E....V   [299]
CpM5       ..R..P...L...FS..H...V....S.........K....S.T......H..H...VG..   [300]
GgalM5     TI..RK-.FLK..FS.KQQN.VK...C.......S....A.V.A...ASTCTE....D..   [299]
DrM5       G--D.QPQKI...FG.KHVSN--------------T.RTQA-----LSHTN.AKTLDD..   [274]
TrM5       VP-QPQKTII...FN.QVNLASRDTNQTSWS....SNAAKS-----VACANDD.S..D..   [292]
LmM5       VT-QPQKTII...FS.KLRSASNDRNQASWS....SNAAKS-----.ATTNDE.S..S..   [270]

HsM5       TTCSSYPSSEDEDKPATDPVLQVVYKSQGKESPGEEFSAEETEETFVKAETEKSDYDTPN   [360]
MmulM5     ............................R..........S.DA.......Q...H.S....   [360]
MmM5       .N...C......A.AT....F.....NEA.....K..NTQ..K....SPR..NN.....K   [360]
RnM5       ............A..T....F.M....EA.....K.SNTQ..K..V.NTR..N......K   [359]
CpM5       A.G.R..VP...AQ.PA...F.A..N..A....R..L..K.AK.KA...QR..T...I.K   [360]
GgalM5     ......A...E...L.....F..A.R.PS.-GKE...NER.SQDIV..EQP.EN.FE.QQ   [358]
DrM5       .NFN..A...E..RSGN-----------------------FQ.SCRHQ.NKSES.EEE.   [310]
TrM5       ..FN..A......R.GSPGGF.S---------SFRKQ.C.ASKSRVGSESEQL.S.EEDS   [343]
LmM5       ..FN..A......R.VSPGGF..P--------SFRNQAC.TMKTGVGSENEQL.S.EEDS   [322]
```

```
HsM5       YLLSPAAAHRPKSQKCVAYKFRLVVKADGNQETNNGCHKVKIMPCPFPVAK-------EP   [413]
MmulM5     .F.........................T..N.............S.....-------..   [413]
MmM5       .F...G....L.................T.......R.......S...S.-------D.   [413]
RnM5       .F........L.................T.......R.......S...S.-------D.   [412]
CpM5       .F........L..........Q.......T.......R.......S...S.-------GT   [413]
GgalM5     .F...--......K..............T..A....R....T..SAALS.-------D.   [409]
DrM5       FFPT.VK.SPT.TK...S...KPKDVS----PLK.TNGDA.PGASS.SS.ESVNAPSSSS   [366]
TrM5       .FQT.PKPQSQ..G...S...KSAA.DGAHA.RQSKNGDT.MASSV.SS.ESVNAPSTSS   [403]
LmM5       FFQT.PKSNSQR.N...S...KP.A.D-THV.HHSKNGDT.MASST.SS.ESMSVPSTSS   [381]

HsM5       STKGLNPNPSHQMTKRKRVVLVKERKAAQTLSAILLAFIITWTPYNIMVLVSTFCDKCVP   [473]
MmulM5     .................M.........................................   [473]
MmM5       .....D.HL.........M.........................................   [473]
RnM5       ....PD..L.........M.........................................   [472]
CpM5       PS...E..L.........M.....................L................T.I.   [473]
GgalM5     .I.SMD..INN.T.....M..I..................................SD.I.   [469]
DrM5       .S.PIDGTLKC.I.....M..I..K..............L..........I....SD.I.   [426]
TrM5       TS.PSDATLKN.I.....M..I.................L..........I....SD.I.   [463]
LmM5       TS.PIDATLKN.I.....M..I.................L..........I....SD.I.   [441]
                                                ----------TM6----------

HsM5       VTLWHLGYWLCYVNSTVNPICYALCNRTFRKTFKMLLLCRWKKKKVEEKLYWQGNSKLP-   [532]
MmulM5     ..........................................................-   [532]
MmM5       ................I..................L.......................-   [532]
RnM5       ................I..................L.......................-   [531]
CpM5       ..................................................--V.EK...A.....-   [530]
GgalM5     L.W.................M......K..........F.Q..............TRM.-   [528]
DrM5       LS..................M......K..Q...R.....Q.R.QRA......C.QNPAVG   [486]
TrM5       LS..................M......K..Q...R.....Q....RI......Y.QNPAVT   [523]
LmM5       LS..................M-------------------------------------------   [461]
              --------------TM7

HsM5       ----    [532]
MmulM5     ----    [532]
MmM5       ----    [532]
RnM5       ----    [531]
CpM5       ----    [530]
GgalM5     ----    [528]
DrM5       SKLT    [490]
TrM5       SKLT    [527]
LmM5       ----    [461]
```

Transmembrane Domains

The amino acid sequence deduced from the bluegill gene fragment was predicted to have six complete transmembrane domains (TM) from TM1 to TM6 and a partial TM7. When compared with the corresponding domains of other vertebrate M_5 receptor proteins, the TMs showed higher percentage identity and similarity than the identity and similarity in the intra- (i) and extracellular (e) loops. Higher identity and similarity was observed with the TMs of the fugu putative M_5 receptor protein than with the similar domains of other vertebrate M_5 receptor proteins. In contrast, the third intracytoplasmic loop (i3) was found to have much less homology with the i3 loops of other vertebrate M_5 receptor proteins. Percentage identity and similarity were considerably higher with the i3 loop of putative fugu M_5 receptor protein than with other vertebrate M_5 receptor proteins. Overall, higher domain identity and similarity was found with putative M_5 receptor proteins in fish than with other vertebrate M_5 receptor proteins (Table 5).

Table 5: Comparison of transmembrane domains (TM) and intra- (i) and extracellular loops (e) of bluegill putative M_5 receptor protein fragment with the corresponding domains of other M_5 receptors.

Domain	Zebrafish M_5	Fugu M_5	Chick M_5	Human M_5
TM1	81 (90)	100 (100)	86 (95)	86 (94)
TM2	95 (95)	100 (100)	90 (94)	90 (94)
TM3	94 (99)	94 (99)	94 (99)	95 (100)
TM4	94 (94)	94 (99)	84 (89)	86 (90)
TM5	95 (95)	100 (100)	100 (100)	100 (100)
TM6	100 (100)	100 (100)	90 (99)	90 (100)
i1	92(92)	92(92)	92(92)	92(92)
i2	75(75)	80(80)	80(80)	100(100)
i3	47 (54)	65 (69)	37 (50)	35 (49)
e1	89(95)	89(95)	74(95)	74(84)
e2	82(82)	82(82)	55(64)	77(81)
e3	89(89)	89(89)	78(89)	NSS
Overall	72 (76)	82 (84)	63 (72)	61 (71)

Genbank accession numbers for M_5 receptor proteins were AAK93794 (Zebrafish), AAF19027 (chick) Zebrafish (Genbank accession no. AAK93794), chick (Genbank accession no. AAF19027) and P08912 (human). Fugu putative M_5 receptor was identified in the present study. Numbers represent percent identity and percent similarity (in parenthesis). NSS-No significant similarity.

Identification of Fugu M_2 Coding Strand

Fugu genomic DNA sequences are publicly available at fugu genomic database website (http://fugu.hgmp.mrc.ac.uk), where the sequences are organized by "clones." The search of fugu genomic database using the zebrafish M_2 coding strand as a query sequence identified clone M000897, which had the highest alignment score with the query. This clone showed very high identity across the whole length of zebrafish M_2 coding strand except in the i3 loop. The overall identity was 85%. The results of the BLASTX of this region of the clone showed very high identity and similarity with human and zebrafish M_2 receptor protein. A 1503 nucleotide long region was identified as a putative fugu M_2 template strand based on the BLASTX and BLASTN alignments with zebrafish M_2 receptor protein and M_2 coding strand, respectively. Upon reverse complementation, the coding strand was found to have an open reading frame encoding 500 amino acids (Figure 8). The deduced amino acid sequence showed highest homology with muscarinic receptor proteins. Among muscarinic receptors, it was most identical to the M_2 subtype. The deduced amino acid sequence was more identical to zebrafish M_2 receptor protein (76%) than to other vertebrate M_2 receptor proteins (Table 4, p. 27). An alignment of fugu M_2 receptor fragment with other known M_2 receptors is shown in Figure 9. The inferred amino acid sequence of the fugu putative M_2 gene was predicted to have seven transmembrane domains. The TMs showed very high identity and similarity with the corresponding domains of other vertebrate M_2 receptor proteins. Much less homology was observed in the i3 loop. For all the domains, higher identity and similarity was observed with the corresponding domains of zebrafish M_2 receptor than with other vertebrate M_2 proteins (Table 6).

```
ATGGACGCGTTCAACTTCACCTACTGGAATGCCTCCGAAGGCAACGAGACGGATGTCGCGGAA 63
 M  D  A  F  N  F  T  Y  W  N  A  S  E  G  N  E  T  D  V  A  E  21
GAGAGCGCGAGCCCCTACAAGACGGTGGAGGTGGTGTTCATCGTGCTGGTGGCCGGGTCCCTC 126
 E  S  A  S  P  Y  K  T  V  E  V  V  F  I  V  L  V  A  G  S  L  42
AGCTTGGTCACCGTCATCGGGAACATCCTGGTCATGCTCTCCATCAAAGTCAACAGGAACCTG 189
 S  L  V  T  V  I  G  N  I  L  V  M  L  S  I  K  V  N  R  N  L  63
CAGACGGTCAACAACTATTTTTTGTTCAGCCTGGCGTGTGCTGACCTCATCATCGGACTCTGC 252
 Q  T  V  N  N  Y  F  L  F  S  L  A  C  A  D  L  I  I  G  L  C  84
TCCATGAACTTGTACACGGTCTACATTGTGATCGGCTACTGGCCTCTGGGCCCGGTGGTGTGC 315
 S  M  N  L  Y  T  V  Y  I  V  I  G  Y  W  P  L  G  P  V  V  C  105
GACCTCTGGTTGGCGTTGGACTATGTTGTCAGCAACGCGTCCGTCATGAACCTCCTCATCATC 378
 D  L  W  L  A  L  D  Y  V  V  S  N  A  S  V  M  N  L  L  I  I  126
AGCTTTGACAGATATTTTTGCGTCACCAAGCCCCTCAGCTACCCTGTCAAGAGGACCACCAAG 441
 S  F  D  R  Y  F  C  V  T  K  P  L  S  Y  P  V  K  R  T  T  K  147
ATGGCGGGAATGATGATCGCGGCGGCCTGGGTCCTTTCCTTCATCCTCTGGGCTCCAGCGATT 504
 M  A  G  M  M  I  A  A  A  W  V  L  S  F  I  L  W  A  P  A  I  168
CTCTTCTGGCAGTTCATCGTTGGTGGGAGGACGGTGCCGGAGAAGGAGTGCTACATCCAGTTT 567
 L  F  W  Q  F  I  V  G  G  R  T  V  P  E  K  E  C  Y  I  Q  F  189
TTCTCAAATGCCGCGGTGACTTTCGGCACCGCCATCGCCGCCTTTTACTTGCCTGTCATCATC 630
 F  S  N  A  A  V  T  F  G  T  A  I  A  A  F  Y  L  P  V  I  I  210
ATGATTCAGCTCTACTGGCAGATCTCCCGAGCGAGCAAGAGCCGCGTGAAGAAGGAGAACCGC 693
 M  I  Q  L  Y  W  Q  I  S  R  A  S  K  S  R  V  K  K  E  N  R  231
AAACCGTCGGGCCCCAATCCAGAGCCCCTGTTACAAGGCCAGAGGAGGAACAACACGCCAAAA 756
 K  P  S  G  P  N  P  E  P  L  L  Q  G  Q  R  R  N  N  T  P  K  252
GCCAACAATAACAACGTACCGGGGGAAGATACAGGATGTTCTCAGAGCCAGAACGCCAACCAC 819
 A  N  N  N  N  V  P  G  E  D  T  G  C  S  Q  S  Q  N  A  N  H  273
GGCGCCAACCAGCACGAGGAAAAACTGCAGAACGGCAAGGGACCGTCCTCCACCACCGCCGAG 882
 G  A  N  Q  H  E  E  K  L  Q  N  G  K  G  P  S  S  T  T  A  E  294
GGAGAAACTGAAGGAGACGACATGACGAGGGAGAACTGCACCACCGCAGAGGAGAAAGAGAGC 945
 G  E  T  E  G  D  D  M  T  R  E  N  C  T  T  A  E  E  K  E  S  315
TCCAACGATTCCACATCGGGCAGCATGGCCAACCAGAAGGAGGAGGAGGCGGCGCCCTCCGCC 1008
 S  N  D  S  T  S  G  S  M  A  N  Q  K  E  E  E  A  A  P  S  A  336
GCCCACACCAGTGCAGAGGCGAGCCAGCCGCTCCCACGCCAGCGGGCGAAGGCGGGCGGTTCG 1071
 A  H  T  S  A  E  A  S  Q  P  L  P  R  Q  R  A  K  A  G  G  S  357
AAGCTGACCTGCATCAAGATCAAGACTAAATCACCCAAGGGGGACTGCTACACGCCCTCCAAC 1134
 K  L  T  C  I  K  I  K  T  K  S  P  K  G  D  C  Y  T  P  S  N  378
GCCACCGTGGAGATCGTCCCGGCCACCGAGCGGCAGAACCACGTGGCGCGGAAGATCGTGAAG 1197
 A  T  V  E  I  V  P  A  T  E  R  Q  N  H  V  A  R  K  I  V  K  399
ATGACGAAGCAGCCGCCCAACAAGAAGAAAAAAGGGCCGCCGTCGCGGGAGAAGAAGGTGACC 1260
 M  T  K  Q  P  P  N  K  K  K  K  G  P  P  S  R  E  K  K  V  T  420
CGCACCATCATGGCCATCCTGGTGGCCTTCGTGGCCACCTGGACTCCGTACAACGTGATGGTG 1323
 R  T  I  M  A  I  L  V  A  F  V  A  T  W  T  P  Y  N  V  M  V  441
CTCATCAACACCTTCTGCTCCAGCTGCATCCCCAACACCGTCTGGACTATCGGCTACTGGCTG 1386
 L  I  N  T  F  C  S  S  C  I  P  N  T  V  W  T  I  G  Y  W  L  462
TGCTACATCAACAGCACCATCAACCCGGCCTGCTACGCCCTCTGCAACGTCACCTTCAAGAAG 1449
 C  Y  I  N  S  T  I  N  P  A  C  Y  A  L  C  N  V  T  F  K  K  483
ACCTTCAAGCACCTCCTCCTCTGCCAGTACAAGAACAGCCGCTCCGCCAGATAG          1503
 T  F  K  H  L  L  L  C  Q  Y  K  N  S  R  S  A  R stop         500
```

Figure 8: Nucleotide sequence of fugu putative M_2 coding strand and the deduced amino acid sequence. The nucleotide and amino acid sequences are 1503 and 500 in length respectively.

Figure 9: Alignment of deduced amino acid sequence of fugu putative M2 receptor with known vertebrate M2 receptors using the program Clustal X (Thompson et al. 1997). The names of the sequences are prefixed with Hs (*Homo sapiens*, human), Ss (*Sus scrofa*, pig), Mm (*Mus musculus*, mouse), Rn (*Rattus norvegicus*, rat), Cp (*Cavia porcellus*, guinea pig), Ggal (*Gallus gallus*, chick), Dr (*Danio rerio*, zebrafish) and Tr (*Takifugu rubripes*, fugu). Genbank accession numbers for these sequences are NP_000730 (HsM2), X04708 (SsM2), AF264049 (MmM2), NM_031016 (RnM2), AY072059 (CpM2), M73217 (GgalM2), AY039653 (DrM2). TrM2 represents fugu putative M2 receptor. Residues identical to that of human M2 are indicated by dots. Dashes in the alignment represent gaps inserted. Transmembrane domains (TM) of fugu M2 are delineated by dashes below the sequences. Amino acid motif and residues which are either conserved across GPCR family or only in muscarinic and other closely related receptors are shown in bold, colored fonts. The conserved aspartic acid-arginine-tyrosine (DRY) motif is boxed. These conserved motif and residues are critical for receptor function. For details see text.

```
HsM2      --------MNNS---TNSSNNSLALTSPYKTFEVVFIVLVAGSLSLVTIIGNILVMVSIK   [49]
SsM2      --------....---.....SG.....................................   [49]
MmM2      --------....---......G..I..................................   [49]
RnM2      --------....---......G..I..................................   [49]
CpM2      --------....---.........I..................................   [49]
GgalM2    --------....TYINS..E.VI..E.....I...........................   [52]
DrM2      MDTINFTFW.A.D--G.ETMETAD-E.....V.............V.......L...   [57]
TrM2      MDAFNFTYW.A.E--G.ETDVAEESA.....V.............V.......L...   [58]
                                          ------------TM1--------

HsM2      VNRHLQTVNNYFLFSLACADLIIGVFSMNLYTLYTVIGYWPLGPVVCDLWLALDYVVSNA   [109]
SsM2      ............................................................   [109]
MmM2      ............................................................   [109]
RnM2      .S..........................................................   [109]
CpM2      .....................................F......................   [109]
GgalM2    ........................I...................................   [112]
DrM2      ...S.....................LC......V.I.....................*..   [117]
TrM2      ...N.....................LC......V.I.....................*..   [118]
                    ----------TM2----------                 ---------TM3

HsM2      SVMNLLIISFDRYFCVTKPLTYPVKRTTKMAGMMIAAAWVLSFILWAPAILFWQFIVGVR   [169]
SsM2      ............................................................   [169]
MmM2      ............................................................   [169]
RnM2      ............................................................   [169]
CpM2      ............................................................   [169]
GgalM2    ............................................................G.   [172]
DrM2      .....................S...N.................................G.   [177]
TrM2      .....................S.....................................G.   [178]
          --------                              ----------TM4----------

HsM2      TVEDGECYIQFFSNAAVTFGTAIAAFYLPVIIMTVLYWHISRASKSRIKKDKKEPVANQD   [229]
SsM2      ............................................................E   [229]
MmM2      ..........................................................E.....   [229]
RnM2      ..........................................................E.....   [229]
CpM2      ..........................................................E.....   [229]
GgalM2    ..P.KD........P.......................Q...........G...AAQ...   [232]
DrM2      ..PEK.............................M....QV.......V...NRK.SGGNL   [237]
TrM2      ..PEK.............................IQ...Q........V..ENRK.SGPNP   [238]
                         ----------TM5----------

HsM2      PVSPS--LVQGRIVKPNNNNMPSSDDGLEHN------------KIQNGKAPRDPV-----   [270]
SsM2      .....--..............G..EA....------------..........A.-----   [270]
MmM2      .....--..............GG.G.....------------.......L..GG-----   [270]
RnM2      .....--..............GG.G.....------------..........G.-----   [270]
CpM2      .....--..............G..SA....------------......T..GV.-----   [270]
GgalM2    .....--....K........I.T.S......------------.V....TTGES.-----   [273]
DrM2      D.AS.NQIRENSAN..T...LTAEETDRGQTQLTDDTINQHDA.L......STASGEAEE   [297]
TrM2      EPLLQGQRRNNTP-.A....V.GE.T.CSQSQNANHGANQHEE.L....G.SSTTAEGET   [297]

HsM2      ------TENCVQGEEKESSNDSTSVSAVASNMRDDEITQDENTVSTS----LGHSKDENS   [320]
SsM2      ------.....................................----.........   [320]
MmM2      ------.....................................----.......D..   [320]
RnM2      ------..T..............SA.....................-----.D..R.D..   [320]
CpM2      ------.....................................I...-----..QT.....   [320]
GgalM2    ------M.........D.........V.P..TKE..AAK.ASQI.A.----QD.L.V...   [323]
DrM2      AG----QA..IPA...........G.GAVT.QKEEAAPPSSAAAND.QTSTRHRA.AGG.   [353]
TrM2      EGDDMTR...TTA...........G.MANQKEEEAAPSAAHTSAEA.QPLPRQRA.AGG.   [357]
```

```
HsM2       KQTCIRIGTKTPKSDSCTPTNTTVEVVGSSGQNGDEKQNIVARKIVKMTKQPA--KKKPP   [378]
SsM2       .....K.V...Q.......A.....L.............................--.....   [378]
MmM2       R....K.V...Q.G.A....S....L.............................--.....   [378]
RnM2       .....K.V..AQ.G.VY...S....L......S......V........P....--.....   [378]
CpM2       .....K.V...Q.G..G..SH....L.............................--.....   [378]
GgalM2     .L.....V..SQ.G.C.A.......I..---T.......S..............--.....   [378]
DrM2       .L...K.I..S..G.CYA.S.A...I.P-----AV.R..H............P-K...A.   [407]
TrM2       .L...K.K..S..G.CY..S.A...I.P-----AT.R..H............PNK...G.   [412]

HsM2       PSREKKVTRTILAILLAFIITWAPYNVMVLINTFCAPCIPNTVWTIGYWLCYINSTINPA   [438]
SsM2       ............................................................   [438]
MmM2       ............................................................   [438]
RnM2       ............................................................   [438]
CpM2       .....................................S.V....................   [438]
GgalM2     ....................T.........S...S...G....................   [438]
DrM2       S.........M...V..VA..T.........A..SS........................   [467]
TrM2       ..........M...V..VA..T..........SS..........................   [472]
                     ----------TM6----------         ----------TM7-----

HsM2       CYALCNATFKKTFKHLLMCHYKNIGATR   [466]
SsM2       ............................   [466]
MmM2       ............................   [466]
RnM2       ............................   [466]
CpM2       ............................   [466]
GgalM2     ............................   [466]
DrM2       ......I.......Q..L.Q....RS..   [495]
TrM2       ......V..........L.Q...SRSA.   [500]
           -----
```

Table 6: Comparison of transmembrane domains (TM), intra- (i) and extracellular loops (e) and amino (N end) and carboxy (C end) terminals of fugu putative M_2 receptor protein with the corresponding domains of other M_2 receptors.

Domains	Zebrafish M_2	Chick M_2	Human M_2
TM1	100 (100)	86 (94)	91 (99)
TM2	100 (100)	86 (94)	86 (94)
TM3	100 (100)	100 (100)	100 (100)
TM4	100 (100)	83 (83)	83 (83)
TM5	91 (95)	87 (87)	87 (87)
TM6	96(96)	74(87)	70 (83)
TM7	100 (100)	96(96)	96 (96)
i1	93(100)	93(100)	92(92)
i2	96(96)	69(73)	69(73)
i3	49 (55)	34 (40)	32 (41)
e1	100(100)	86(86)	57(57)
e2	100(100)	79(89)	68(74)
e3	100(100)	78(78)	78(78)
N end	74(85)	NSS	NSS
C end	82(86)	69(77)	69(77)
Overall	76 (80)	65(72)	63(69)

Genbank accession numbers for M_2 proteins were NP_840086 (zebrafish), P30372 (chick) and P08172 (human). Numbers represent percent identity and percent similarity (in parenthesis). NSS-No significant similarity.

Phylogenetic Analyses

The tree topologies obtained using both protein and nucleotide alignments employing different methods were similar. In all the trees vertebrate muscarinic receptors formed one ingroup. Within the ingroup, two monophyletic groups were observed, one formed by odd numbered muscarinic receptors and another by even numbered muscarinic receptors. Within these groups, receptors belonging to the same subtype formed monophyletic clades. The bluegill putative M_5 receptor fragment formed a monophyletic unit with other M_5 receptors and within the clade it was within a terminal clade with fugu putative and zebrafish putative M_5 receptors. This terminal clade formed a sister group to the other vertebrate M_5 receptors which were grouped together. The fugu putative M_2 receptor was also found to be grouped with its own subtype. Within the M_2 group, it formed a terminal clade with zebrafish M_2 receptor. This clade formed a sister group to the other vertebrate M_2 receptors which were grouped together. In the trees obtained using nucleotide alignment employing parsimony, maximum likelihood, Bayesian and neighbor joining methods, the monophyletic clades formed by M_1 and M_3 receptors formed one monophyletic unit with the monophyletic clade formed by M_5 receptors forming its sister group (Figure 10). High bootstrap support (>70) was observed for all the monophyletic groups described above using different methods except parsimony and maximum likelihood bootstrap support for the monophyletic unit formed by M_1 and M_3 clades (<70) (Figure 11). The tree topologies obtained using protein alignment employing maximum likelihood and neighbor joining methods also resolved the same intragroup arrangement within the monophyletic group formed by clades of odd numbered receptors with high bootstrap support (Figure 12) but the tree topology obtained using protein

alignment under parsimony criterion differed in the intragroup arrangement of odd numbered muscarinic receptors. In this topology, the monophyletic M_1 and M_5 clades formed one monophyletic unit with the monophyletic M_3 receptor clade as its sister group (Figure 13).

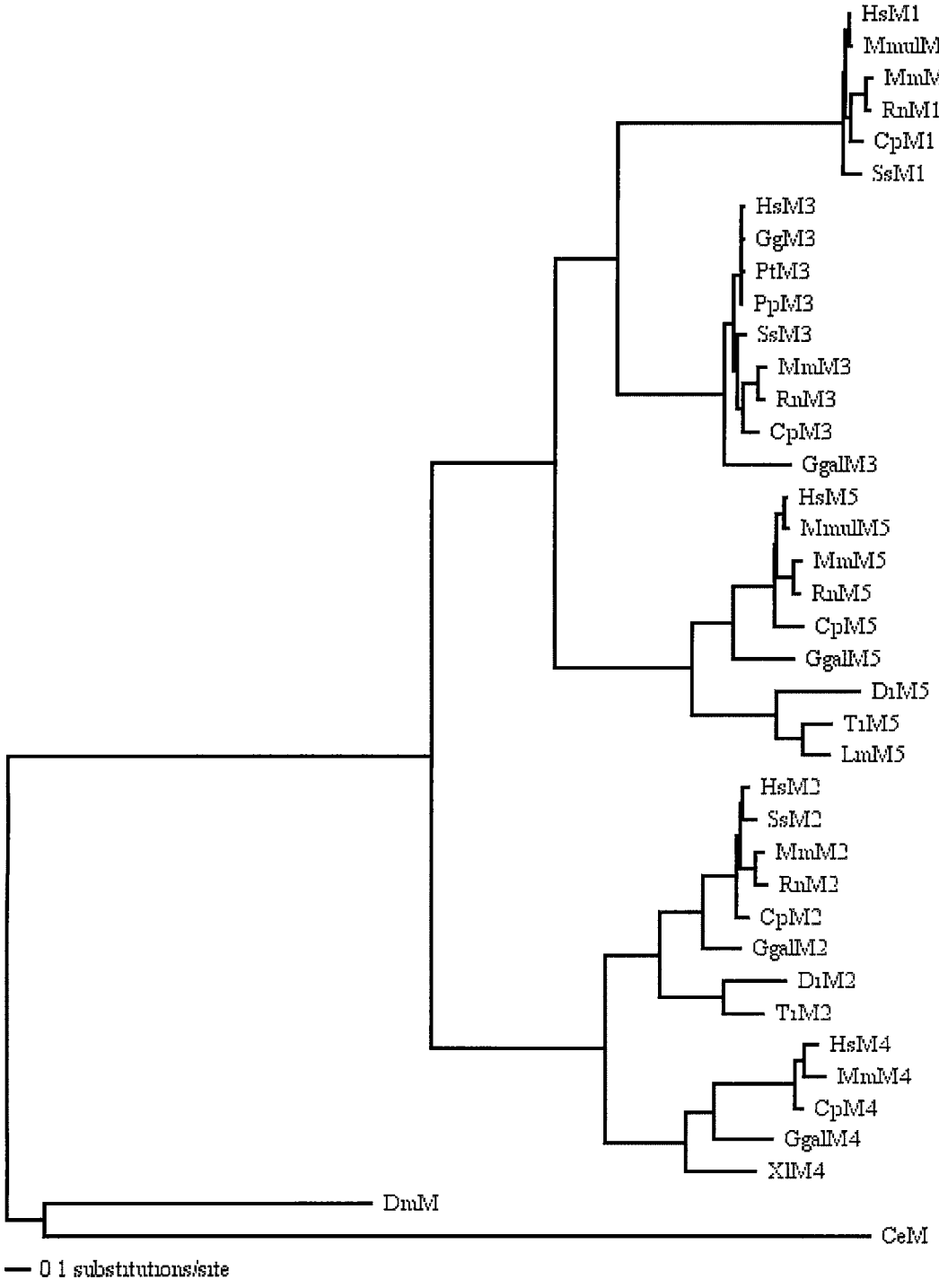


Figure 10: Phylogram obtained using PAUP* under maximum likelihood criterion using nucleotide alignment. The tree topologies obtained employing parsimony, Bayesian and neighbor joining methods using nucleotide alignment showed similar clade arrangement.

Figure 11: The representative tree topology obtained using nucleotide alignment under parsimony, maximum likelihood, Bayesian and neighbor joining criteria. The numbers at each node represent support values. The numbers in red font are parsimony bootstrap values. The numbers in blue font are maximum likelihood bootstrap values. The numbers in green font are Bayesian support values while numbers in black font are neighbor joining bootstrap values. The support values for terminal clades except for those formed by bluegill putative M_5 receptor fragment (LmM5) and fugu putative M_2 receptor (TrM2) are not shown as they differed in arrangement in different tree topologies.

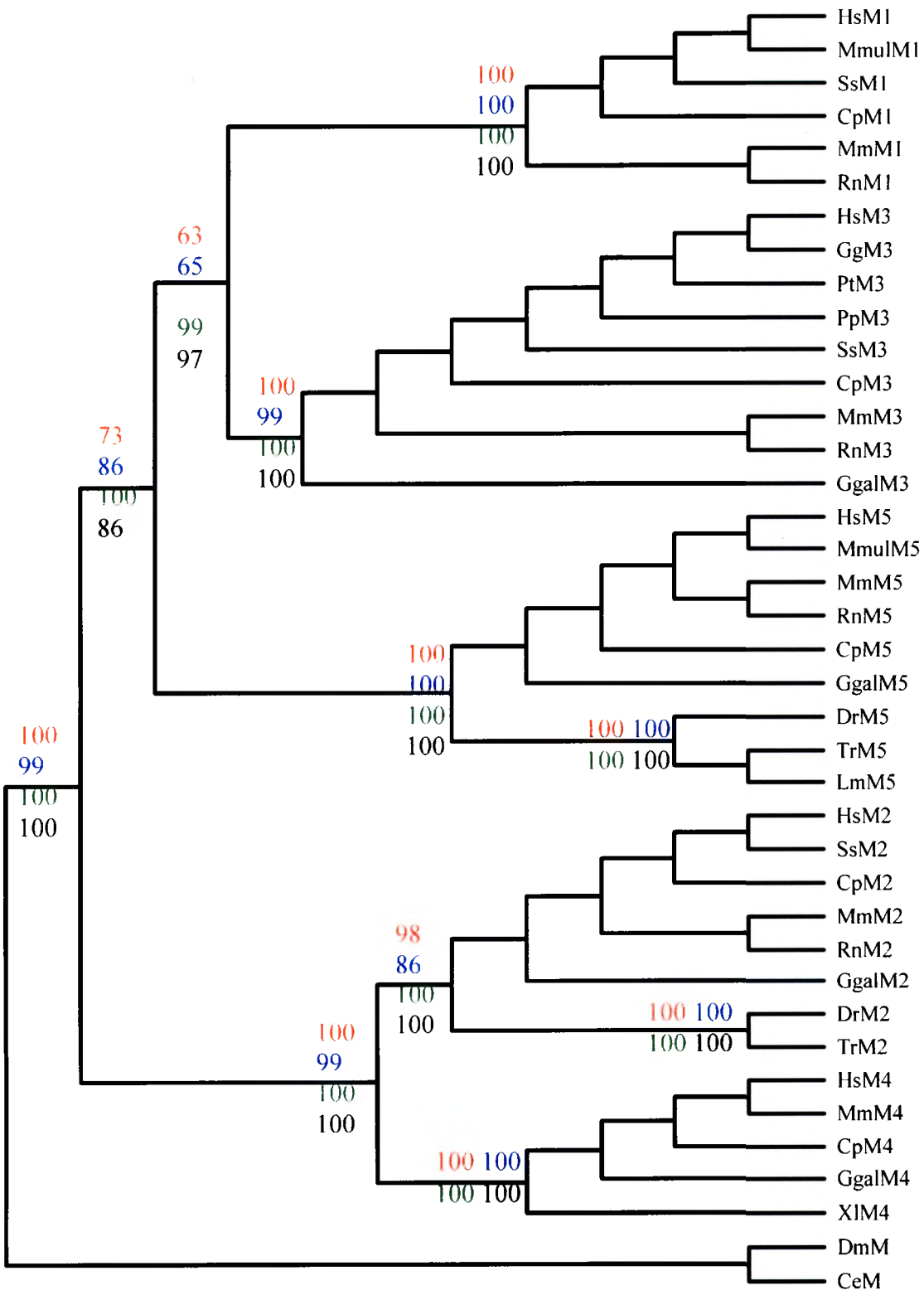
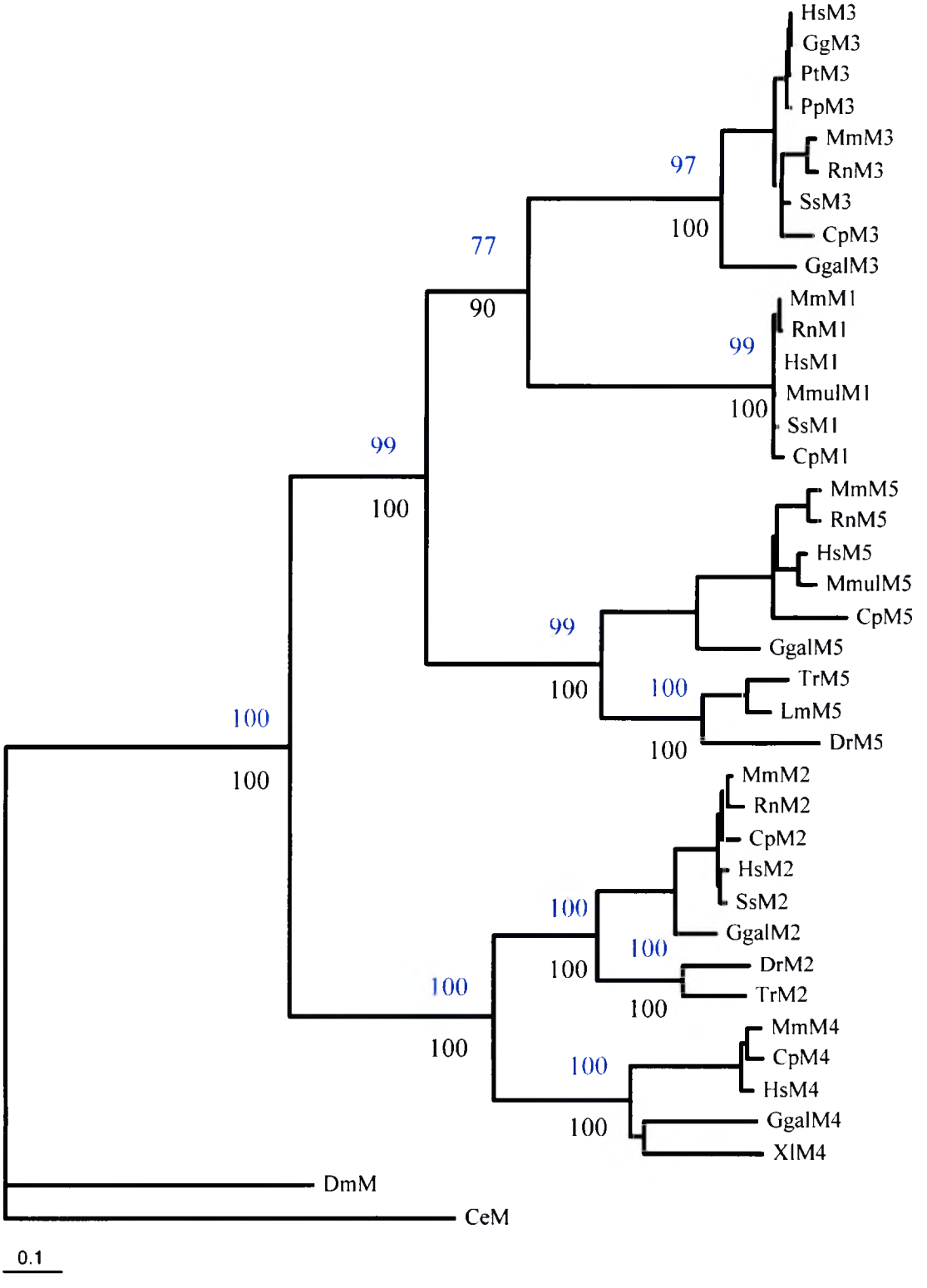

HsM1
MmulM1
SsM1
CpM1
MmM1
RnM1
HsM3
GgM3
PtM3
PpM3
SsM3
CpM3
MmM3
RnM3
GgalM3
HsM5
MmulM5
MmM5
RnM5
CpM5
GgalM5
DrM5
TrM5
LmM5
HsM2
SsM2
CpM2
MmM2
RnM2
GgalM2
DrM2
TrM2
HsM4
MmM4
CpM4
GgalM4
XlM4
DmM
CeM

Figure 12: Phylogram obtained using Puzzle 5.0 under maximum likelihood criterion using protein alignment. The tree topology obtained employing neighbor joining method using protein alignment also showed similar clade arrangement. The numbers at each node represent support values. The numbers in blue font are maximum likelihood support values while numbers in black font are neighbor joining bootstrap values. The support values for terminal clades except for those formed by bluegill putative M_5 receptor fragment (LmM5) and fugu putative M_2 receptor (TrM2) are not shown as they differed in arrangement in different tree topologies.


HsM3
GgM3
PtM3
PpM3
MmM3
RnM3
SsM3
CpM3
GgalM3
MmM1
RnM1
HsM1
MmulM1
SsM1
CpM1
MmM5
RnM5
HsM5
MmulM5
CpM5
GgalM5
TrM5
LmM5
DrM5
MmM2
RnM2
CpM2
HsM2
SsM2
GgalM2
DrM2
TrM2
MmM4
CpM4
HsM4
GgalM4
XlM4
DmM
CeM
97
100
77
90
99
100
99
100
99
100
100
100
100
100
100
100
100
100
100
100
100
0.1

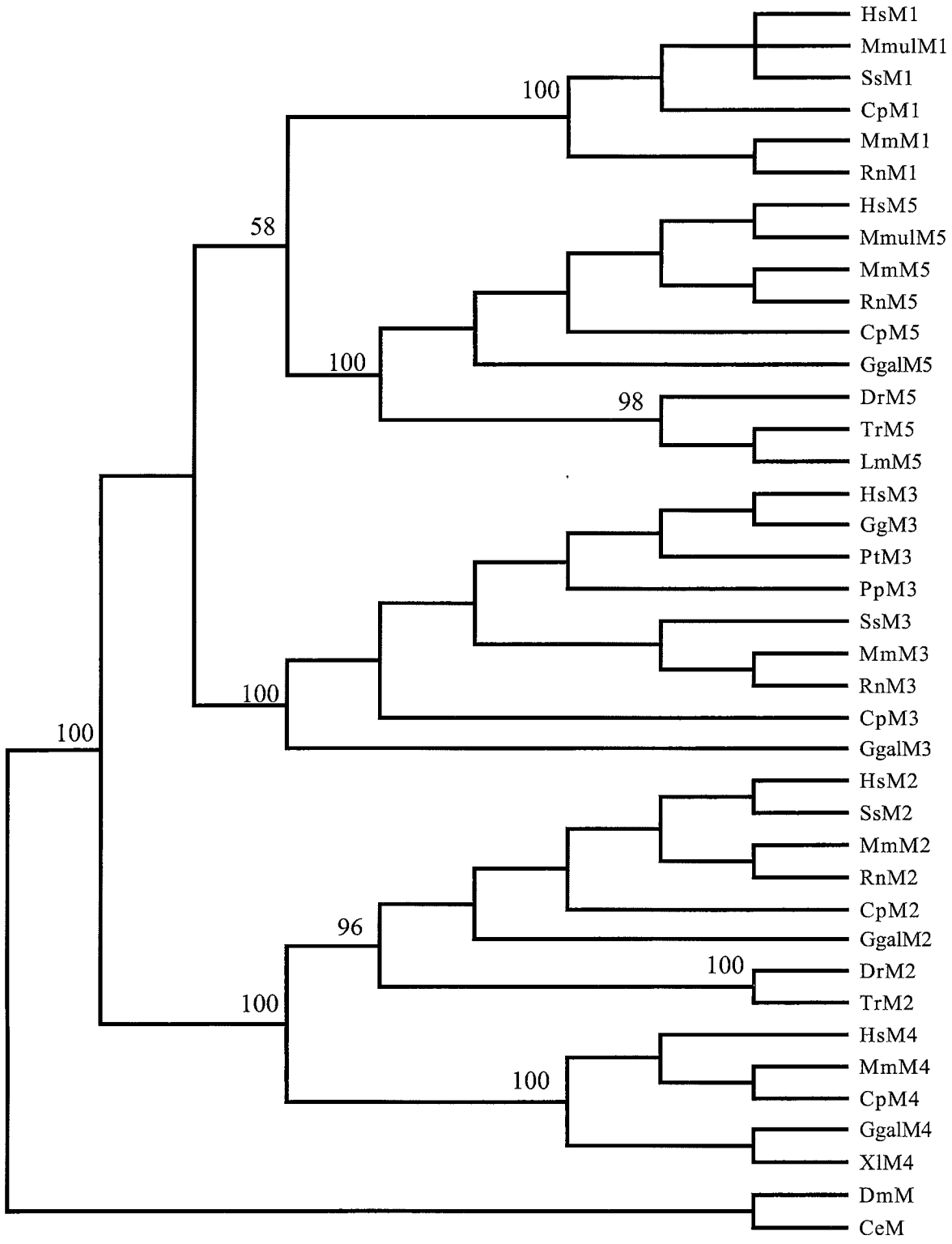


Figure 13: Tree topology obtained using PAUP* under parsimony criterion using protein alignment. The numbers at the nodes represent bootstrap values.

DISCUSSION

I have amplified a 1385 nucleotide-long fragment of bluegill putative M_5 receptor gene using primers based on the homologous regions near the 5' and 3' ends of zebrafish M_5 coding strand, fugu putative M_5 coding strand and homologous fragments of *Tetraodon nigroviridis* whole genome sequence. The nucleotide sequence has an open reading frame encoding 461 amino acids. In addition I have identified a fugu putative M_2 coding strand from fugu genomic database by BLAST search using zebrafish M_2 coding strand as a probe. The fugu M_2 coding strand is 1500 nucleotides long. This sequence encodes a protein of 500 amino acids.

When compared with the protein database on NCBI, the nucleotide sequences and their deduced amino acid sequences showed the highest degree of homology to known muscarinic receptors. The bluegill putative M_5 receptor protein fragment is more homologous to M_{odd} subtypes than M_{even} subtypes, while the fugu putative M_2 receptor shows exactly reverse trend in homology. When compared with zebrafish M_2 and M_5, fugu M_2 and M_5, chick M_2-M_5 and human M_1-M_5 receptor proteins, the bluegill putative M_5 receptor protein fragment showed higher homology with M_5 receptors than with any other subtype, while fugu putative M_2 receptor protein showed highest homology to M_2 subtypes. The degree of homology for bluegill putative M_5 receptor fragment was fugu M_5 > zebrafish M_5 > chick M_5 > human M_5, while for fugu putative M_2 receptor it was zebrafish M_2 > chick M_2 > human M_2. Overall, both the sequences showed more

homology in the transmembrane domains than intra- and extracellular loops when compared with the corresponding domains of their respective subtypes. Phylogenetic analysis supported the assignment of the receptors from bluegill and fugu described herein to the M_5 and M_2 subtypes, respectively.

Critical Amino Acids

Amino acids conserved across GPCR family

Only about 20 amino acids are conserved across the entire G-protein coupled receptor superfamily (Hulme et al., 1990; Wess et al., 1993). One of the most remarkable homologies among various GPCRs is a triplet of amino acids, aspartate-arginine-tyrosine, located at the interface of TM3 with the second intracytoplasmic loop (Jones et al., 1995; Zhu et al., 1994). An alignment of over 200 sequences shows that the arginine residue is invariant, whilst the neighboring residues are sometimes replaced (Baldwin, 1993). Among known muscarinic receptors this triplet is fully conserved except for the M_4 receptor in *Xenopus laevis* where aspartic acid is replaced by glutamic acid, a conservative substitution. Within this triplet, arginine has been found to be critical in coupling to intracellular transduction mechanism in both odd and even muscarinic receptors (Zhu et al., 1994; Jones et al., 1995). Missense mutations affecting the neighboring aspartic acid and tyrosine residues have also been found to decrease receptor coupling to intracellular transduction mechanisms (Fraser et al., 1989; Zhu et al., 1994). Thus this triplet seems to play a role in muscarinic receptor coupling to intracellular transduction mechanisms. Both, bluegill putative M_5 receptor fragment and fugu putative M_2 receptor have this motif at the predicted interface of TM3 and second intracytoplasmic loop. This motif is shown boxed and in bold font in Figures 7 and 9.

The residues conserved among all the G-protein coupled receptors, most of which are located within the seven TM helices, have been suggested to play key roles in protein folding and/or receptor function (Wess et al., 1993; Hulme et al., 1990). There are four proline residues, one each in TM4, 5, 6, and 7 that are particularly well conserved among GPCRs. Their likely effect is to introduce kinks into the TM helices which might be essential for receptor activation (Hulme et al., 1990). Site-directed mutagenesis studies in rat M_3 in which the conserved prolines were mutated to alanine to straighten the kinked helices, indicated a role for proline residues in inducing a proper protein fold that allows efficient intracellular receptor trafficking and/or stable plasma membrane integration (Wess et al., 1993). In the same study, the proline residue in TM7 was found to play a role in agonist-induced changes in receptor conformation that trigger G-protein activation. A small polar region immediately preceding proline residue in TM7 was speculated to be critically involved in receptor activation. Fugu putative M_2 receptor and bluegill putative M_5 receptor fragment have these four proline residues at the corresponding positions in TM4, 5, 6, and 7. The conserved proline residues are depicted in bold, gold-colored font in Figures 7 and 9.

In contrast to proline residues in TM5, 6, and 7, the proline in TM4 is not considered to influence protein folding critical for receptor expression but has been found to affect agonist and antagonist affinities (Wess et al., 1993). This residue is predicted to be located at a level within the outer leaflet of the plasma membrane similar to the level of a series of residues that are involved in binding of muscarinic agonists. However, the proline in TM4 doesn't face the central pore-like cavity enclosed by seven TM helices where ligand-binding is thought to occur (Wess et al., 1993; Hulme et al., 1990; Curtis et

al., 1989). It is interesting to note that there are two consecutive proline residues at the corresponding position in the TM4 of the bluegill putative M_5 receptor fragment which are observed in the other two putative fish M_5 receptors, namely zebrafish M_5 (Liao et al. 2001, submitted to NCBI but unpublished) and fugu putative M_5. This structural feature is also shared by muscarinic receptors in *Drosphila melanogaster* and *Caenorhabditis elegans* (Hwang et al. 1999). Expression and mutagenesis studies would show whether this extra proline residue produces more kinks thereby influencing protein folding required for receptor expression and how an extra proline affects agonist and antagonist affinities.

Three conserved tryptophan residues, one each in TM4, 6, and 7 are predicted to be directly involved in ligand binding (Hibert et al., 1991). Wess et al. (1993), using site-directed mutagenesis, showed that these residues are not critically important for receptor activation; nevertheless, substitution of tryptophan residues in TM4 and 6 resulted in reduced ligand binding affinities. These tryptophan residues might be involved in recognition of the receptor by the ligand (Wess et al., 1993). Fugu putative M_2 receptor and bluegill putative M_5 receptor fragment have these three residues in TM4, 6, and 7. The conserved tryptophan residues are shown in Figures 7 and 9 in blue, bold font.

Amino acids typical of muscarinic receptors and other closely related receptors

Transmembrane domains of all muscarinic receptors contain a series of conserved serine, threonine and tyrosine residues, which, with very few exceptions do not occur in any other G-protein coupled receptor (Wess et al., 1991). To elucidate their role in ligand-binding, particularly hydrogen bonding interactions with electron rich moieties in biogenic amine ligands, Wess et al. (1991) mutated serine and threonine residues to

alanine and tyrosine to phenylalanine in rat M_3 receptor. Out of the nine residues, mutation of conserved tyrosine residues in TM3, 6, and 7 and threonine residues in TM5 resulted in strong reduction in agonist binding affinities. It was proposed that threonine residues in TM5 and tyrosine in TM6 interact with muscarinic agonists through hydrogen bonds, thereby causing conformational change in the third intracytoplasmic loop which contains the structural determinants for G-protein recognition and activation (Wess et al., 1991). In the same study, the conserved serine residue in TM2 was found to influence antagonist binding affinities. All these serine, threonine and tyrosine residues are present at the corresponding positions in the transmembrane domains of fugu putative M_2 receptor and bluegill putative M_5 receptor fragment. The conserved tyrosine residues in TM3, 6 and 7, threonine residues in TM5 and serine residue in TM2 are depicted in Figures 7 and 9 in green, black and orange fonts, respectively. All the fonts are bold and threonine residues are depicted on a yellow background.

Four aspartic acid residues, one each in TM2, the first extracellular loop, TM3 and at the interface of TM3 and second intracytoplasmic loop have been found to be conserved among known muscarinic receptors. Mutagenesis studies have implicated a role for these residues in ligand binding and agonist induced activation of the β_2-adrenergic receptor (Chung et al., 1988; Fraser et al., 1988; Strader et al., 1988). Fraser et al. (1989) examined the roles of these conserved aspartic acid residues in rat M_1 muscarinic receptor by site-directed mutagenesis wherein the aspartic acid residues were replaced by asparagine to eliminate the negative charge at each position. The mutation of aspartic acid in TM2 increased affinity for agonist but decreased the ability of agonist to activate the intracellular signaling pathway. This residue was suggested to be involved in

agonist-induced receptor activation (Fraser et al., 1989). Mutation of aspartic acid residue at the interface of TM3 and second intracytoplasmic loop also increased affinity for agonist, but decreased the efficiency of receptor-effector coupling. It was suggested that this residue might be critical for normal receptor-G-protein interactions and alignment. Based on the reduced affinity for antagonist and agonist observed in the receptor mutated at the aspartic acid residue in the first extracellular loop, this aspartic acid residue and the aspartic acid residue at the proximal end of TM3 were suggested as likely sites of ligand binding (Fraser et al., 1989). All four aspartic acid residues implicated directly or indirectly in ligand binding are observed at their corresponding positions in the fugu putative M_2 receptor and bluegill putative M_5 receptor fragment. In Figures 7 and 9, these residues are shown in bold, violet font.

There are nine cysteine residues that are conserved among muscarinic receptors. These include one each in the TM2, the first and the second extracellular loop and two each in the third extracellular loop, TM7 and in the carboxy terminal (Savarese et al., 1992). Site directed mutagenesis studies in rat M_1 muscarinic receptor showed that out of the two cysteine residues in the third extracellular loop, the one nearer to the N-terminal end of the loop and the two cysteine residues in the carboxy terminus of the receptor did not affect ligand binding and the ability of the receptor to increase phospholipase C activity (Savarese et al., 1992). In other words, replacing these cysteines with serine had no detectable effect on receptor function. It was suggested that these residues do not directly participate in receptor-ligand or receptor-G-protein interaction. Biochemical evidence and the peptide sequence suggest that the two extracellular cysteine residues, one in the first and the other in the second extracellular loop, form a disulphide bond

(Curtis et al., 1989; Kurtenbach et al., 1990). This disulphide bond is postulated to be an important structural requirement to enable the ionic interaction between the negatively charged side chain of aspartic acid in the TM3 and the positively charged group of muscarinic ligands (Hulme et al., 1990). The disulphide bond has also been suggested to be critical for proper protein folding thereby preventing premature degradation of translated polypeptide (Karnik et al. 1988). Cysteine residues in TM7 have been found to affect agonist and antagonist affinities as well as stimulation of intracellular signal transduction (Savarese et al., 1992). Their effect is believed to be due to their proximity and influence on the tyrosine residues in TM7 which are suggested to interact with aspartic acid residues in the TM2, TM3 and the second intracytoplasmic loop. These aspartic acid residues in turn are critical for ligand binding and G-protein activation (Savarese et al., 1992). The fugu putative M_2 receptor has all the conserved cysteine residues at the corresponding positions. The bluegill putative M_5 receptor fragment has the conserved cysteine residues in the first, second and third extracellular loops and one of the cysteine residues in TM7 (The bluegill putative M_5 receptor gene fragment isolated in the present study contains only a partial sequence of TM7.), but the cysteine residues in the TM2 has been replaced by alanine. The fugu putative M_5 receptor also has alanine in the corresponding position while zebrafish putative M_5 has phenylalanine (Liao et al. 2001, submitted to NCBI but unpublished). This cysteine residue in TM2 is not critical for agonist or antagonist binding (Savarese et al., 1992). The conserved cysteine residues are depicted in red, bold font in Figures 7 and 9.

Amino acids involved in ligand-binding

Ligand binding to muscarinic receptors and other biogenic amine receptors is predicted to occur in a pocket formed by the ring-like arrangement of the seven TM domains (Wess, 1993). Ligand binding is thought to be initiated by ionic interactions between the positively charged amino group present in virtually all muscarinic receptor ligands and a conserved aspartic acid residue located in TM3 (Wess 1993). This residue is conserved among all the receptors that bind biogenic amine ligands. The specificity of the given amine ligand for a particular receptor is determined by the additional interactions between the hydroxyl groups of serine, threonine and tyrosine residues and the ligand. These conserved residues are located in the hydrophobic core made up of the seven TM domains and are thought to interact with the ester moiety in acetylcholine (Wess et al., 1991). As mentioned earlier threonine and tyrosine residues are important for acetylcholine binding (Wess et al., 1991). In an extension of their mutagenesis studies, Wess et al. (1992) showed that substituting the ester moiety in acetylcholine with ether or ketooxygen atoms restored agonist affinity of the mutated rat M_3 receptor lacking the critical threonine and tyrosine residues, supporting the concept that the majority of those residues are important for the interaction of the acetylcholine with the receptor protein.

It is predicted that the seven TM domains form a central binding cavity and the critical amino acids in ligand binding are located in the TM3-7 facing the cavity. These critical amino acids are located near the extracellular region of the TM domains (Wess, 1993). The fugu putative M_2 receptor protein sequence is predicted to have all the seven TM domains and the critical amino acids involved in ligand binding. This prediction

implies that if expressed, these putative receptors should be able to bind muscarinic agonists, although this needs to be verified through functional studies. The bluegill putative M_5 receptor protein fragment contains only six transmembrane domains but still has the critical amino acids for ligand binding in the TM3-6. Amplification and sequence analysis of the complete coding strand would reveal whether this receptor is functional.

i3 loop

The third intracytoplasmic loop (i3) located between the TM5 and TM6 is a determinant of G-protein selectivity. This conclusion is based on the observations using chimeric muscarinic receptors. In the M_1/M_2 chimeric receptors, exchanging the i3 domain between M_1 and M_2 caused a reversal in the ability of the resulting hybrid receptors to couple to specific ion channels (Kubo et al., 1988). Also the M_2/M_3-i3 and M_3/M_2-i3 hybrid receptors displayed the same functional profile as M_3 and M_2, respectively (Wess et al., 1990). Differences in the i3 loop account for most of the sequence diversity between muscarinic receptor subtypes and also between muscarinic receptors from different species (Eglen and Nahorski, 2000). The i3 loop is variable in size among different subtypes and M_3 and M_5 possess larger i3 loops than the other subtypes. The i3 loop of bluegill putative M_5 receptor fragment contains 222 acids, which are in the range of i3 loops of other M_5 receptors (220-229). The i3 loop of fugu putative M_2 receptor protein has 205 amino acids, about twenty-five amino acids more than the i3 loops of mammalian and chick M_2 receptors. The zebrafish M_2 receptor protein also has 201 amino acids (Hsieh and Liao, 2002). The larger i3 loops of these two M_2 receptors in fish mainly account for their larger size. Both these receptors contain 30-35 more residues than mammalian and chick M_2 receptors. Except for the first and the last 15-20

amino acids, there is no apparent homology in the remainder of the i3 loops of different subtypes. Even within the same subtype this domain is the most divergent. Among the five subtypes, the M_5 subtypes has been shown to share the least homology in the i3 region when comparisons are made between human and rat sequences (Bonner et al., 1988). But the i3 loop of bluegill putative M_5 receptor protein fragment shows equal and sometimes even higher homology with i3 loops of other M_5 receptors compared to the homology of the i3 loop of fugu putative M_2 receptor protein with i3 loops of other M_2 receptors.

Amino acids critical for G-protein coupling

As mentioned earlier the i3 loops of different subtypes are homologous only at their N-terminal (N-i3), proximal to TM5, and C terminal (C-i3), proximal to TM6, regions. These regions are involved in the activation of G-proteins (Wess et al., 1990; Kunkel and Peralta, 1993). Deletion mutants of M_1 and M_3 receptors, in which most of the i3 loop was deleted except the two portions proximal to the membrane, were found to be functional (Shapiro and Nathanson, 1989; Kunkel and Peralta, 1993). Using substitution mutagenesis studies, Kunkel and Peralta (1993) found that lysine-arginine-threonine-lysine-glutamic acid (KRTKE) motif at N-i3 is important for activation of signal transduction by M_3. Out of this motif arginine is conserved among all known muscarinic receptors and is important for signaling as shown by restoration of signaling ability of the M_3 receptor by restoring just arginine (Kunkel and Peralta, 1993). The N-i3 region of bluegill putative M_5 receptor fragment has this motif. The only difference is the conservative substitution of aspartic acid for glutamic acid. In Figure 14 (p. 60), this motif is boxed and shown in bold font. Apparently this motif is critical only for odd

muscarinic receptors, and even among odd muscarinic receptors it is conserved only in M_3 and M_5 subtypes. By employing random saturation mutagenesis technique, Burstein et al. (1995) and Hill-Eubanks et al. (1996) identified the critical amino acids for selectivity of G-protein coupling in the C-i3 and N-i3 regions of M_5 receptor, respectively.

The motifs isoleucine-tyrosine-threonine-arginine (IYTR) at N-i3 and lysine-alanine-alanine (KAA) at C-i3 were identified as functionally important in M_{odd} receptors. Further studies by Burstein et al. (1996, 1998) elucidated specific roles of these amino acids. Based on the results of point mutations of IYTR motif, Burstein et al. (1996) suggested that tyrosine in IYTR motif is the key determinant of G-protein coupling specificity. Burstein et al. (1998) proposed a model in which the i3 loop forms a G-protein coupling pocket. This pocket basically comprises N-i3 and C-i3 regions which form α-helical extensions of TM5 and TM6, respectively. These α-helical extensions start in the TM domains with hydrophobic amino acids and end in the portions of the i3 loop with basic amino acids proximal to the membrane. At N-i3, the α-helical extension of TM5 starts with leucine and tyrosine and ends with arginine, the conserved residue in the IYTR motif, while at C-i3 the α-helical extension of TM6 starts with leucine and ends with the lysine residue which is part of the KAA motif. The basic amino acids at the ends of α-helical extensions are predicted to make high affinity ionic interactions with G-proteins, thus recruiting them (Burstein et al., 1998). The protein-protein interaction between the receptor and the G-protein is proposed to be predominantly hydrophobic, and the hydrophobic amino acids of the α-helical extensions are thought to be involved in this interaction. Thus the G-protein coupling pocket has a positively charged lip and a

hydrophobic core. Both the motifs, i.e. IYTR and KAA, are present in the bluegill putative M_5 receptor protein fragment. In the fugu putative M_2 receptor, tyrosine and threonine are replaced by serine, a conservative substitution, and juxtaposed alanine residues in the KAA motif are substituted by valine and threonine, respectively. In Figure 14 (p. 60), the IYTR motif in bluegill putative M_5 receptor fragment and its corresponding motif in fugu putative M_2 receptor are shown in orange and green, bold fonts, respectively. The KAA motif in bluegill putative M_5 receptor fragment and its corresponding motif in fugu putative M_2 receptor are boxed in Figure 14.

The alignment performed in this study for phylogenetic analysis revealed that the tyrosine which is predicted by Burstein et al. (1996) to be the key residue for G-protein selectivity in odd numbered muscarinic receptors is replaced by serine in even numbered muscarinic receptors. Burstein et al. (1996) found that substituting tyrosine with serine in human M_5 receptor resulted in the greatest decrease in its function as determined by reporter gene assay. Thus serine might be the key determinant of G-protein coupling selectivity in even numbered muscarinic receptors. Replacement of threonine with serine in fugu putative M_2 receptor is a conservative substitution while replacement of alanine-alanine in KAA motif by valine and threonine is one conservative substitution and another non-conservative substitution. The first alanine residue in the KAA motif is thought to be involved in hydrophobic interaction with G-protein (Burstein et al., 1998) and so replacement of this residue with valine retains hydrophobicity. The second alanine residue in the KAA motif has been identified as another major determinant of G-protein coupling specificity in the odd numbered muscarinic receptors. It probably functions as a G-protein contact site that allows for efficient recognition of G-proteins by receptors

(Burstein et al., 1998). Burstein et al. (1998) found that mutation of this residue to the analogous residue in the even numbered muscarinic receptors, i.e. threonine, abolishes signaling by M_5.

Both, the bluegill putative M_5 receptor fragment and the fugu putative M_2 receptor have the conserved leucine-tyrosine sequence in TM5 just before the start of the i3 loop. In Figure 14 (p. 60), this motif is boxed and depicted in bold font. The α-helical extension of TM5 which extends into N-i3 starts with these residues (Burstein et al., 1998). The bluegill putative M_5 receptor fragment has leucine at the interface of C-i3 and TM6. The α-helical extension of TM6 which extends into the C-i3 region supposedly starts at this residue (Burstein et al., 1998). This residue is replaced by isoleucine in the fugu putative M_2 receptor, a conservative substitution. The leucine in bluegill putative M_5 receptor and isoleucine in fugu putative M_2 receptor are shown in blue and pink, bold fonts, respectively, in Figure 14. The contact between the muscarinic receptor and G-protein is through C-i3 region of the receptor and C-terminus region of Gα subunit of the heterotrimeric G-protein (Blin et al., 1995; Liu et al., 1995). Wess et al. (1997) have shown that the ability of M_2 receptor to interact with G_i protein specifically depends on the presence of a four amino acid motif, valine-threonine-isoleucine-leucine (VTIL), located at the i3 loop/TM6 junction. The fugu putative M_2 receptor has all these residues at the corresponding position, except leucine which is substituted by methionine. But point mutation studies of the residues in VTIL motif has shown that valine, threonine and isoleucine are engaged in specific interaction with G_i protein and contribute to the specificity and efficiency of receptor/G-protein coupling while leucine is not critical for determining the specificity of the interaction (Wess et al., 1997). Thus replacement of

leucine with methionine may not be of any consequence as far as G-protein coupling is concerned. The corresponding motif to VTIL in M_{odd} receptors is alanine-alanine-leucine-serine (AALS) located at the interface of i3 loop and TM6. The bluegill M_5 receptor fragment has this motif at the corresponding position. This motif overlaps with the KAA motif discussed earlier, i.e. these are the same alanine residues. The AALS motif plays the same role in M_{odd} receptors as that of VTIL in M_{even} receptors (Wess et al., 1997). The VTIL and AALS motifs are part of the α-helical extension of TM6, which extends into C-i3, in M_{even} and M_{odd} receptors respectively. In Figure 14, VTIM motif in fugu putative M_2 receptor is shown in pink, bold font while AALS motif in bluegill putative M_5 receptor fragment is shown in blue, bold font.

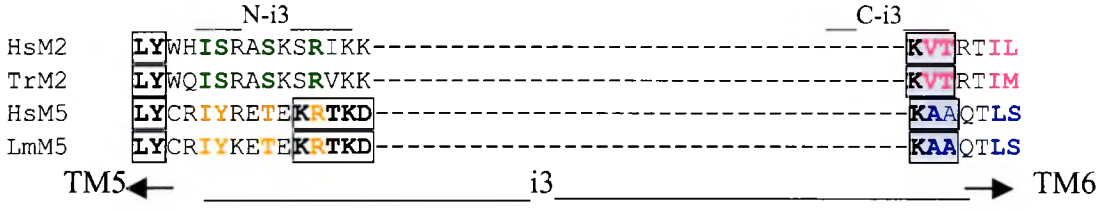


Figure 14: Alignment of deduced amino acid sequence of N-i3 and C-i3 regions of fugu putative M2 (TrM2) and bluegill putative M5 receptor fragment (LmM5). HsM2 and HsM5 represent human M2 and M5 receptors respectively. The dashes represent sequence of i3 loop which is not shown. The arrows represent end of i3 loop and start of Transmembrane domains. Critical amino acid motifs at the interface of TM5 and N-i3 and TM6 and C-i3 are shown in bold font and either boxed or shown in colored font or both. These motifs are critical for G-protein coupling. For details see text.

Muscarinic Receptor Phylogeny

Based on the results of molecular cloning muscarinic acetylcholine receptors are divided into five subtypes (M_1-M_5). Odd numbered receptors (M_1, M_3 and M_5) preferentially couple to the G_q family of G-proteins to mediate stimulation of phospholipase C while even numbered receptors preferentially couple to G_i/G_o family of G-proteins to mediate inhibition of adenylyl cyclase (Hsieh and Liao, 2002). The molecular and functional distinctions among receptors were reflected in the tree topologies obtained using both protein and nucleotide alignments employing different methods. In all the trees, receptors belonging to the same subtype formed a monophyletic clade and clades of odd numbered receptors formed one monophyletic unit while clades of even numbered receptors formed another monophyletic group. Except the most parsimonious tree obtained using protein alignment, in the rest of the tree topologies clades of M_1 and M_3 receptors formed one monophyletic group with the clade formed by M_5 receptors forming its sister group. Hulme et al. (1990) also noted that M_1 and M_3 protein sequences are more closely related to one another than to M_5. The grouping of M_1 and M_5 receptors in the same monophyletic unit in the most parsimonious tree obtained using protein alignment might have resulted due to the overall similarity in the three odd numbered subtypes at amino acid level. As the criterion in parsimony method is minimum changes, different grouping might result in closely related sequences. In any case, the M_1-M_5 group had poor bootstrap support.

The bluegill putative M_5 receptor fragment was grouped with other M_5 receptors and within that group it formed a terminal clade with zebrafish putative and fugu putative M_5 receptors. Fugu and bluegill receptors formed one monophyletic unit with zebrafish

receptor as their sister group. This grouping is consistent with the close taxonomic relationship between fugu and bluegill as both are placed under the same superorder *Acanthopterygii* while zebrafish belongs to *Neopterygii*. The fugu putative M_2 receptor was also found to be grouped with the other M_2 receptors and within that group it formed a terminal clade with zebrafish M_2 receptor. Both the terminal clades formed by the fish muscarinic receptors formed a sister group to the group formed by the rest of the receptors in their respective clades. All these clades had high bootstrap and Bayesian support.

Conclusion

I have isolated and sequenced bluegill putative M_5 receptor gene fragment from genomic DNA. I have also identified the coding strand of fugu putative M_2 receptor gene from fugu genomic database. Both the sequences are more identical and similar to known muscarinic receptor genes than to any other known gene. Upon virtual translation into amino acids, the bluegill putative M_5 receptor gene fragment shows more identity and similarity with the other known M_5 receptor proteins while fugu putative M_2 receptor gene shows more identity and similarity with the other known M_2 receptor proteins. Both the receptors have the critical amino acids shown by others to be required for ligand binding and G-protein coupling. In phylogenetic trees obtained under parsimony, maximum likelihood, Bayesian and distance criteria using protein and nucleotide alignment, the bluegill putative M_5 receptor fragment and fugu putative M_2 receptor were found to be grouped with their respective subtypes with high bootstrap and Bayesian support. All this evidence supports the identity of the sequences isolated from bluegill

genomic DNA and fugu genomic database as muscarinic receptor genes and furthermore establishes their identity as a bluegill putative M_5 receptor gene (incomplete fragment) and fugu putative M_2 receptor coding strand, respectively. Though it is not known whether these genes are expressed, this is a definitive step towards identifying muscarinic receptor genes in bluegill. The next step will be the sequencing of the complete coding strand of bluegill M_5 using reverse transcription-PCR (RT-PCR) and rapid amplification of complementary DNA ends (RACE) techniques. Identification of fugu M_2 coding strand will help in designing primers for amplification of M_2 coding strand from bluegill genomic DNA. Efforts are being made in that direction in our laboratory.

REFERENCES

Akiba, I., T. Kubo, A. Maeda, H. Bujo, J. Nakai, M. Mishina, and S. Numa. 1988. Primary structure of porcine muscarinic acetylcholine receptor III and antagonist binding studies. FEBS Lett 235:257-61.

Allard, W. J., I. S. Sigal, and R. A. Dixon. 1987. Sequence of the gene encoding the human M1 muscarinic acetylcholine receptor. Nucleic Acids Res 15:10604.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.

Ashkenazi, A., J. W. Winslow, E. G. Peralta, G. L. Peterson, M. I. Schimerlik, D. J. Capon, and J. Ramachandran. 1987. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. Science 238:672-5.

Baldwin, J. M. 1993. The probable arrangement of the helices in G protein-coupled receptors. Embo J 12:1693-703.

Baumgold, J., and P. H. Fishman. 1988. Muscarinic receptor-mediated increase in cAMP levels in SK-N-SH human neuroblastoma cells. Biochem Biophys Res Commun 154:1137-43.

Birdsall, N. J., N. M. Nathanson and R. D. Schwarz. 2001. Muscarinic receptors: it's a knockout. Trends Pharmacol Sci 22:215-19.

Blin, N., J. Yun, and J. Wess. 1995. Mapping of single amino acid residues required for selective activation of Gq/11 by the m3 muscarinic acetylcholine receptor. J Biol Chem 270:17741-8.

Bonner, T. I., N. J. Buckley, A. C. Young, and M. R. Brann. 1987. Identification of a family of muscarinic acetylcholine receptor genes. Science 237:527-32.

Bonner, T. I., A. C. Young, M. R. Brann, and N. J. Buckley. 1988. Cloning and expression of the human and rat m5 muscarinic acetylcholine receptor genes. Neuron 1:403-10.

Bonner, T. I. 1989. New subtypes of muscarinic acetylcholine receptors. Trends Pharmacol Sci Suppl:11-5.

Burnside, B., and B. Nagle. 1983. Retinomotor movements of photoreceptors and retinal pigment epithelium: mechanism and regulation. In Osborne, N., and G. Chader ed. Progress in Retinal Research, Pergamon Press, NY. 2:67-109.

Burstein, E. S., T. A. Spalding, D. Hill-Eubanks, and M. R. Brann. 1995. Structure-function of muscarinic receptor coupling to G proteins. Random saturation mutagenesis identifies a critical determinant of receptor affinity for G proteins. J Biol Chem 270:3141-6.

Burstein, E. S., T. A. Spalding, and M. R. Brann. 1996. Amino acid side chains that define muscarinic receptor/G-protein coupling. Studies of the third intracellular loop. J Biol Chem 271:2882-5.

Burstein, E. S., T. A. Spalding, and M. R. Brann. 1998. Structure/function relationships of a G-protein coupling pocket formed by the third intracellular loop of the m5 muscarinic receptor. Biochemistry 37:4052-8.

Caulfield, M. P., and N. J. Birdsall. 1998. International Union of Pharmacology. XVII. Classification of muscarinic acetylcholine receptors. Pharmacol Rev 50:279-90.

Chung, F. Z., C. D. Wang, P. C. Potter, J. C. Venter, and C. M. Fraser. 1988. Site-directed mutagenesis and continuous expression of human beta-adrenergic receptors. Identification of a conserved aspartate residue involved in agonist binding and receptor activation. J Biol Chem 263:4052-5.

Creason, S., K. M. Tietje, and N. M. Nathanson. 2000. Isolation and functional characterization of the chick M5 muscarinic acetylcholine receptor gene. J Neurochem 74:882-5.

Curtis, C. A., M. Wheatley, S. Bansal, N. J. Birdsall, P. Eveleigh, E. K. Pedder, D. Poyner, and E. C. Hulme. 1989. Propylbenzilylcholine mustard labels an acidic residue in transmembrane helix 3 of the muscarinic receptor. J Biol Chem 264:489-95.

Eglen, R. M., and S. R. Nahorski. 2000. The muscarinic M(5) receptor: a silent or emerging subtype? Br J Pharmacol 130:13-21.

Felder, C. C. 1995. Muscarinic acetylcholine receptors: signal transduction through multiple effectors. Faseb J 9:619-25.

Flynn, D. D., G. Ferrari-DiLeo, A. I. Levey, and D. C. Mash. 1995. Differential alterations in muscarinic receptor subtypes in Alzheimer's disease: implications for cholinergic-based therapies. Life Sci 56:869-76.

Fraser, C. M., F. Z. Chung, C. D. Wang, and J. C. Venter. 1988. Site-directed mutagenesis of human beta-adrenergic receptors: substitution of aspartic acid-130 by asparagine produces a receptor with high-affinity agonist binding that is uncoupled from adenylate cyclase. Proc Natl Acad Sci U S A 85:5478-82.

Fraser, C. M., C. D. Wang, D. A. Robinson, J. D. Gocayne, and J. C. Venter. 1989. Site-directed mutagenesis of m1 muscarinic acetylcholine receptors: conserved aspartic acids play important roles in receptor function. Mol Pharmacol 36:840-7.

Gadbut, A. P., and J. B. Galper. 1994. A novel M3 muscarinic acetylcholine receptor is expressed in chick atrium and ventricle. J Biol Chem 269:25823-9.

García, D. M., and B. Burnside. 1994. Suppression of cAMP-induced pigment granule aggregation in RPE by organic anion transport inhibitors. Invest Ophthalmol Vis Sci 35:178-88.

García, D. 1998. Carbachol-induced pigment granule dispersion in teleost retinal pigment epithelium. Cytobios 94:31-7.

Gether, U. 2000. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. Endocr Rev 21:90-113.

Gish, W., and D. J. States. 1993. Identification of protein coding regions by database similarity search. Nat Genet 3:266-72.

Gocayne, J., D. A. Robinson, M. G. FitzGerald, F. Z. Chung, A. R. Kerlavage, K. U. Lentes, J. Lai, C. D. Wang, C. M. Fraser, and J. C. Venter. 1987. Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors obtained by automated DNA sequence analysis: further evidence for a multigene family. Proc Natl Acad Sci U S A 84:8296-300.

Gonzalez, A. I., E. Crittenden, and D. M. Garcia. 2004. Activation of muscarinic acetylcholine receptors elicits pigment granule dispersion in retinal pigment epithelium isolated from bluegill. BMC Neuroscience. 5:23.

Growdon, J. H. 1997. Muscarinic agonists in Alzheimer's disease. Life Sci 60:993-8.

Hall, B. G. 2001. Phylogenetic trees made easy: A how-to manual for molecular biologists. Sinauer, Sunderland, MA.

Herrera, L., P. Carvallo, M. Antonelli, and J. Olate. 1994. Cloning of a *Xenopus laevis* muscarinic receptor encoded by an intronless gene. FEBS Lett 352:175-9.

Hibert, M. F., S. Trumpp-Kallmeyer, A. Bruinvels, and J. Hoflack. 1991. Three-dimensional models of neurotransmitter G-binding protein-coupled receptors. Mol Pharmacol 40:8-15.

Hill-Eubanks, D., E. S. Burstein, T. A. Spalding, H. Brauner-Osborne, and M. R. Brann. 1996. Structure of a G-protein-coupling domain of a muscarinic receptor predicted by random saturation mutagenesis. J Biol Chem 271:3058-65.

Hillis, D. M., B. K. Mable, A. Larson, S. K. Davis, and E. A. Zimmer. 1996. Nucleic Acids IV: Sequencing and cloning. In. Hillis, D. M., C. Moritz and B. K. Mable, ed. Molecular Systematics, Sinauer, Sunderland, MA. 2nd ed. 321-81.

Hsieh, D. J., and C. F. Liao. 2002. Zebrafish M_2 muscarinic acetylcholine receptor: cloning, pharmacological characterization, expression patterns and roles in embryonic bradycardia. Br J Pharmacol 137:782-92.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-5.

Hulme, E. C., N. J. Birdsall, and N. J. Buckley. 1990. Muscarinic receptor subtypes. Annu Rev Pharmacol Toxicol 30:633-73.

Hwang, J. M., D. J. Chang, U. S. Kim, Y. S. Lee, Y. S. Park, B. K. Kaang, and N. J. Cho. 1999. Cloning and functional characterization of a *Caenorhabditis elegans* muscarinic acetylcholine receptor. Receptors Channels 6:415-24.

Jones, S. V., C. J. Heilman, and M. R. Brann. 1991. Functional responses of cloned muscarinic receptors expressed in CHO-K1 cells. Mol Pharmacol 40:242-7.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275-82.

Jones, P. G., Curtis, C.AM., and Hulme, E.C. 1995. The function of a highly-conserved arginine residue in activation of the muscarinic M_1 receptor. Eur J Pharmacol 288:251-257.

Karnik, S. S., T. P. Sakmar, H. B. Chen, and H. G. Khorana. 1988. Cysteine residues 110 and 187 are essential for the formation of correct structure in bovine rhodopsin. Proc Natl Acad Sci U S A 85:8459-63.

Kubo, T., K. Fukuda, A. Mikami, A. Maeda, H. Takahashi, M. Mishina, T. Haga, K. Haga, A. Ichiyama, K. Kangawa, M. Kojima, H. Matsuo, T. Hirose, and S. Numa. 1986a. Cloning, sequencing and expression of complementary DNA encoding the muscarinic acetylcholine receptor. Nature 323:411-6.

Kubo, T., A. Maeda, K. Sugimoto, I. Akiba, A. Mikami, H. Takahashi, T. Haga, K. Haga, A. Ichiyama, K. Kangawa, H. Matsuo, T. Hirose, and S. Numa. 1986b. Primary structure of porcine cardiac muscarinic acetylcholine receptor deduced from the cDNA sequence. FEBS Lett 209:367-72.

Kubo, T., H. Bujo, I. Akiba, J. Nakai, M. Mishina, and S. Numa. 1988. Location of a region of the muscarinic acetylcholine receptor involved in selective effector coupling. FEBS Lett 241:119-25.

Kunkel, M. T., and E. G. Peralta. 1993. Charged amino acids required for signal transduction by the m3 muscarinic acetylcholine receptor. Embo J 12:3809-15.

Kurtenbach, E., C. A. Curtis, E. K. Pedder, A. Aitken, A. C. Harris, and E. C. Hulme. 1990. Muscarinic acetylcholine receptors. Peptide sequencing identifies residues involved in antagonist binding and disulfide bond formation. J Biol Chem 265:13702-8.

Liu, J., B. R. Conklin, N. Blin, J. Yun, and J. Wess. 1995. Identification of a receptor/G-protein contact site critical for signaling specificity and G-protein activation. Proc Natl Acad Sci U S A 92:11642-6.

Migeon, J. C., and N. M. Nathanson. 1994. Differential regulation of cAMP-mediated gene transcription by m1 and m4 muscarinic acetylcholine receptors. Preferential coupling of m4 receptors to Gi alpha-2. J Biol Chem 269:9767-73.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY.

Offermanns, S., T. Wieland, D. Homann, J. Sandmann, E. Bombien, K. Spicher, G. Schultz, and K. H. Jakobs. 1994. Transfected muscarinic acetylcholine receptors selectively couple to Gi-type G proteins and Gq/11. Mol Pharmacol 45:890-8.

Olianas, M. C., and P. Onali. 1991. Muscarinic stimulation of adenylate cyclase activity of rat olfactory bulb. II. Characterization of the antagonist sensitivity and comparison with muscarinic inhibitions of the enzyme in striatum and heart. J Pharmacol Exp Ther 259:680-6.

Peralta, E. G., J. W. Winslow, G. L. Peterson, D. H. Smith, A. Ashkenazi, J. Ramachandran, M. I. Schimerlik, and D. J. Capon. 1987a. Primary structure and biochemical properties of an M2 muscarinic receptor. Science 236:600-5.

Peralta, E. G., A. Ashkenazi, J. W. Winslow, D. H. Smith, J. Ramachandran, and D. J. Capon. 1987b. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. Embo J 6:3923-9.

Peralta, E. G., A. Ashkenazi, J. W. Winslow, J. Ramachandran, and D. J. Capon. 1988. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. Nature 334:434-7.

Posada, D., and K. A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817-18.

Savarese, T. M., C. D. Wang, and C. M. Fraser. 1992. Site-directed mutagenesis of the rat m1 muscarinic acetylcholine receptor. Role of conserved cysteines in receptor function. J Biol Chem 267:11439-48.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic anaysis using quartets and parallel computing. Bioinformatics 18:502-504.

Shapiro, R. A., N. M. Scherer, B. A. Habecker, E. M. Subers, and N. M. Nathanson. 1988. Isolation, sequence, and functional expression of the mouse M1 muscarinic acetylcholine receptor gene. J Biol Chem 263:18397-403.

Shapiro, R. A., and N. M. Nathanson. 1989. Deletion analysis of the mouse m1 muscarinic acetylcholine receptor: effects on phosphoinositide metabolism and down-regulation. Biochemistry 28:8946-50.

Strader, C. D., I. S. Sigal, M. R. Candelore, E. Rands, W. S. Hill, and R. A. Dixon. 1988. Conserved aspartic acid residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. J Biol Chem 263:10267-71.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0 Beta. Sinauer, Sunderland, MA.

Tatusova, T. A., and T. L. Madden. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247-50.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-82.

Tietje, K. M., P. S. Goldman, and N. M. Nathanson. 1990. Cloning and functional analysis of a gene encoding a novel muscarinic acetylcholine receptor expressed in chick heart and brain. J Biol Chem 265:2828-34.

Tietje, K. M., and N. M. Nathanson. 1991. Embryonic chick heart expresses multiple muscarinic acetylcholine receptor subtypes. Isolation and characterization of a gene encoding a novel m2 muscarinic acetylcholine receptor with high affinity for pirenzepine. J Biol Chem 266:17382-7.

van Koppen, C. J., W. Lenz, and N. M. Nathanson. 1993. Isolation, sequence and functional expression of the mouse m4 muscarinic acetylcholine receptor gene Biochim Biophys Acta 1173 (3):342-4.

Wess, J., T. I. Bonner, F. Dorje, and M. R. Brann. 1990. Delineation of muscarinic receptor domains conferring selectivity of coupling to guanine nucleotide-binding proteins and second messengers. Mol Pharmacol 38:517-23.

Wess, J., D. Gdula, and M. R. Brann. 1991. Site-directed mutagenesis of the m3 muscarinic receptor: identification of a series of threonine and tyrosine residues involved in agonist but not antagonist binding. Embo J 10:3729-34.

Wess, J., R. Maggio, J. R. Palmer, and Z. Vogel. 1992. Role of conserved threonine and tyrosine residues in acetylcholine binding and muscarinic receptor activation. A study with m3 muscarinic receptor point mutants. J Biol Chem 267:19313-9.

Wess, J., S. Nanavati, Z. Vogel, and R. Maggio. 1993. Functional role of proline and tryptophan residues highly conserved among G protein-coupled receptors studied by mutational analysis of the m3 muscarinic receptor. Embo J 12:331-8.

Wess, J. 1993. Molecular basis of muscarinic acetylcholine receptor function. Trends Pharmacol Sci 14:308-13.

Wess, J. 1996. Molecular biology of muscarinic acetylcholine receptors. Crit Rev Neurobiol 10:69-99.

Wess, J., J. Liu, N. Blin, J. Yun, C. Lerche, and E. Kostenis. 1997. Structural basis of receptor/G protein coupling selectivity studied with muscarinic receptors as model systems. Life Sci 60:1007-14.

Zhu, S. Z., S. Z. Wang, J. Hu, and E. E. el-Fakahany. 1994. An arginine residue conserved in most G protein-coupled receptors is essential for the function of the m1 muscarinic receptor. Mol Pharmacol 45:517-23.

Zinn, K. M., and M. F. Marmor. 1979. The retinal pigment epithelium. Harvard University Press, Cambridge, MA.

VITA

Prasad V. Phatarpekar was born in Savantwadi, Maharashtra, India, on October 20, 1972 to Padma V. Phatarpekar and Vijaykumar N. Phatarpekar. After graduating from Kalsulkar English School, Savantwadi, in 1988, he entered D.G. Ruparel College, University of Mumbai, Mumbai, India. He received the degree of Bachelor of Science in Zoology from University of Mumbai in April 1993. He attended Goa University, Goa, India, during 1993-1995 and earned a Master of Science in Marine Science in April 1995. During the following years he was employed as a research fellow with the National Institute of Oceanography, Goa, India. He enrolled as a graduate student in the Biology Department at Texas State University-San Marcos in Spring 2002. While at Texas State, he worked as an instructional assistant for Organismal Biology and Vertebrate Physiology and as a research assistant.

Permanent Address: 'Sneh,' Salaiwada, Savantwadi

Maharashtra 416510

India

This thesis was typed by Chhaya P. Phatarpekar.