Volume 255, number 2, 219-225 FEB 07614 September 1989

Cloning, sequence analysis and chromosome localization of a *Drosophila* muscarinic acetylcholine receptor

Takeshi Onai, Michael G. FitzGerald, Shoji Arakawa, Jeannine D. Gocayne, Deborah A. Urquhart*, Linda M. Hall*°, Claire M. Fraser†, W. Richard McCombie and J. Craig Venter

Section of Receptor Biochemistry and Molecular Biology, Laboratory of Molecular and Cellular Neurobiology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA and **Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA*

Received 3 August 1989

Two cDNA clones (3.7 kb and 4.8 kb) encoding a *Drosophila* muscarinic acetylcholine receptor were isolated from a *Drosophila* head cDNA library and characterized by automated DNA sequence analysis. The *Drosophila* muscarinic receptor contains 788 amino acids with a calculated M_r of 84 807 and displays greater than 60% homology with mammalian muscarinic receptors. The muscarinic receptor maps to the tip of the right arm of the second chromosome of the *Drosophila* genome.

Receptor evolution; Sequence homology; Signal sequence; cDNA; DNA sequence analysis, automated; Receptor supergene family

1. INTRODUCTION

Muscarinic acetylcholine receptors have been shown to be members of a supergene family that includes adrenergic receptors and opsins (see [1-3] for review). The most striking structural feature of all of these proteins is the presence of 7 regions of hydrophobic amino acids that are presumed to represent membrane spanning helices [2,3]. Based on gene cloning and sequence analysis, 5 distinct mammalian muscarinic acetylcholine receptor subtypes have been identified and expression studies have demonstrated that each of these receptor subtypes displays unique pharmacological and biochemical responses [4-12].

A number of studies have provided evidence for muscarinic acetylcholine receptors in lower phyla, particularly in molluscs and arthropods (see [13] for a review). Limited proteolysis of affinity labeled *Drosophila* muscarinic acetylcholine receptors suggested considerable conservation of receptor structure between mammalian and insect receptors [14,15]. Here we describe the cloning and sequence analysis of a cDNA coding for a *Drosophila* muscarinic acetylcholine receptor. As predicted from earlier biochemical, pharmacological and immunological data [14-17], this receptor exhibits a high degree of structural homology to mammalian muscarinic receptors.

Correspondence address: J.C. Venter, Section of Receptor Biochemistry and Molecular Biology, Laboratory of Molecular and Cellular Neurobiology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

° *Present address:* Department of Biochemical Pharmacology, School of Pharmacy, SUNY at Buffalo, Buffalo, NY 14260, USA

† *Present address:* Section of Molecular Neurobiology, LPPS, NIAAA, 12501 Washington Avenue, Rockville, MD 20852, USA

2. MATERIALS AND METHODS

2.1. *Library screening*

A cDNA library in lambda zap was obtained from Dr.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

Thomas Schwarz (Howard Hughes Medical Institute, Stanford University School of Medicine). The library was made from poly A^+ mRNA isolated from the heads of the Oregon-R strain of *Drosophila melanogaster*. The library was plated at a density of approximately 5×10^4 plaques per 150 mm plate (10^6 plaques total) using the bacterial host strain BB4 and was probed with a ^{32}P-radiolabeled 0.8 kb TaqI fragment of a rat M2 muscarinic acetylcholine receptor cDNA clone containing bases 22–856 which correspond to amino acids 2–286 [5]. The probe was radiolabeled using random primer extension to a specific activity of $>10^9$ dpm/μg DNA. Hybridization was carried out for 16 h at 42°C in a solution containing 6 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 5 × Denhardt's solution, 10 μg/ml sheared, heat-denatured salmon sperm DNA, 30% formamide (deionized) and 50 mM sodium phosphate, pH 7.4. Filters were washed with 6× SSC/0.1% SDS for 15 min at room temperature, 2× SSC for 15 min at room temperature, and 1× SSC for 15 min at room temperature plus 10 min at 42°C, dried and autoradiographed overnight at −70°C using Kodak X-Omat AR5 film.

2.2. *Preparation of cDNA clones for DNA sequencing*

Positive lambda zap clones containing cDNA inserts were amplified in vivo as plasmids as described by Stratagene, Inc. (La Jolla, CA). cDNA fragments were removed from each clone using *Eco*RI, the ends of each fragment were filled in using Klenow DNA polymerase and subcloned into the *Sma*I site of pVZ1 [18]. Insert orientations were determined by restriction digest analysis. Opposite orientations of each fragment in pVZ1 were isolated for use in DNA sequencing. Restriction digest analysis indicated that there were no internal Asp718 or SacI sites in the insert fragments. Therefore, these enzymes were used to linearize the pVZ1 subclones prior to exonuclease III digestion which creates a series of deletions for use in sequencing. Exonuclease III deletion generation was carried out according to the strategy of Henikoff [19]. At 30 s intervals, aliquots were removed from exonuclease III reactions and blunt ends were created with S_1 nuclease and DNA polymerase. The partially deleted plasmids were ligated and used to transform the bacterial host strain, BSJ72. Overlapping deletion subclones were amplified and the appropriate inserts were selected for sequencing based on their mobilities on agarose gels.

2.3. *DNA sequencing*

Single-stranded DNA was rescued from cells transformed with pVZ1 deletion plasmids by superinfection with M13KO7 (Promega, Inc., Madison, WI). The single-stranded DNA was used as a template for dideoxy sequencing as described by Sanger et al. [20] using modified T7 (Sequenase) polymerase and fluorescent dye primers with an Applied Biosystems 370A DNA sequencer [5].

2.4. *Chromosome localization*

The 4.8 kb *Eco*RI *Drosophila* cDNA insert was biotinylated using the BRL nick translation system and biotin-11-dUTP (Enzo Chemical). It was hybridized to larval salivary gland polytene chromosomes and detected using the BRL BluGENE detection system according to the procedure described by Engels et al. [21].

3. RESULTS AND DISCUSSION

3.1. *DNA cloning and sequencing*

A *Drosophila* head cDNA library was screened using a 0.8 kb TaqI probe derived from a cDNA clone encoding a rat M2 muscarinic acetylcholine receptor [5]. Initial positive clones were rescreened twice using the same probe. Two positive clones were excised from lambda zap as plasmids and characterized further. Restriction endonuclease digest analysis of these clones indicated that one contained a 3.7 kb *Eco*RI insert, the second contained a 4.8 kb *Eco*RI insert. The *Eco*RI inserts of the two clones were subcloned into pVZ1 for DNA sequence analysis which revealed that they were overlapping clones displaying significant homology with mammalian muscarinic acetylcholine receptors. The 3.7 kb cDNA clone lacked a complete 5′ coding region which was found in the 4.8 kb cDNA clone. The sequencing strategies for the 3.7 kb and 4.8 kb cDNA inserts are illustrated in fig.1A. The complete nucleotide sequence of these cDNA clones is illustrated in fig.1B.

The 4.8 kb *Drosophila* cDNA clone contains a single, large open reading frame that extends from an initiation codon (ATG) at bases 1–3 to a termination codon (TAG) at bases 2362–2364 (fig.1B) and codes for a protein of 788 amino acids with a calculated molecular weight of 84 807. There are, however, other possible candidate initiation codons within the 5′ sequence of this cDNA clone. Further experiments will be necessary to identify the actual initiation codon. The amino acid sequence of the *Drosophila* receptor shows a high degree of sequence homology with mammalian muscarinic acetylcholine receptors (fig.2), strongly suggesting that this cDNA codes for a *Drosophila* muscarinic acetylcholine receptor. The *Drosophila* muscarinic receptor displays approximately 60% homology with mammalian muscarinic receptors (M1: 69%, M2: 64%, M3: 62%, M4: 64% and M5: 64%) when the entire sequence is compared and favored amino acid substitutions are considered. The transmembrane domains have a higher degree of amino acid homology with a maximum of 88% in transmembrane domains III and IV (fig.2).

This *Drosophila* receptor contains several features that are characteristic of the muscarinic/adrenergic receptor supergene family. There are 3 potential sites for *N*-linked glycosyla-

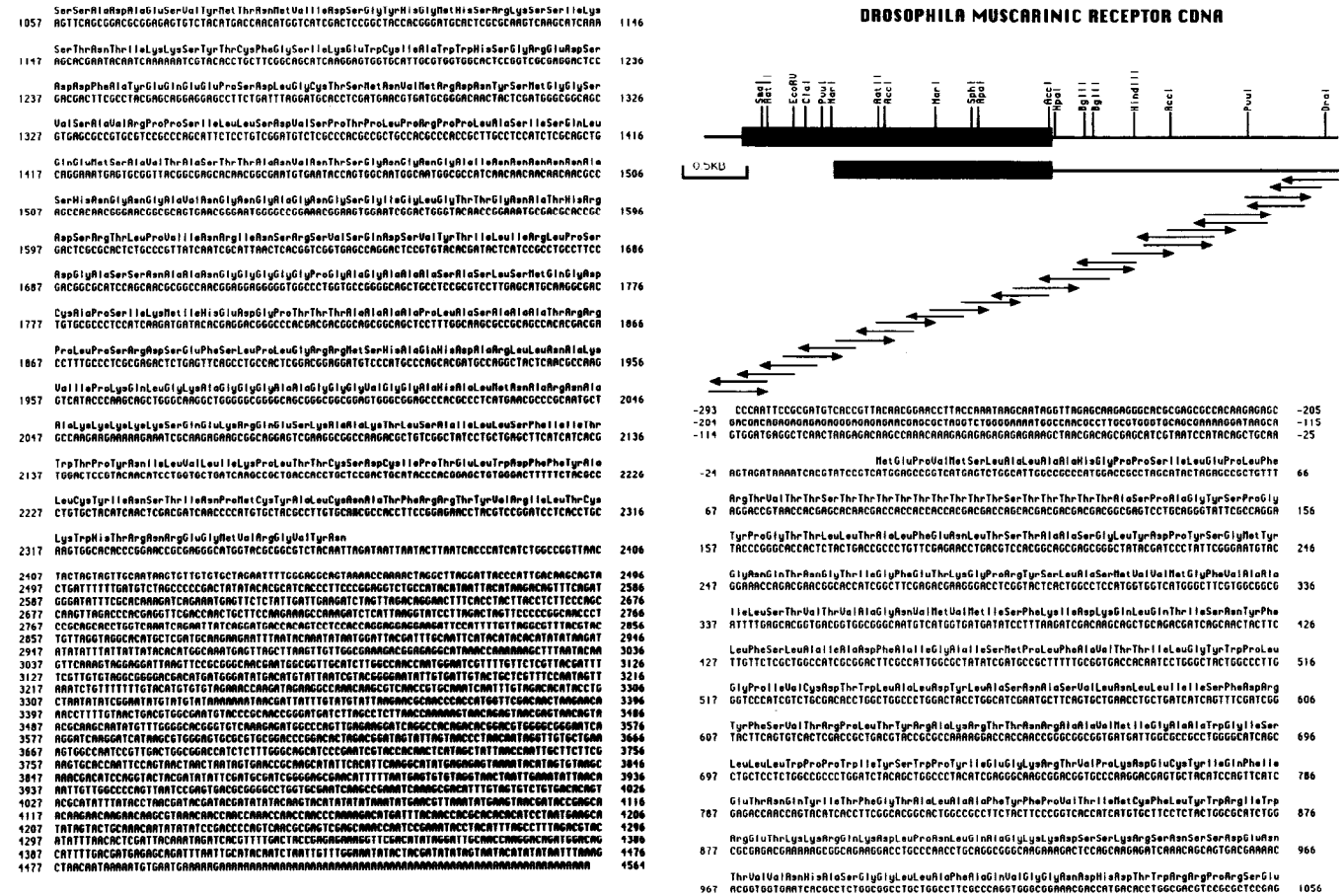

Fig.1. Sequencing strategy and primary structure of the cDNA encoding muscarinic acetylcholine receptor from *Drosophila*. (A) Schematic drawing and partial restriction endonuclease map of the *Drosophila* muscarinic receptor cDNA. Beneath is shown a schematic of the ordered deletion sequencing strategy. Arrows represent overlapping sequencing reactions and indicate the direction of sequencing; many of the fragments were sequenced more than once. (B) Nucleotide and deduced amino acid sequence of the cDNA encoding a *Drosophila* muscarinic acetylcholine receptor. Bases are numbered from the proposed initiation ATG of the open reading frame. The corresponding amino acid sequence of the *Drosophila* receptor is shown above the DNA sequence. This sequence was submitted to GenBank on 4-3-89 and assigned accession no. M23412.

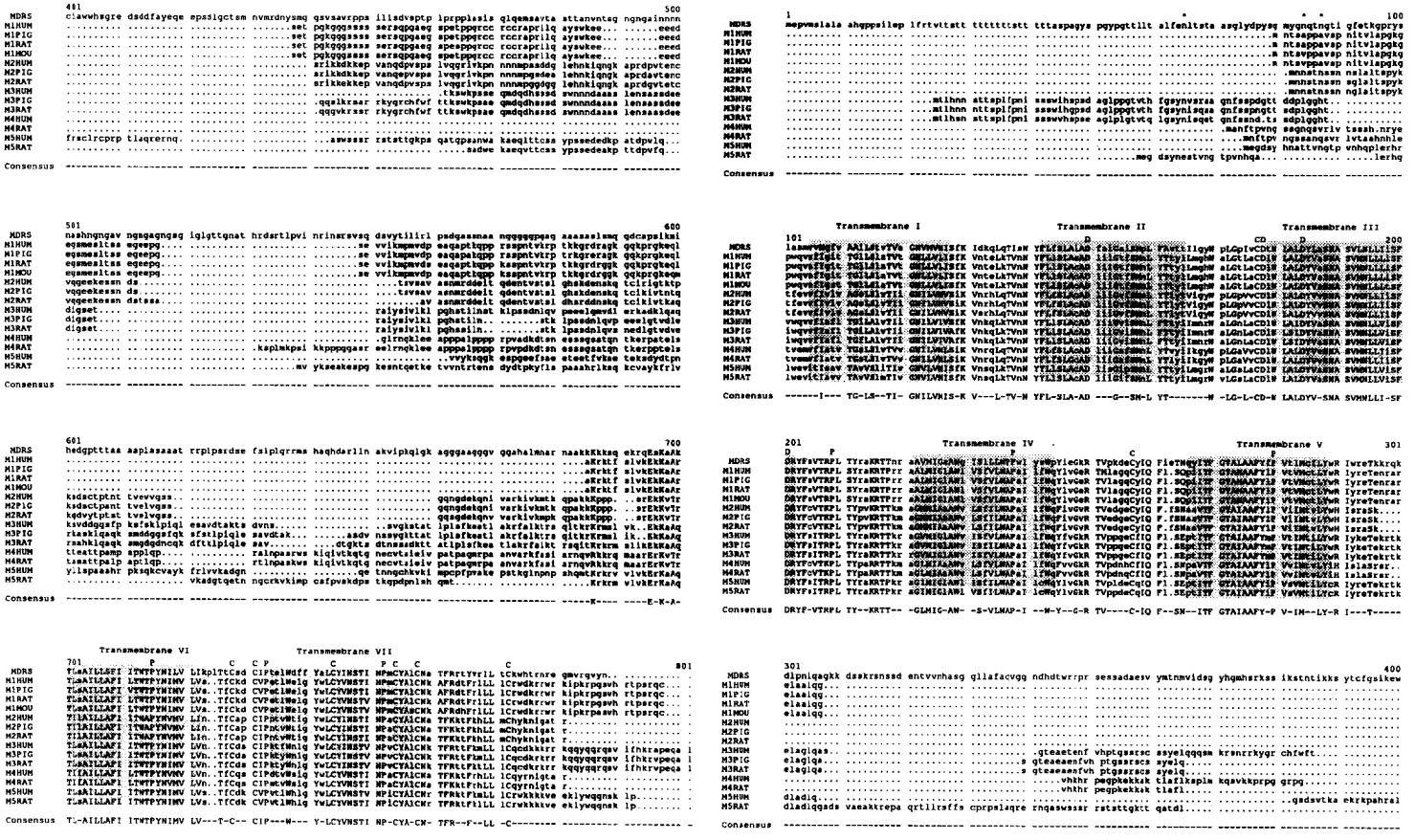

Fig.2. Comparison of the primary structure of *Drosophila* and mammalian muscarinic acetylcholine receptors. Sequences were aligned by sequential application of the GAP program in the UWGCG package [27] which uses the method of Needleman and Wunsch [28]. All conserved residues are indicated in the consensus line below the sequences, conserved aspartic acids, cysteines and prolines are indicated above the sequences, putative membrane spanning regions are shaded, putative sites for *N*-linked glycosylation are indicated by asterisks. Conservative amino acid substitutions, as defined by Dayhoff et al. [29] were considered to be equivalent. Data are from [4–12,30,31].

tion (Asn-X-Ser or Thr) as seen in the other cloned muscarinic receptors at Asn-65, -84 and -87 in the *Drosophila* sequence. Furthermore, the *Drosophila* protein contains 3 aspartic acids in the putative second and third transmembrane spanning helices that are conserved in all members of this gene family known to bind amine ligands [22,23]. All but one of the cysteine residues found in mammalian muscarinic receptors is conserved in the *Drosophila* muscarinic receptor.

The predicted hydropathy plot of the *Drosophila* receptor is illustrated in fig.3 and is compared to those of mammalian muscarinic acetylcholine receptors. As seen in mammalian muscarinic receptors, the *Drosophila* receptor contains 7 stretches of hydrophobic amino acids that most likely represent transmembrane spanning domains. In addition, the *Drosophila* receptor also contains a stretch of hydrophobic residues at its amino terminus that is similar in its hydropathy to signal se-

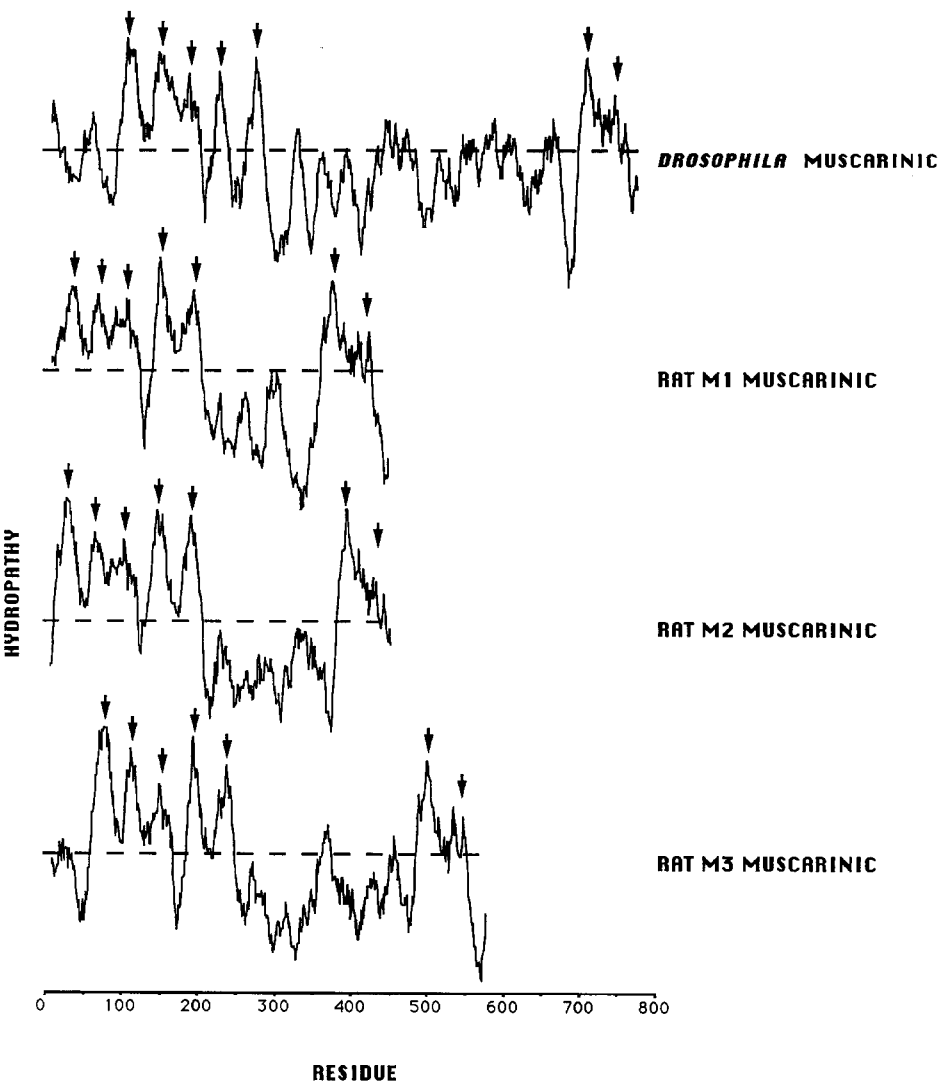

Fig.3. Hydropathy analysis of *Drosophila* and mammalian muscarinic acetylcholine receptors. Hydropathy plots of Kyte and Doolittle [32] were calculated using a window of 20 residues for the *Drosophila* muscarinic receptor or for rat M1–M3 muscarinic acetylcholine receptors [4,5]. Conserved hydrophobic regions presumed to represent transmembrane spanning domains are indicated by the arrows.

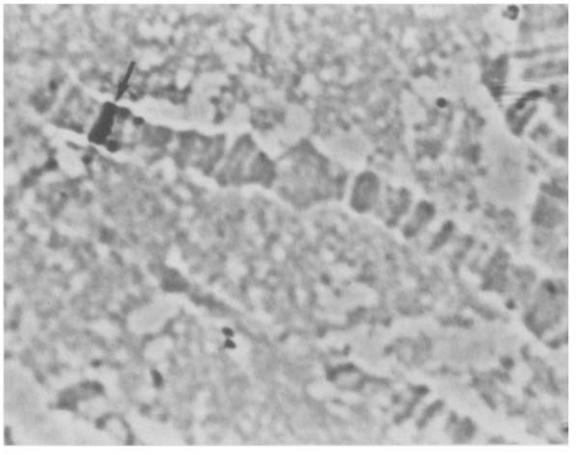

Fig.4. Chromosome localization of the *Drosophila* muscarinic receptor gene. In situ hybridization of the biotinylated cDNA insert to salivary gland chromosomes was detected by binding steptavidin-coupled alkaline phosphatase to the hybridized clone and then detecting the binding by the presence of a colored precipitate formed by the action of alkaline phosphatase. The position of the colored band is indicated by the arrow. Direct examination of the chromosomes showed that the hybridization was proximal (i.e., closer to the centromere) to the dark band which defines the start of the 60 Da region.

quences in other membrane proteins [24] but is not found in mammalian muscarinic receptors. Cleavage of the protein in this region would leave a mature protein containing glycosylation sites very close to the amino terminal end of the protein, as is seen in mammalian muscarinic receptors; however, there is no evidence as yet for such a cleavage. The *Drosophila* receptor also contains a significantly longer intracellular loop between putative transmembrane segments V and VI than do any of the mammalian receptors. The amino acid sequence of this region of the *Drosophila* muscarinic receptor shows little similarity to any of the corresponding sequences found in mammalian muscarinic receptors.

3.2. *Chromosome localization*

To determine where this putative muscarinic receptor maps in the *Drosophila* genome, the biotinylated cDNA was hybridized to *Drosophila* salivary gland polytene chromosomes. As shown in fig.4, the probe hybridized to a single site at the tip of the right arm of chromosome 2 in the region of 60C5-8. This muscarinic receptor gene maps to the right (i.e., is farther from the centromere) of the *seizure* locus which affects saxitoxin binding to the voltage-sensitive sodium channel [25]. The muscarinic receptor gene maps to the left of a gene cloned on the basis of its high homology to voltage-sensitive sodium channels [26]. Thus, there appear to be a number of interesting, neuronally expressed genes in this chromosome region. Studies are in progress to determine whether any of the known mutations which map to this general region represent the muscarinic receptor and whether this is the only gene coding for muscarinic receptors in *Drosophila*.

Acknowledgements: We thank Dr Anthony R. Kerlavage for generation of the hydropathy plots. This work was supported in part by NIH grant HL39369 to L.M.H. D.A.U. was supported by a New York City Heart Association post-doctoral fellowship award.

REFERENCES

[1] Nathanson, N.M. (1987) Ann. Rev. Neurosci. 10, 195-236.
[2] Schimerlik, M.I. (1989) Annu. Rev. Physiol. 51, 217-227.
[3] Kerlavage, A.R., Fraser, C.M. and Venter, J.C. (1987) Trends Pharmacol. Sci. 8, 426-431.
[4] Bonner, T.I., Buckley, N.J., Young, A.C. and Brann, M.R. (1987) Science 237, 527-532.
[5] Gocayne, J., Robinson, D.A., FitzGerald, M.G., Chung, F.-Z., Kerlavage, A.R., Lentes, K.-U., Lai, J., Wang, C.-D., Fraser, C.M. and Venter, J.C. (1987) Proc. Natl. Acad. Sci. USA 84, 8296-8300.
[6] Peralta, E.G., Ashkenazi, A., Winslow, J.W., Smith, D.H., Ramachandran, J. and Capon, D.J. (1987) EMBO J. 6, 3923-3929.
[7] Kubo,T., Fukuda, K., Mikami, A., Maeda, A., Takahashi, H., Mishina, M., Haga, T., Haga, K., Ichiyama, A., Kangawa, K., Kojima, M., Matsuo, H., Hirose, T. and Numa, S. (1986) Nature 323, 411-416.
[8] Kubo, T., Maeda, A., Sugimoto, K., Akiba, I., Mikami, A., Takahashi, H., Haga, T., Haga, K., Ichiyama, A., Kangawa, K., Matsuo, H., Hirose, T. and Numa, S. (1986) FEBS Lett. 209, 367-372.
[9] Allard, W.J., Sigal, I.S. and Dixon, R.A.F. (1987) Nucleic Acids Res. 15, 10604.
[10] Liao, C.-F., Themman, A.P.N., Joho, R., Barberis, C., Birnbaumer, M. and Birnbaumer, L. (1989) J. Biol. Chem. 264, 7328-7337.
[11] Bonner, T.I., Young, A.C., Brann, M.R. and Buckley, N.J. (1988) Neuron 1, 403-410.
[12] Peralta, E.G., Winslow, J.W., Peterson, G.L., Smith, D.H., Ashkenazi, A., Ramachandran, J., Schimerlik, M.I. and Capon, D.J. (1987) Science 236, 600-605.
[13] Venter, J.C., di Porzio, U., Robinson, D.A., Shreeve, S.M., Lai, J., Kerlavage, A. R., Fracek, S.P., jr, Lentes, K.-U. and Fraser, C.M. (1988) Progr. Neurobiol. 30, 105-169.

[14] Venter, J.C. (1983) J. Biol. Chem. 258, 4842–4848.
[15] Venter, J.C., Eddy, B., Hall, L.M. and Fraser, C.M. (1984) Proc. Natl. Acad. Sci. USA 81, 272–276.
[16] Dudai, Y. and Ben-Barak, J. (1977) FEBS Lett. 81, 134–136.
[17] Haim, N., Nahum, S. and Dudai, Y. (1979) J. Neurochem. 32, 543–552.
[18] Henikoff, S. and Eghtedarzadeh, M.K. (1987) Genetics 117, 711–725.
[19] Henikoff, S. (1987) Methods Enzymol. 155, 156–165.
[20] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[21] Engels, W.R., Preston, C.R., Thompson, P. and Eggleston, W.B. (1985) Focus 8, 6–8.
[22] Venter, J.C., Fraser, C.M., Kerlavage, A.R. and Buck, M.A. (1989) Biochem. Pharmacol. 38, 1197–1208.
[23] Strader, C.D., Sigal, I.S. and Dixon, R.A.F. (1989) FASEB J. 3, 1825–1832.
[24] Walter, P., Gilmore, R. and Blobel, G. (1984) Cell 38, 5–8.
[25] Jackson, F.R., Gitschier, J., Strichartz, G.R. and Hall, L.M. (1985) J. Neurosci. 5, 1144–1151.
[26] Salkoff, L., Butler, A., Wei, A., Scarvada, N., Giffer, K., Ifune, C., Goodman, R. and Mandel, G. (1987) Science 237, 744–749.
[27] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[28] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.
[29] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) In: Atlas of Protein Sequence and Structure (Dayhoff, M.O., ed) vol. 5, suppl. 3, pp. 345–352, National Biomedical Research Foundation, Silver Spring, MD.
[30] Shapiro, R.A., Scherer, N.M., Habecker, B.A., Subers, E.M. and Nathanson, N.M. (1988) J. Biol. Chem. 263, 18397–18403.
[31] Akiba, T., Kubo, T., Maeda, A., Bujo, H., Nakai, J., Mishina, M. and Numa, S. (1988) FEBS Lett. 235, 257–261.
[32] Kyte, J. and Doolittle R.F. (1982) J. Mol. Biol. 157, 105–132.