Volume 16 Number 17 1988

Nucleotide sequence of the rat guanidinoacetate methyltransferase gene

Hirofumi Ogawa and Motoji Fujioka

Department of Biochemistry, Toyama Medical and Pharmaceutical University, Faculty of Medicine, Sugitani, Toyama 930-01, Japan

Submitted July 29, 1988

Accession no. X08056

Several genomic clones for rat guanidinoacetate methyltransferase [EC 2.1.1.2] were isolated from a Charon 4A library (1) by plaque hybridization with a cDNA previously isolated (2). One of these clones had 4.3- and 7.5-kbp EcoRI fragments. The entire nucleotide sequence of the former (4321 bp) and the partial sequence at 5' side of the latter (EcoRI-KpnI segment, 643 bp) are determined on both strands by the method of Sanger et al. (3). The gene is found to consist of at least 6 exons and 5 introns, whose boundary sequences follow the gt..ag rule (4,5). There is no difference in the nucleotide sequences of the cDNA and the corresponding regions of the gene; this suggests the presence of a single gene for guanidinoacetate methyltransferase.

```
GAATTCTGAT CTCTGCTCTG CACACATTCA CTGCCCCTCT GCACAACACC CATGTACATT TAAAAAAAAT TAAGAAACAA ACCAAAGGGG TTGGGGATTT   100
AGCTCAGTGG TAGAGCGCTT GCCTAGGAAG CGCAAGGCCC TGGGTTTGGT CCCCAGCTCT GAAAAAAAGA ACCAAAAAAA AAAAAAAAGA GAGAGAAACA   200
AACCAAAAAC CCCAAAACAA AAGTGTTAAC TGTTTCAATT CTAAACTGGA GGCCCTGGTC TCCACTCCTG GTTTTTAGCA GTTTTCCCTG CCTGTAAAAT   300
GAGAGTACTA ACAGTTACTG CTGTAGCATC TTCTATGGGT TAAGAGTTTC ACTGTTCTTC TAGCAGTTCT CAGAAAAATG ATCATCTTCT AAAGGTTAAG   400
AGTTTCACTG TTCTTCTAGC AGTTCTCAGA AAAATGATCC GATCACTACA CCTAACCTGC CCAAGTGACC CAGGAGCTAC AGCGGGTGAA TGACAGGTGT   500
GGATTGGACT AGCTGGGAAG GACCGGCTTC TTGGACCTCA CCAGGCCAAC GAGTGTCTAA ACTTGAGGCA TCCAGCCAGG CATGGCGGCA CACGCCTTTA   600
ATCCCAGCAG TTCAGAGGTA GAATCAGACA GAACTCAGAG AAAGGGCCAT CTTGGCCTAC ACAGCAAGCT CCAGGCCAGC CAGGAGCTCC TGTCTCTAAA   700
TAAACAGATT AAAGAGAAAT AAAAAAAATC AACTTTTTGA TTTGAGGAGC ATCGTGGAGG GCAGCTAGGA TCAGCAAGTT TCACAGCCTC TTCAGACAGT   800
AAGTAGAGTA CCCTTGTCAG CCAGCCAAGC AAATGCAGAA AGTCAAGCTT CATTCCAAAT GTCTGGTACC AGGGAGCTTG CCCCGCACCT CGCAGAAGCC   900
ACAACAGTAC TTCTTTCGGT GGCACTGGTC CCACTGGTAG TCTCTAAGCG TCCCTCCCTA CACCGCACGC ACTTCGCCCC TCAAGGCTAC AACTTCCTGT  1000
CTAGATAACT AGGCTCCGCC CAGGCCCGTA GGCTCCGCCC CACTCATAAT AAAAGACTGG ATTCAGCTCC ACCCACGAAC AACTACCCCG GCCCTAGGCC  1100
GGCCTCCTAG GCGCCAACCT CACCCTAGCG GAAAGCTAGG CCGGGGCTCC AGTTTTGCCT CTACCAGGCT AGGCCCCACC CGTGGGAGAA GACTAGGCTT  1200
CGCCACTAGC AGAAGGCTTG GATTTCATCC TATCCCATAA ACACCTACCC CATGGCTCTG TCCTCATCCG TAGCAGAAAG CTGGACCATC CACCGTCCCA  1300
GCCGGCTGGC CCCGTTCCTT AGCTTCTTCT CTCATACTTA GCAGAAAGAT GACCCCACCC CATTGCTCCG GTCCCGCCCC TAGCAGAAGG CTGGACTCTG  1400
CCCTACCCTA CAGGCTCCAT CCTGTATGGC TCCTGCCTGC ATCCCTCATA CAAGTCTGGA TTCTGTCCAG CCCACAGGTC TGGCCCCATC CATAGCAAAA  1500
AGGTTGGACT CCTCCCCAGC CCACCACTAG GTATGCGACT CTAGAGATCG GGTACCGAGC TCGGTACCCG GGGTTCCTCT ACACCCCCCC TGCAGGACCT  1600
ATCCCATTCC ATAGATCCGC CCCTCACAGA ACTTTACTCC GGGTTCCAGC CAGTAGCCCC TACCCTGAGC ACGCTCCACC CTGCCCTCAG ACTCCATCCA  1700
CCCAGGCCCA GGCCCAGTAC TAGGCTGGCC ACCCTCCCGG ATCAAGGCCC AACCCTTTTC TAATGACTGG ACCCTCCCTA TCCCAGCAAG GCCCGCCCCA  1800
GCTAATAGGC CCCGCCACCC GGCCCAGCAC TGGACTCCCT CGGGGCACAC AGGTCCCGCC CCCGCCTCAG GCTCCCACCC ACCCCAGTGC AAGGCTAAAC  1900
TCCTCTTTGC AGAGGCGCCG CCCCACCCCG CCTCCTCCAG CCTGGTGGTT CCGCAGCCGT ACTCTCCTGG CCTGGTGTGC ACAGCCTCAC CATGAGTTCT  2000
                                           ▲                                                        MetSerSer
TCTGCAGCCA GCCCGCTTTT CGCGCCTGGC GAGGACTGCG GCCCCGCGTG GCGCGCGGCC CCCGCGGCCT ATGATACGTC TGACACGCAC CTGCAGATCC  2100
SerAlaAlaS erProLeuPh eAlaProGly GluAspCysG lyProAlaTr pArgAlaAla ProAlaAlaT yrAspThrSe rAspThrHis LeuGlnIleL
TGGGCAAGCC AGTAATGGAG CGTTGGGAGA CCCCCTACAT GCATTCGCTG GCGGCTGCTG CTGCCTCCAG AGGTACTCTG CCCAGAAGAA ACGGAGGCTA  2200
euGlyLysPr oValMetGlu ArgTrpGluT hrProTyrMe tHisSerLeu AlaAlaAlaA laAlaSerAr gG
CTGTTTAAGG AGTGTAGGAG TTCCCATGCA TGCTGCATGC CTGAGTTTCC TTACCTGAGG CACTGAGCTG CCTTGGGGTT ACCTTGGCAG GGAAGGGAAG  2300
AGTGTTGGCA AAATGTAGAC GCAGGCAAGA GAAGCCGGCT CCGAACTAGA ACTTCCTCTC CTGGGGAAAA AGACTTCTAC TTCCCGGAGC TGGGTTGGAG  2400
GGTGGATCTT CGGACCACAG CCTCGAGTGG AGCTTAGAGA ACATGGAGGG GCGTGGGACT TGTGCGGGTC GGGAACCTCC GAGTTTCCCC AGGCTGTTGC  2500
TAAGAATAGC TTGGGAAAAC GGAGCAGAGA GGCAGCTGTC CGGTGCTATG GCCCTGAACT TCTGGACGAG GGGTGGTGTC TGGGACACCC CAGCCTCAAA  2600
CAGGTCAATA GCCAAAGGGC TTCAAAGTTC ACCCGGAGCA CATGAGGACC TGGTATCAAG TGCGGAATGA AACATGGAGG GGTTACAGTG TCGGGCCTCA  2700
GAACCAGGCT TGGCCCTTTA TAACCCCACA TAGGGCCATA GCACAAATAG CGTGGGTGGG CATGGGCTCA GAGTCCAGGG ATCCCTGGGG CCAAGCCCAA  2800
AAGCACGTGG CCAGTCCCTC AGGCTCCCAG GCTCGCCGCT CTGATTGGCA GCTGGCCAGG GCCCACACTC TCTGTGCTAC TTTGTTCTGT CCACCCACTG  2900
CTGGCTGGCT GTCCTTGAGG CATGTGTGAG ACCGGTCCAC CCTAACCGTC CCCCTCCCCC ACCCATGCCA GCTGTAGTGT TTTTGTCCTC CTAGGGGGCC  3000
                                                                                                       lyGlyA
GGGTCCTGGA AGTGGGCTTT GGGATGGCCA TTGCAGCCTC CAGGGTGCAG CAGGCCCCCA TAAAGGAACA CTGGATTATT GAATGCAACG ATGGGGTCTT  3100
rgValLeuGl uValGlyPhe GlyMetAlaI leAlaAlaSe rArgValGln GlnAlaProI leLysGluHi sTrpIleIle GluCysAsnA spGlyValPh
CCAGCGTCTA CAAAACTGGG CCCTGAAGCA GCCACATAAG GTACCTTCTG TCACTGGAGT AGGAGGCTGG GAGTACCCCC CCGGGACTAG GGAGATCATC  3200
eGlnArgLeu GlnAsnTrpA laLeuLysGl nProHisLys
GGGGGTACCA ATGTGACCCC TTATATCCTT CTCTTGCTCC CTTCTCCTCT CCAAGGTTGT TCCCTTGAAA GGCCTGTGGG AGGAGGAGGC ACCTACACTG  3300
                                                            ValVa lProLeuLys GlyLeuTrpG luGluGluAl aProThrLeu
CCTGATGGTC ACTTTGATGT GAGGGCACTT TGGGAGCCTG GCAAGGGGGT TGGCCTCTGG GTACTGATTG GCAGCCTCTT GGGCTATCAT ACCCTCCTTA  3400
ProAspGlyH isPheAsp
GCTCAGTGGA GCTTGCCGTC AAGAATGAGG ACCAGAGTTT GATCCCCAGG ATGCATGTGG TGTAAGGAGA GAATGGGTGC TCATAAGTTG TCCCCTACCC  3500
CAACAAATGC ACCTCCAATG TGCACTGGGG ACATAGCCTC CGGGAAAAGG CTCCATGAAC TGGCTAGCAG AGGTGCTAAG GGCAGCTGGG CGGGCTGGGC  3600
AGTTTGGCCT CCAAGGCCAT GGGATAGGTG GGATGAAGCT TGGGCAAGGC TGTGCCAGCC CTGACTGGCA GGCCTGTGTT CACGCTGCAG GGGATTCTAT  3700
                                                                                                   GlyIleLeuT
ACGACACATA TCCACTGTCT GAAGAGACCT GGCACACTCA CCAGTTCAAC TTTATTAAGG TGGGACTGCA GGGCCGGGGA CCGAGGGCCT TGGACGGTCA  3800
yrAspThrTy rProLeuSer GluGluThrT rpHisThrHi sGlnPheAsn PheIleLys
GGTGGGCAGG GTGTGATATC TACAGACCAG CAGCTCTCTG CTCATGGTCC CTCCACCCCT TCTTGTCCAT GCCCTTTCTC TCTCTCTGCA GACTCATGCT  3900
                                                                                                    ThrHisAla
TTCCGTTTGC TGAAGCCTGG GGGTATCCTC ACTTACTGCA ACCTCACGTC CTGGGGGGAA CTCATGAAGT CCAAGTACAC AGACATCACT GCCATGTTTG  4000
PheArgLeuL euLysProGl yGlyIleLeu ThrTyrCysA snLeuThrSe rTrpGlyGlu LeuMetLysS erLysTyrTh rAspIleThr AlaMetPheG
AGGTACGCTG CCTCATAGCA AGTGGGGGCT CCGACCCAAA AGCCAGGCAG TGGCCCGGCA AGTGTGGTGT TGAGACCCAT GACATGGGGG AGAGGTCCAG  4100
lu
```

```
TCAGGGCTTC TCTGCAGCCC GAACCTACTG CTCTCTGCCT GTGGGGCGGT GCCTGAGAGA TGCCCACCAG GGCCATCAGG ACAGCTTCCC TTCAGCGCAT  4200
GAGGCAGCTC TTGGCACAGA AAACTCCTCT TGCTCGAGAG GCCTACAGCA GGCCTGTGCA AGGATGCTGG TCTGAGACAC AGCTGAGGTC ATCAGGAAAA  4300
CTCGTTGCCT GCATAGAATT CACACTCACA CTAGGCCCTG TCCCGCATCA TCAGTGGTGA CCTGTGACCC CTCCCTACAG GAGACTCAGG TGCCTGCACT  4400
                                                                                       GluThrGlnV alProAlaLe
GCTGGAAGCT GGCTTCCAGA GAGAAAACAT CTGTACAGAG GTGATGGCGC TGGTGCCCCC AGCCGACTGC CGCTACTATG CCTTCCCTCA GATGATCACA  4500
uLeuGluAla GlyPheGlnA rgGluAsnIl eCysThrGlu ValMetAlaL euValProPr oAlaAspCys ArgTyrTyrA laPheProGl nMetIleThr
CCCCTGGTCA CCAAGCACTG AGCGGCTGGC CCAGGGCTAC AAGGAGAATA TGTCCTCCTC AGTGCCTTTG TAGCTGGAGT GTGGGCTCCA GCCTCTCCAC  4600
ProLeuValT hrLysHis** *                                                                    ▼
TATCCCTGCA GTGTGACATC CTAACCTCTG CCTGGTACGG CCATCTCCCC AGAGCTCAGG AGTAAAATAA ATGCTACCAA GACTAGTAGT GGGTTGGCTG  4700
GCGGTGAGCC GGAGGTGAGG CCCGCATCTG GTTCAGAGCC TGCCTCTCCC TGCCTCCATA GATGTGGCTA TTCCATCACT GGTGCTCAGT GGCCCCCGAC  4800
CCAGACATGA CCAATCAACC ATTGTATGTT TTCAACAGTG GCCCAGGAAG AGCCTGGGGT CTCACCAGGG TCAGGGTAGG GAGAGCCTCC ACGGGTCCCA  4900
GCCCAGCACA CAGCAAGTCA TGCAACCCTC ATGGCACGAC CCTACTCTGA ACATCTCTGG TACC                                   4964
```

Fig. 1. Nucleotide sequence of the guanidinoacetate methyltransferase gene and its deduced amino acid sequence. The positions of the upward and downward arrowheads indicate the 5' and 3' ends of the largest cDNA clone (GAT9 (2)), respectively.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research (#62570102) from the Ministry of Education, Science, and Culture of Japan. We thank Tomoharu Gomi in our laboratory for expert technical assistance.

References

1. Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, R. B., Teyes, A. A. and Bonner, J. (1979) Proc. Natl. Acad. Sci. USA 85, 3255-3260
2. Ogawa, H., Date, T., Gomi, T, Konishi, K., Pitot, H. C., Cantoni, G. L., and Fujioka, M. (1988) Proc. Natl. Acad. Sci. USA 85, 694-698
3. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467
4. Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-383
5. Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472