

Nucleic Acids Research

2901 TCTATGAAGCCAGGCTCAAATTCGGATCTTTGCTGCCACCCCTCTAAATGCTGAGTTATAGTGTATAGGCTAGCCGATCTGGCTTTGTTCCGTTGTT
3001 TCACAGGATGACCTGGAAATTTGTATCTAGCTGGCCTTGAACCTGGTGATATATACCTCAACCTCTCAAGTGTCTGGGATTAAGGTATTCACCACCCT
3101 GCCTGGCTCAGACAGCTCTTAGCTGGGATTTCTATAGATGACAACTCTTGTGCACAGGTTGAAACATGTCTGCCTGTATATAGCTGGCCATGATCTGC
3201 CCTCACTGTGTCTCAGCTTCCCTGCTGGGATACAGGTATGGGTCACCTCTTGACATCTCTFAGACATCTCTTGGGGCTCTAGCTGGGATTA
3301 CTTCATTGTCTCTGGTTCTTTTAAATGCTTGGTTTTTGGAGCTGTCTAACTATGTAGACAGAGCTGGCCCTACTACTCGTGATTCCCTGCCCTCTG
3401 CCTTCAGGATGCTGGGATGAGAGGCTGTACCACAAATGATCCAGGGTAGGACTTTTGGTTTTGTTCTACTGTTTTTTTTTTTTTTTGGACAGGGTT
3501 CACTGTCTACTTTATGCTTGGTTAGCTGGCATTTCCTTTGTAGATCAGGGTGGCTTGAACCTATGATCTCTCTGCTTGGACCTCTGAGTGTCTGG
3601 ACTGCAAGGATGCATTCGCCACACATGCTATGATGTCCAGGGATGGAATGTCCACACAGGCAAGCACTCTTGAACATGTCTCCACAGCTGCCCTTT
3701 TGAATGGAGTCAGAATGGAGTATTCCTGTGTCTTGCCTCAGCCCTCCACAGCTGGGTGGTAACGCCATGCCACTCTATGGCTCTGGCTCTGGTT
3801 GCCTTCACAGGCTCAAGTTGATAGAGACCAGACTTCAGCTTACTAGTGGAGCTTCTGCCTCAGCCCTGTCTATCCGTCTTGGTGTCTGCCCA
3901 TGGTTGGATGTCTCTCTGTCTCAAGTGTAGGCCTCAGGGGTTAAGGAACAGCCAGCCTTTGGGGACCTTGGCTACTGAGTTATCCATCTCTGA
4001 GATCTCTGTAATAGGTAGGATATGATTCACACTGCCCCGGATGGTTTTATATGGAACAGCTGGAGTTTTCCCCACAGGcagggcaatgtgagcag
4101 ccacctatccagcccacctgggtccctccagagccactgtgctccagaggagaactggctctctctctctgcctgatggaggGTAGTCTAGAGGGC
4201 TGGAGGCATTGTAGGGGGAATGCAGATGTGCTAGGCTGTCCCTAGCCCTCTCTGTACCTTTCCAGaagaaactggaatactgagaaactcggctccacc
4301 tccacctgggagaggtgagccacctccagggcagaagtccagactggaagccacctccgctgcagctgtttgtggaccactcgtggccacgacctca
4401 cttgccaagccgaactcctccccctatccttcatcgtggactccaaggtGTAGGCTGGCTGCTTCTTGGAGCAGGGTTCTGTGATCCCTGGAACCTCTGTAGCAGGA
4501 TTAAGTATCTTATAGCATACATTAACAATGAATCGGGTGGCTTATGACACTTTCATATGTGTGCATTTTGGCCAGATACCCTCTTTATCTCTTTTC
4601 CTTGCTCTCCAGTTGATCTTTCTCTTCAACAAGTTTCTCTACATTTGTCCAAAAAATAAAAAAAAAAAAAAAAAAATCCTTGTCTCAGTCACTGTTGGG
4701 ATCACAGGATGAGCCACTATGCTGGAGTGAAGTGAATGAGCATGACACACTGTGAAAGAGGGAGCCCTCAGCTGGGCACTTCTGCAATCTAAGCA
4801 TGTAGGAGGAGCCAGCTTACTGGAGCAGTAATAGAGAGGCTGCCCTCAGCAGTTCACAGCAGTGGCTGCTCTTCTGGAGGACCAGGGTTCAATCCCCAGA
4901 CAGCTCACAATGCTGTAACTCCAGTTCAGGGGCTCCACACACACACATAGGCACCAACCAATGAACAAAAGTAAATGATTTTTAAATGTAAAC
5001 TTAATGCAAGCCTTAGCTGCTTCCCTCTGTCTGTCTGCCTCAGTCTCTGATCTCAGTCCCTTGGATTTCTTCACTTGGTTTCTTTTATTTTTTAAAT
5101 TTTAAATTTTTAAATTTTTTAAATTTTTTAAATTTTTTGGTTTTTGGAGCAGGGTTTTCTGTGATCCCTGGAACCTTCTTGTAGCAGCTTGTAGCAGGA
5201 TGGCCTCAAACCTCAGAAATCCGCTGCCCTGCCCTCCGAATGCTGGGATCAAAGGGTGGCCACCAGCCGGGCTCTACCTTGGTTTTTACCAGAGA
5301 TAACTGTGACCTTAATGTAAGCCGAGCAGTGTCTTCCCTGCAAGTGGTATGGGTGAGGTTTTATTTTGGTTTTTGGTTTTTAAAGCAGGGTT
5401 TCTCTGTAGCTTGGCTGTCTGGAAATTAAGTCTGTAGACCAGGCTGGCTCGAACCTCAGATGTCTGCTGCCCTGCCCTCCCAAGTGTGGGATTA
5501 AGACACGATTAACACCACTCAACTGCTATGGTCAATTTTTGGTTTTTGGAGCAGGGTTTTCTGTGATCCCTGGAACCTTCTTGTAGCAGCTTGTAGCAGGA
5601 ACTCAGTTTTATGGGTCAACTTTAAATCACAGCAATGGAAGCAAACCTAAGCACCTGGCTTCCAGTCCCAATGGCAATGAGATAGCTTTAGCTGTT
5701 ATTTTTTACTTCTCCCTTCCAAGTgacctgtggatggtctatctgagagctttccgacttccaagtccccagaccggccagagactctccagttc
5801 acggtggatgattccatttggccaacagctccagaactcagGTAAAGCCTTTCCACAGCCGCTGGCTCAGCTGAACTCAGTCTGCTCTGCTTCTGCTT
5901 CCTGCTCTCCACACTGTCTCCTTTAAGCAAAATGGCTTACCATGGTAGAAATTAATGAAAGTAGACAAAGAGCTATGGGTTCAGAGAGGGCCCTCAG
6001 GTAAGTATTTGGCTGCTTAAACATGAGAATGTTAAATAGGGGCTGGAGAGATTTGCTATAGTATTAGATACCAGCCGCTTTCTCAGAGGACCGA
6101 CGCATGGTGGCTCAGACAGCTGTAACCTCCGTTCTAAAGGATCCAATGTTCTTTTTGGCTCTCTTAAAGCAGGCAATGATGACAGCTGGACATACA
6201 TGCAGGCAAAACACACAGTAATAAAATTAATAAAGAAAGAAAGCTAAATAGCTGGCATACATACAGATAAATAATAATAAATAACAGCTGGCTGATGACAAA
6301 TTAATGTACGGCTCCAGAGTGTAGAGCTGGGGTGGGGTGGGTTCCGAGGATCTCTCTTCAAGCCAGCCAGGTTTACATAGTGTAGTCCAGGCTAG
6401 CCAAGGCAGCATAGTGTAGAGCTGTCTTCCCTCAGAACCCCAAAACAAGAAAGTAAAGTTAAGTTCTGAAGCTGAGGGCATGGCTCAACAGCTAAGA
6501 ACTCTTGAAGAGACTGACATCTAGTTCTAGGCCCACTGTCTGGGGCTGTGAGAGCTATCTCAAAGGTGAAAGCAATCCGGTGTGTAGAGCTCTGTGG
6601 AGTGCTGTGTAGACGGGCTCTAGGATTGACCCAGTATCTAAACATGAATCGCCCAAAAAATTAATTTTTTAAATTAACCTTTGCTTTAATATGTCTTT
6701 TTTCTGAGACAGGTTCTGCTATGTATCTGACAGCCAGAGCTGAACCTAAGTAGATTAGGCTGGCTTGAATTTGGAGATCTGCTGGGACTAAAGGC
6801 AGGTGGCACCCTGCCTTTTTTTTTTTTTTGGTTTTTGGAGCAGGGTTTCTGTGTAGCCCTGACTGTCTAGAACCTCACTCTGTAGACAGCCGCTG
6901 GCCTCAACTCAGAAATCGCCTGCCTGCTGCTGCTGCTGAGTGTGGGATTAAGGGTGTGCCACCACCCAGCCTTTTTTTTTTTTTTTTTTTTT
7001 TTTGAGACAGAACTCACTACTTTTATGGCTATGGCTGGCTGAAACTTGACTGTAGTCCATGCTGGCTTGAACCTGTATCTGCTGCCCTGCCCTCTG
7101 AATGCTGTGCCATCACATCTGGCTCTCCGCAAACTTTTCAAAAAGCTGTGCACACTCTTCCAACTCGGTCTGCTCACAACCTTCTTTCAGctcaat
7201 caactgcatctcaagtccgcccagctaaacagatccccgatgaagctcaacaaggctgtgctgctcaacaagactcccagagGTGAGGAGCAGGTTG
7301 CTTGCTGTAGCCACCCGGGCTACTCAAATGATTTCTCAATATACAATGGCAAACTCTCTGTGCTAGCTGAGCTAAGTAAGCTTTTGTCTT
7401 GTTACTCAGTtgggtgcccagtagaggggatgctgacatctgtgattgctgcagccatggcaactgtagtaactcaagctcttcacagttccagatccat
7501 ggaccccggcagtggtccaagctagttctcgaaccggcagggcagcgtATGTGCGGGGACCGGCACTTATCTCTCTCTTTAAGATCTCTGCTTC
7601 ACTTCTTATGTGCCATGAGATGGTGTGACTGAGGGAATGCTTCTGGGTCAAGTGTGCTATCTCAGCTAGGAGAGTGTCTGAGTGTCTCTCCAGTGG
7701 CTTGTGGGTTAGGTTAGCCCTGTCAATCCGTAGATTTCTCTCAAGACATTAAGGGTGAAGAGAGACTAGATGGATGATCAGGTGATACCTCAATCCAC
7801 TGATGCAGACAAATTTGAACTCTTGGGAAGCTGACTTACTGTAGAGCTGAGATGGCTTACATTTGCTATCCCGTCTCTGGGGAGCTTAAGAAAGA
7901 AATATAGTTCAAGGGCAGCTTGGGGGAGCCCTGTCTCAAGAAACCAACATGAAGGCTGGGAGAAAGGTTAGTGGTAGATAGAGCTCTGCTAGAAAT
8001 CCCCAGTGGAGGGTTGGGGTGTGGCTCAGTGGTCAAGCCCTGCTAGAAATCCCAAGGCTGGGAGAAAGGTTAGTGGTAGATAGAGCTCTGCTAGAAAT
8101 TAGAATCCCCAGTTAGGGGCTGGGGTGTGGCTCAGTGGTGAAGCCCTGCTAGAAATCCCAAGGCTGGGAGAAAGGTTAGTGGTAGATAGAGCTCTGCTAGAAAT
8201 TCTGCCATACTCTGAGGGCCTTGGTATGATTTCCAGCACTAAGCAAAAATTTCTCCATGTTCTGGTCTTTCTTAGTgacctgatgaagctgat
8301 gcaactgagggcccctgatattccttggaaagccacagcagctgtggaagctggactctgctcaaacctctggtgctctgggttagggctaggcc
8401 tggccacagtggaactcctgaccctggcagctatagctcctgctgtcaccaggaagtgctcactcctcttccactccttctgataccttccgatacagagaa
8501 gaaactcaGTTCTGAGTGTGTTAATGCTATGCTGTAGATGGGGCAGGGTGGGATGGGGTGGGATGGGTTGGTGGGATGGTGGGATGGTGGG
8601 TGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGG
8701 ATTGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGGTGGTGGGATGG

8801 GTGGTGGTTGCATTGGGGCAGGGTGAAGTGGACATCACAGTGCATGTGTGCTAATAGCGTAATGCAATAAGAGCAGATGGACCTTCAGTGAGCAGCCATC
 8901 TAGATGGCGAGAGGCTAAGCTGGTTGTAGAGACTTACAGAGACTTGTAGATCCTGCTTTGGGCAACATAGCTGCCCAAGTAGCCGTGTATGTGACT
 9001 CCAAGGAGATCAATATGGAGATGGGAGCTGATTGAAAGAGCTAGGTAGGCCAGCCAAACTAACTCTTAGGATATGAGGAATCTTGTAAACGCTGGCCC
 9101 CACGGATCTGACTCCAGTGGGATAAGGATCTGCGCTCTGCCTTATTAGTGTGGGATTTAAGATCTGCACCCTCTGCGCAGGCTGAACAAAGCCTT
 9201 TCTTTTTTTGAGACTGTGTAGAATCCAGGGCCTTGGCACATGCTAAGCAAAGTCTCCACCATGCTACATGGCCAGTCCCTTTCCCTTTGAGATAGCAATTT
 9301 CATGTAGCCAGCTAGACTCAGGCTTGGCATAACTGAGGACAACCTTGAAGTTCTGCTGCCCTTTGCTCTTCTCCCAAGAGCTAGACCAAGTGTTTCC
 9401 TTTCTAGCATGCTCCAGGCATGTACCACTAGGCTCAGCAGCCGCTCAGCCTTTCTAACAACTTGCTATGACACTCAGGGCTTTAATGGGCCAAATGGAT
 9501 GTTTGCCTGGTAGGCACAGTTGTAGTGTAAACCTATGGGAACAGAAAGATGCTGGGATAGAGCAGCTGGGAGAGGGGCACACTGGAGGCAGGTGCAT
 9601 GACTGGGCTTGGGCTGTAGTTAGGAGCAGAATGTTGCCTAGCATGCACAAGACTGATTTCCCGACTGTTTGAGATGAGGTCTCACGAACTAAGACTAG
 9701 TCTTAAATCCCAAGTCTCTCCGACCCCTCCCTCCAATCCATCTCCAAAGTGTAGTATTATAAACCTGTGTCATCACATGACTCTAGGTGTCTAAGTA
 9801 TATATAGAACAAGATTGATTTCTTGGTCAACTGAGGAGGGGAAGCCCTTGATACTCCAGTGTGTCAGCAATGTGATTTGGTAGCAAGTGTAGAGGGTG
 9901 TCAAAGATAGATCAGGCTGGGGAGATCTGAATTC

Fig. 1. The sequence of 10,833 nt at the ZP3 genomic locus of CD-1 mice. Sequence was obtained from a series of overlapping subclones and confirmed by analysis of both DNA strands. Eight exon sequences, encoding the entire ZP3 polypeptide chain (7), are denoted by lower case letters. Seven introns and the 3' and 5' flanking sequences are denoted by upper case letters. The putative transcription initiation site is designated +1, the 'TATA' box and polyadenylation signal are wavy-underlined, and the initiation and termination codons are boxed. A number of repeated sequences have been identified and are underlined. In the 5' flanking region between -826 and -508 nt are 6 copies of a 53 nt repeat; 4 copies of the same repeat also are found between 7,983 and 8,194 nt in the seventh intron. These repeats are more identical to each other (>75% homology) than those found at the ZP3 locus of Balb/C mice (8) and are present in the 5' flanking region of the ZP3 locus in hamsters (B. Ruiz-Seiler, R. Kinloch, P. Wassarman, unpublished results). In the 3' flanking region, between 8,554 and 8,805 nt, are 21 copies of a perfect 12 nt repeat. Eleven copies of this sequence have been found in the 3' flanking region of the ZP3 locus of Balb/C mice (8), and the sequence is present in the 3' flanking region of the ZP3 locus in hamsters and humans as well (B. Ruiz-Seiler, C. Fimiani, R. Kinloch, P. Wassarman, unpublished results).

References

1. Wassarman, P.M. (1987) *Science* 235, 553-560.
2. Wassarman, P.M. (1987) *Annu. Rev. Cell Biol.* 3, 109-142.
3. Wassarman, P.M. (1988) *Scientific Amer.* 255 (Dec.), 78-84.
4. Salzmann, G.S., Greve, J.M., Roller, R.J., and Wassarman, P.M. (1983) *EMBO J* 2, 1451-1456.
5. Wassarman, P.M., Bleil, J.D., Florman, H.M., Greve, J.M., Roller, R.J., Salzmann, G.S., and Samuels, F.G. (1985) *Cold Spring Harbor Symp. Quant. Biol* 50, 11-19.
6. Wassarman, P.M. (1988) *Annu. Rev. Biochem.* 57, 415-442.
7. Kinloch, R.A., Roller, R.J., Fimiani, C.M., Wassarman, D.A., and Wassarman, P.M. (1988) *Proc. Natl. Acad. Sci., USA* 85, 6409-6413.
8. Chamberlin, M.E. and Dean, J. (1989) *Devl. Biol.* 131, 207-214.