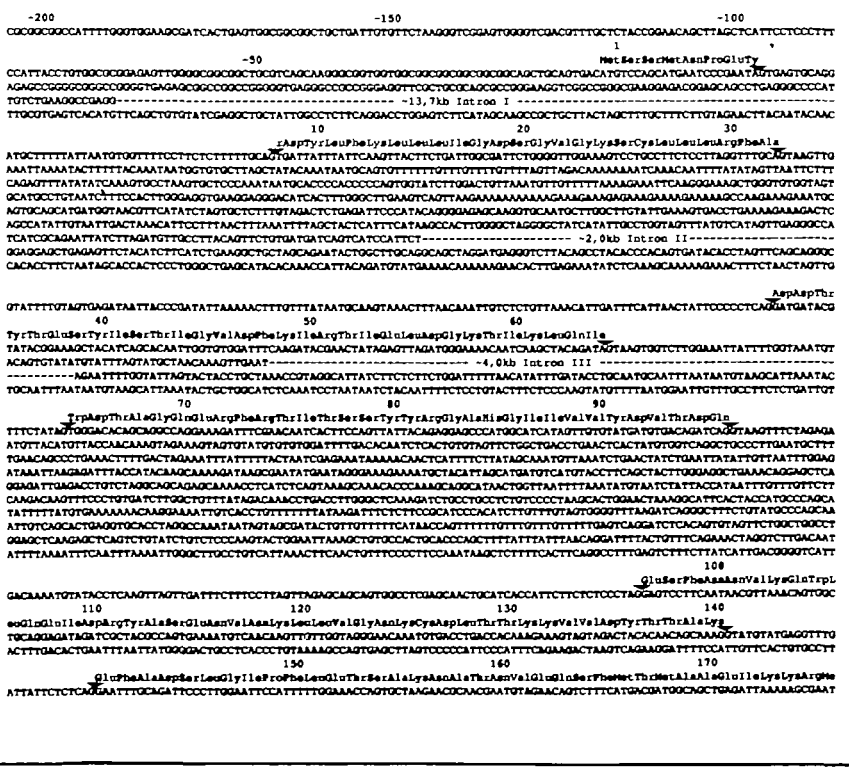
Nucleotide sequence of the mouse *ypt1* gene encoding a *ras*-related GTP-binding protein

Hendrik Wichmann, Christine Disela*, Heinz Haubruck and Dieter Gallwitz

Max-Planck-Institute for Biophysical Chemistry, Department of Molecular Genetics, PO Box 2841, D-3400 Göttingen, FRG

Submitted July 17, 1989 EMBL accession nos X15744–X15747 (incl.)

Mostly by cDNA cloning, well over 20 significantly related, GTP-binding proteins have been identified in mammalian cells. According to distinct structural features, these proteins, constituting the *ras* superfamily, can be separated into at least three families: *ras*, *rho* and *ypt* proteins (1). The genomic sequences of the H-, K-, N- and R-*ras* genes only have been elucidated (2,3,4,5). Here we present the genomic sequence of the mouse *ypt1* gene which was determined from 3 overlapping fragments of recombinant λ-phages identified by hybridization with the previously cloned mouse *ypt1* cDNA (1). The *ypt1* gene is roughly 24 kb long, with 5 introns of about 14 kb (I), 3 kb (II), 4.3 kb (III), 1187 bp (IV) and 149 bp (V) interrupting the protein-coding region of 205 codons. None of the intron positions matches those of the H-, K- and N-*ras* genes (3 introns) on the one hand or those of the R-*ras* gene (5 introns) on the other. Interestingly, the position of intron III perfectly matches the location of one of the three introns of the *ypt1* gene of the fission yeast *Schizosaccharomyces pombe* (6). A star above residue +682 signifies the beginning of the poly (A) tail of the 1.6 kb mRNA, the end of the 3.2 kb mRNA (1) has not been identified.



Downloaded from http://nar.oxfordjournals.org/ at MPI Study of Societies on January 4, 2017

```
        180                                 190                                 200                 205
tGlyProGlyAlaThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGlyGlyGlyCysCysSTP
GGGTCCTGGGACTACACGCTGGTGGTGCCGGAGAAGTCCAATGTTAAAATCCAGACACTCCAGTCAAGCAGTCAGGTGGGAGGCTGCTGCTAAAATCTGCCTCCGTCCTTTTCTCACAGCAA
                +50                                                 +100
TGAATTCGCAATCTGAACCCAAGTGAAAAAACAAAATTGCCTGAATTGTACTGTATGTAGCTGCACTACAACAGATTCTTACCGTTTCCACAAGGTCAGAGATTGTAAATGGTCAATACT
                                                        +200                                                        +250
GACTTTTTTTTTTATTCCCTTGACTCAAGACCGCTAACTTCAATTTCAGAACGTGTTTTAAACCTTTGTGTGCTGGTTTATAAAATAATGTGTGTAATCCTTGTTGCTTTCCTGATACCA
                        +300                                                        +350
GATCGTTTCCCGTGGTTGGTTAGAATATATTTTGTTTTGATGTTTATATTGGCATGTTTAGATGTTGGGTTTAGTCTTCTGAAGATGAAGTTCAGCCATTTTGTATCACACAGCACAACC
        +400                                                +450                                                        +500
AGTGTCTGTCAGTTTCCACGTCATAAAGTTTAGTGAGACGTATATGTAAGATTCGATTTGCTAGATCTTCCTGGTAGAGTTATAAAGGGACAGAGCACACTATCTGATTAATAGTTTCTT
                                +550                                                        +600
CATACTCTGCATATAATTTGTGGCTGCAGAATATTGTATTTGTTGCACACTATGTAACAAAACTGAAGATATGTTTAATAAATATTGTACTTATTGGAAGTAATATCAAACTGTATGGTG
                +650                                                +700
ATAAGTATTGTTTTAATTCCTATGATTTAAAGGGAAATAGAGCTTTGCATTCTACTCAGCAGCCACGTGTGTGCAGCAAGGGCACTGTAGACCTATCTTAATAGCACAGCATCCAGGGTTT
                                                        +800                                                        +850
GCTTTCTAGTAATAGGTCTGATGAGTGACCTAGAGAAGCTTTTGTGCTGTGTAGAGACTTCACCACCTTCATTGGTTCAAGGAACTCAAACTCTAGTGCTTCATCTCCAGCCCAGCCAC
                        +900                                                        +950
CTGAAGTGGCTTCACCAGCTACCCTTTTTTCCCTCACTCTGTGGTAGAGATGGAGAGCAGCCCATGCTTCCTCCCAGGCAGTGATCTGGTAATAGTGCTGGTGTTTCTTCATGCTCAGTC
        +1000                                               +1050                                                       +1100
TGATGGCGCACACACCTTTTTCTTTCTTTTTCTTTTTGGCTATTGGAGATAGGGTTTCTCTGTGTATCCCTGGCTGTCCTGGAACTCTCTCTGAGACCAGGCTGGCCTCGAACTCAGGG
                                        +1150                                                       +1200
ATCCACCTAGCTCTGTCTCTGAGTGCTGAGACTGAAAGGTGTGCGCCGCCACGCCTGGCTTTTTTCCCCTTTCTTTTTTGAGACAGGGTCTTACTATGTAGCCCTGGCTAGCCTGGAACTT
                +1250                                               +1300
AAGTATGTACAACAGGCTGCCTCCCCCTCCCCCATTCACCTGCCTTTTCCTCAAAAGTGCTGGGATTAAATGATGTCCCAGACTGCACCTATTTTTTGAGACAATGTCTCATATAGCTCA
                                                        +1400                                                       +1450
AGAGGCCTCAAATTCACTTGTAACCAAGGATGCCCTTGAACTCCTGATCCTCCTGCCTCCACTTCCAGGTGTCAACTTTATTCTTTCCAATTGTATGAATCATGTTGGTGGAGAGTTTAA
                                +1500                                                       +1550
GCATTAGAATAGAAAGCAGGAAATGTGATTGCCCTAATTGTAGAAATTCTATCTCCCCCCCCCCCAAATTAAATGAAAACATGAACGCTAGATGGGAATGATGGGAGAAATAATATATT
        +1600                                               +1650                                                       +1700
TATTTCGGTGAAATTTTGCATGTGTGAAAGTTTGTTAACTAGGATCTTCGAAATCCTAGACTTTGCTTTTGTCTTACTGTGGGAAACAAAGTCTCTTGCAGACACTAGCATTTTAAGGTG
                                        +1750                                                       +1800
ACAAAGGTCAGAGATTGTTTCCTTTGGTTTAAGGAGCAGCCAAGCATTCTTCAATTGGGGGCTAGGGTAGGGTATGGCTGTTGATAACACATTTTATCATGGGGTCTTTAATGATAAAAT
                +1850                                               +1900
ATTAAACTAAAAATAAGCCAGAAATTACAGGTTTATTTTTCACTAAACAGACAATTGAAATACATGGTACAAAAATAAGTGGTAAGATTATTGTAAATTGAAAAATGGACAGAATAAATA
                                                        +2000                                                       +2050
TTCAATTTAATTTTCCATCTATGAGAATTTCACAATAAAATCATAGTTTACTTTGTATTATAGATGCGCTTGTTGGGTCTATTCATCCTTATATATAAAAAGGTAACTGGAAACTCCTCA
                        +2100                                               +2150
AGTCACCAGGAATTGGGTCAATATCTTTTATTTTTTTGTAGCTGTTGGTTTTTTGTCAATATAAAGGGAGAAAGAGTTCATAAACTGTTTTCTATCATAACTAAATGATACAACACCAAT
        +2200                                               +2250                                                       +2300
ATAGGATGAAATGGTGCCAAGCACTTTAATTTTACACAGCAAAAGTTGTTCAAACTGTCAGCTTATTCTGTTAGCAGAAAACATTACATGAATTTATTTTTAATATAAACATATTTACAG
                                        +2350                                                       +2400
ACTAAAATTCTAGTGCTCTTTCCCCACTGGCAGCAAGATTAGGGCTTACCAGACTATTTCAGCATGCTTCACTATTTGGTTTCATGTATGTGCTGATAATTGCTGTTTTAGAGTGGGAAA
                +2450                                               +2500
TAACCCATTTTTCAAGTATAAATATTTCAAGACTCTTGGGGGCTCTAATATGTTCACATAATAGAAAACTGTCCCTTGATATATTTTAAAAATGACTGCCATAGAGAAACGTGCATAATG
                                                        +2600                                                       +2650
AAAAGCACTCTCTCAGTTGGCCACGGCTCCCCTAAGGTCCAGTGCTCTTTCGGACAGTTCCCTCAGGCTCATATTTGGCCACATCTTGTACCACGATGTCATTTTAAGTTTCAAAGCTGT

TTGAATTAT
```


ACKNOWLEDGEMENTS

We thank S. Ruppert and G. Schütz, Heidelberg, for providing the mouse genomic library and H.P. Geithe for synthesizing primers for sequence analysis.

*Present address: Research Institute of Molecular Pathology (IMP), A-1030 Wien, Austria

REFERENCES

1. Haubruck,H., Disela,C., Wagner,P. and Gallwitz,D. (1987) EMBO J., **6**, 4049-4053.
2. Capon,D.J., Chen,E.Y., Levinson,A.D., Seeburg,P.H. and Goeddel,D.V. (1983) Nature, **302**, 33-37.
3. McGrath,J.P., Capon,J.D., Smith,D.H., Chen,E.Y., Seeburg,P.H., Goeddel,D.V. and Levinson,A.D. (1983) Nature, **304**, 501-506.
4. Taparowsky,E., Shimizu,K., Goldfarb,M. and Wigler,M. (1983) Cell, **34**, 581-586.
5. Lowe,D.G., Capon,D.J., Delwart,E. Sakaguchi,A.Y., Naylor,S.L. and Goeddel,D.V. (1987) Cell **48**, 137-146.
6. Fawell,E., Hook,S. and Armstrong,J. (1989) Nucl. Acids Res., **17**, 4373.

Downloaded from http://nar.oxfordjournals.org/ at MPI Study of Societies on January 4, 2017