Sequence of cDNA for murine *Zfy*-1, a candidate for *Tdy*

Alan Ashworth*, Sally Swift and Nabeel Affara

Chester Beatty Laboratories, The Institute of Cancer Research, Fulham Road, London SW3 6JB and [1]Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
Submitted February 22, 1989 EMBL accession no. X14382

Sex in mammals is thought to be regulated by gene(s) present on the Y chromosome called the Testis Determining Factor (TDF in humans, Tdy in mice). Recently a candidate gene for human TDF has been isolated and shown to encode a zinc-finger protein (1). In the absence of definitive proof that this gene encodes TDF it has been named ZFY (2). A probe derived from the finger region of this gene (Affara, unpublished) was used to isolate from an adult mouse testis cDNA library (kindly supplied by K. Willison) a full-length cDNA clone for Zfy-1, one of the two Zfy homologues present on the mouse Y chromosome (3). The cDNA encodes a protein of 88,279 daltons the N-terminal 370 amino acids of which is extremely rich in acidic residues and may constitute a transcriptional activation domain (4). Residues 1270 - 1298 encode a string of basic amino acids that might be a signal for nuclear localization. The amino acid sequence of the zinc finger domain encoded by residues 1367-2501 is 79% homologous to that of the human gene (1).

```
GCATCAAACGGTCGTTACTCATTTTCAGGTGTTCTGGGTTTTCAGGCGTTCTGGGCTGTAAGAGACCGGCCCAGTACTCTGAAGTCTGCAGTACTTGTCG  100
TCATTGCCTACCCCTCTACACTGGTCTGGAGCTGACTTAGTAAGAACTGAAGGCCATGGATGAAGATGAAATTGAATCGACCCCAGAAGAAGAAAAGTCA  200
                                                    M  D  E  D  E  I  E  S  T  P  E  E  E  K  S
TTCTTTGATGGAATAGGAGCTGATGCAGTACACATGGATAGTGACCAGATTGTTGTGGAAGTACAAGAAACTGTTTTTTTTAGCTAATTCAGATGTAACTG  300
 F  F  D  G  I  G  A  D  A  V  H  M  D  S  D  Q  I  V  V  E  V  Q  E  T  V  F  L  A  N  S  D  V  T
TGCATAATTTTGTTCCTGATAATCCAGGCTCAGTTATAATTCAAGATGTTATTGAAAATGTTCTTATTGAAGATGTTCACTGTTCACATATTTTAGAAGA  400
V  H  N  F  V  P  D  N  P  G  S  V  I  I  Q  D  V  I  E  N  V  L  I  E  D  V  H  C  S  H  I  L  E  E
AACAGATATATCTGACAATGTCATTATTCCTGAGCAAGTTCTCAATTTAGGTACAGCAGAAGAAGTGTCTTTAGCACAGTTCTTAATTCCAGACATTTTA  500
  T  D  I  S  D  N  V  I  I  P  E  Q  V  L  N  L  G  T  A  E  E  V  S  L  A  Q  F  L  I  P  D  I  L
ACATCTGGTATTACGTCAACCTCATTGACTATGCCTGAACATGTCTTGATGAGTGAAGCTATACATGTGTCTGATGTAGGACATTTTGAACAAGTGATTC  600
  Y  S  G  I  T  S  T  S  L  T  M  P  E  H  V  L  M  S  E  A  I  H  V  S  D  V  G  H  F  E  Q  V  I
ATGATAGCCTTGTAGAAACAGAAGTCATCACTGATCCTATAACAGCTGACACTTCAGATATACTGGTAGCAGATTGTGTTTCTGAAGCAGTCTTAGATTC  700
H  D  S  L  V  E  T  E  V  I  T  D  P  I  T  A  D  T  S  D  I  L  V  A  D  C  V  S  E  A  V  L  D  S
CAGTGGGATGCCTCTGGAGCAGCAAGATAATGACAAAATCAACTGTGAGGATTATCTGATGATGTCTTTGGATGAACCTAGCAAAACAGATCTTGAAGGT  800
  S  G  M  P  L  E  Q  Q  D  N  D  K  I  N  C  E  D  Y  L  M  M  S  L  D  E  P  S  K  T  D  L  E  G
TCCTCTGAAGTTACCATGAATGCAGAGTCAGAAACTGATTCTTCTAAATTGGATGAAGCATCTCCAGAAGTTATCAAGGTGTGCATTCTTAAAGCTGACT  900
  S  S  E  V  T  M  N  A  E  S  E  T  D  S  S  K  L  D  E  A  S  P  E  V  I  K  V  C  I  L  K  A  D
CAGAAGTGGATGATGTAGGAGAAACTATACACGCAGTAGAGAGTGAGACCAAAAATGGCAATGAAGCTGAAGTCACAGATCAGAGCACTAGCATTCGTC  1000
S  E  V  D  D  V  G  E  T  I  H  A  V  E  S  E  T  K  N  G  N  E  A  E  V  T  D  Q  S  T  S  I  R  L
TCCCAGAGTCAACATTTATATGTCAGCCAGTGATTCTCAAAAGGAAGAAGAAGATACTGAAGTAATTGTAGGAGATGAAGATGCTGGTGGCACAGCTGCA  1100
  P  R  V  N  I  Y  M  S  A  S  D  S  Q  K  E  E  E  D  T  E  V  I  V  G  D  E  D  A  G  G  T  A  A
GATACTCCTGAGCATGAGCAACAGATGGATGTCAGTGAAATAAAAGCAGCTTTCCTACCTATTGCATGGACAGCAGCTTATGATAATAATTCTGATGAAA  1200
  D  T  P  E  H  E  Q  Q  M  D  V  S  E  I  K  A  A  F  L  P  I  A  W  T  A  A  Y  D  N  N  S  D  E
TTGAAGATCAGAATGTCACTGCCAGTGCTCTCTTAAACCAAGATGAGTCTGGTGGCCTTGACAGAGTACCAAAACAAAAATCAAAAAAGAAAAAAAGACC  1300
I  E  D  Q  N  V  T  A  S  A  L  L  N  Q  D  E  S  G  G  L  D  R  V  P  K  Q  K  S  K  K  K  K  R  P
TGAATCCAAACAGTACAAGTCAGCAATATTTGTTGCTCCTGATGGACAAACTTTACGTGTCTATCCTTGCATGTTTTGTGGGAAAAAATTTAAGACCAAA  1400
  E  S  K  Q  Y  K  S  A  I  F  V  A  P  D  G  Q  T  L  R  V  Y  P  C  M  F  C  G  K  K  F  K  T  K
AGGTTTTTGAAAAGACATACAAAAAACCATCCTGAATACCTTGCTAATAAAAAATATCACTGTACTGAGTGTGATTACAGTACCAACAAGAAGATAAGCT  1500
 R  F  L  K  R  H  T  K  N  H  P  E  Y  L  A  N  K  K  Y  H  C  T  E  C  D  Y  S  T  N  K  K  I  S
TACATAATCACATGGAGAGCCACAAGCTAACCATTAAGACAGAAAAGACCACCGAATGTGATGACTGTAGGAAGAATCTTTCTCATGCTGGGACTTTGTG  1600
L  H  N  H  M  E  S  H  K  L  T  I  K  T  E  K  T  T  E  C  D  D  C  R  K  N  L  S  H  A  G  T  L  C
TACTCACAAAACAATGCATACAGAAAAAGGAGTCAACAAAACATGTAAGTGTAAGTTCTGTGACTATGAAACAGCTGAACAGACATTATTGAATCACCAC  1700
  T  H  K  T  M  H  T  E  K  G  V  N  K  T  C  K  C  K  F  C  D  Y  E  T  A  E  Q  T  L  L  N  H  H
CTTTTGGTGGTCCACAGGAAGAAATTTCCTCACATTTGTGGAGAATGTGGTAAAGGTTTCCGTCACCCATCAGCACTCAAAAAGCACATACGAGTTCACA  1800
L  L  V  V  H  R  K  K  F  P  H  I  C  G  E  C  G  K  G  F  R  H  P  S  A  L  K  K  H  I  R  V  H
CAGGAGAGAAGCCCTATGAATGTCAGTATTGTGAGTACAAGTCTGCAGACTCTTCCAACTTGAAAACTCATATAAAATCTAAGCATAGTAAAGAGATACC  1900
T  G  E  K  P  Y  E  C  Q  Y  C  E  Y  K  S  A  D  S  S  N  L  K  T  H  I  K  S  K  H  S  K  E  I  P
ACTGAAGTGTGGCATCTGTCTCCTGACTTTCTCAGATACCAAAGAGGCTCAGCAACATGCCGTTCTGCACCAAGAAAGCAGAACACATCAATGTTCACAT  2000
  L  K  C  G  I  C  L  L  T  F  S  D  T  K  E  A  Q  Q  H  A  V  L  H  Q  E  S  R  T  H  Q  C  S  H
TGCAACCATAAGAGTTCAAACTCAAGTGATTTAAAGCGACACATAATTTCCGTTCACACAAAGGCGTATCCTCATAAATGTGACATGTGCAGCAAAGGAT  2100
 C  N  H  K  S  S  N  S  S  D  L  K  R  H  I  I  S  V  H  T  K  A  Y  P  H  K  C  D  M  C  S  K  G
TTCATAGGCCTTCAGAACTCAAGAAGCATGTGGCTACCCATAAAAGTAAAAAAATGCACCAATGTAGACACTGTGACTTTAATAGTCCAGATCCATTTCT  2200
F  H  R  P  S  E  L  K  K  H  V  A  T  H  K  S  K  K  M  H  Q  C  R  H  C  D  F  N  S  P  D  P  F  L
GCTTAGTCACCATATTCTCTCAGCTCACACAAAGAATGTTCCATTCAAGTGTAAGAGATGTAAAAAGGAATTTCAACAACAGTGTGAGCTTCAAACGCAT  2300
 L  S  H  H  I  L  S  A  H  T  K  N  V  P  F  K  C  K  R  C  K  K  E  F  Q  Q  Q  C  E  L  Q  T  H
ATGAAGACCCACAGTAGCCGAAAAGTCTATCAGTGTGAGTACTGTGAATATAGCACCAAGATGCCTCAGGTTTTAAGCGCCACGTTATCTCCATTCATA  2400
M  K  T  H  S  S  R  K  V  Y  Q  C  E  Y  C  E  Y  S  T  K  D  A  S  G  F  K  R  H  V  I  S  I  H
CGAAAGACTATCCTCACAGCTGTGACTTCTGCAAGAAAGGATTCCGGAGACCCTCGGAAAAGAATCAACACATAATGCGACATCATAAAGTTGGCCTGCC  2500
T  K  D  Y  P  H  S  C  D  F  C  K  K  G  F  R  R  P  S  E  K  N  Q  H  I  M  R  H  H  K  V  G  L  P
CTAAGTGGTATAATCCTAACATATAGGGATATTGGCATTGAGACAGTAAATTCATTTTAAAAGCAGGCATCCTTGTTCACATCCAATATCATGAATTTGT  2600
CAAGTGAATATAAAGATGATATTGCACCTATTTGACTTTATGTTTTATTAAATAGGGTATTCAATATATCCTTTACAGACTGACAAAGTGTGAGAAATAA  2700
ATTTTAACAACACpolyA
```

* To whom correspondence should be addressed

REFERENCES

(1) Page, D.C. et al. (1987) Cell 51, 1091-1104. (2) Page, D.C. et al. (1989) Phil. Trans. Roy. Soc. in press. (3) Nagamine, C.M. et al. (1989) Science 243, 80-83. (4) Ptashne, M. (1988) Nature 335, 683-689.