Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 2405–2408, April 1990
Cell Biology

Characterization of cDNA encoding a human sperm membrane protein related to A4 amyloid protein

YUAN CHANG YAN*, YUN BAI†, LINFANG WANG†, SHIYING MIAO†, AND SAMUEL S. KOIDE‡§

*Shanghai Institute of Cell Biology, Academia Sinica, Shanghai, China; †Institute for Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing, China; and ‡Center for Biomedical Research, The Population Council, 1230 York Avenue, New York, NY 10021

Communicated by Roy Hertz, January 2, 1990 (received for review November 9, 1989)

ABSTRACT **A rat testis λgt11 cDNA library was screened with a monoclonal antibody raised against a human sperm membrane protein designated YWK-II. A clone was found with a cDNA insert composed of 1837 base pairs that contained an open reading frame coding for 191 amino acid residues. The deduced polypeptide contained a segment with high homology to the transmembrane–cytoplasmic domains of the A4 amyloid protein found in brain plaques of Alzheimer disease patients. A sequence of basic amino acid residues, Arg-Lys-Arg, was found instead of Lys-Lys-Lys at the probable membrane–cytoplasmic junction that may be a unique property of sperm membrane proteins.**

In previous studies, monoclonal antibodies (mAbs) were raised against specific human sperm proteins to determine their role in spermatogenesis and their involvement in the causation of infertility (1, 2). One of the mAbs, designated YWK-II, interacted with 60- and 72-kDa proteins determined by Western blot analysis (1, 2). The interacting antigens were located at the midpiece, tail, and equatorial region of the human sperm head (1, 2) by use of an immunocytochemical method. Localization of the antigens to the equatorial region is of importance since this site may be involved in the adherence of the sperm to the egg during fertilization (3, 4). The YWK-II mAb blocked the penetration and fertilization of zona-free hamster eggs by human sperm (5). The antigen was also located on rat sperm head and germ cells of rat testis by use of an immunocytochemical method (6).

In the present study a 1.8-kilobase (kb) cDNA fragment encoding a sperm protein that interacted with the YWK-II mAb was isolated from a rat testis λgt11 cDNA expression library and its nucleic acid sequence was determined.¶ Data are presented showing high homology in the amino acid sequence of the deduced polypeptide with that of the transmembrane and cytoplasmic domains of the A4 amyloid protein found in brain plaques of Alzheimer disease patients (7–9).

EXPERIMENTAL

Epitope Selection with YWK-II Antibodies. A positive clone interacting with YWK-II mAb was isolated after screening a rat testis λgt11 cDNA expression library. The YWK-II mAb that eluted from positive clones on nitrocellulose filter paper interacted with human sperm proteins with estimated molecular masses of 60 and 72 kDa, as determined by Western blot analysis (data not shown).

Identification of the YWK-II cDNA as a 1.8-kb DNA Fragment. Phage DNA was isolated, cut with *Eco*RI, and analyzed by agarose gel electrophoresis. Insertion of 1.8-kb DNA was detected. The DNA fragment was inserted into pBR322, and the recombinant DNA was transfected into JM83 cells. When the recombinant pBR322 was cut with *Eco*RI and analyzed by agarose gel electrophoresis, the same 1.8-kb cDNA insert was identified (data not shown).

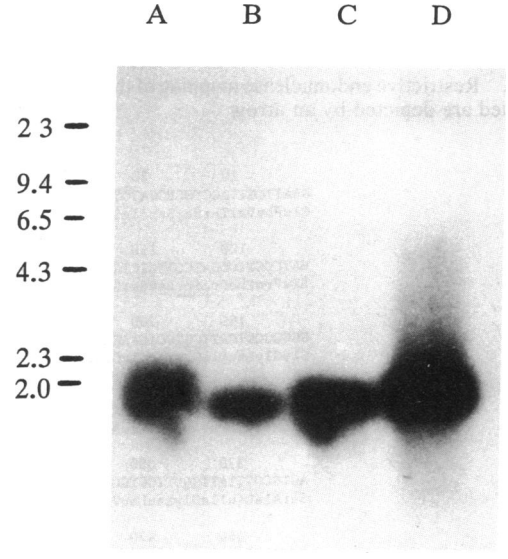

FIG. 1. Homogeneity of insert cDNA validated by Southern blot. The 1.8-kb cDNA fragment isolated from plasmid transformant was radiolabeled with [^{32}P]dATP by nick-translation and hybridized with the phage cDNA insert. Hybridization was performed overnight at 42°C. Autoradiography was carried out at −60°C for 6 hr. Lanes A–D are separate preparations of *Eco*RI-cut transformed phage cDNA insert. Sizes are given in kb.

Southern Blot Analysis. The 1.8-kb plasmid DNA fragment was radiolabeled by nick-translation using [^{32}P]dATP and hybridized with the phage DNA insert at 42°C for 12 hr. Autoradiography was performed at −60°C for 6 hr. Hybridization occurred indicating homology of the phage and plasmid DNA fragments (Fig. 1).

Sequencing of the 1.8-kb cDNA. The nucleotide sequence was determined according to Sanger's dideoxynucleotide chain-termination method (10) using the no. 409 M13 ^{32}P-sequencing pack purchased from New England Biolabs. For DNA sequencing, Sequenase was used according to the method published by United States Biochemical.

RESULTS

Restriction endonuclease analysis of the cloned 1.8-kb cDNA insert was performed (Fig. 2). By using *Pst* I, *Bgl* II, and *Pvu* II, 10 fragments were isolated. Nucleotide sequencing of these fragments showed overlaps of adjacent segments. The

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ORF, open reading frame; mAb, monoclonal antibody.

§To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M31322).

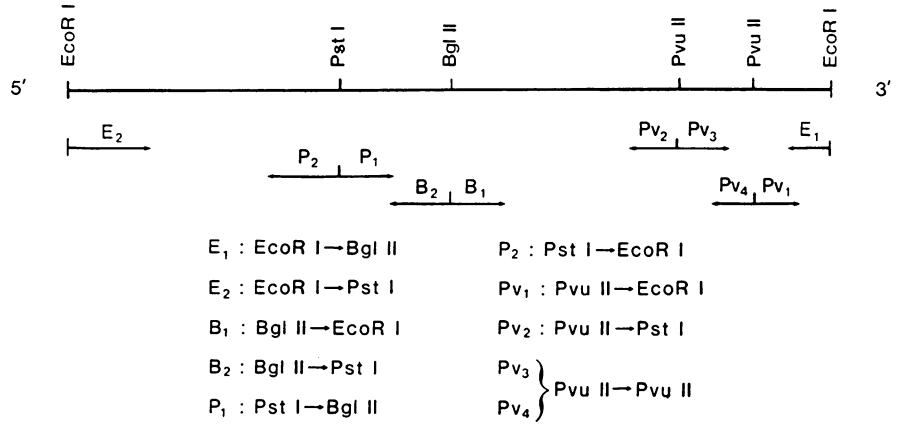

FIG. 2. Restrictive endonuclease mapping of the 1.8-kb DNA fragment and strategy for sequencing. The direction and length of each sequence determined are depicted by an arrow.

```
      10        20        30        40        50        60        70        80        90
GAATTCGTTACCTCCTCCATCTCAGAGAACCCTGTGGATGTCCGGGTGAGCTCTGAGGAAAGTGAGGAGATCCCCCCCGTTCCACCCTTTC
GluPheValThrSerSerIleSerGluAsnProValAspValArgValSerSerGluGluSerGluGluIleProProPheHisProPhe

     100       110       120       130       140       150       160       170       180
CATCCCTTCCCATCCTTGTCTGAGAACGAAGACACTCAGCCGGAGTTGTACCACCCAATGAAAAAAGGATCTGGAATGCCAGAGCAAGAT
HisProPheProSerLeuSerGluAsnGluAspThrGlnProGluLeuTyrHisProMetLysLysGlySerGlyMetAlaGluGlnAsp

     190       200       210       220       230       240       250       260       270
GGGGGCCTGATTGGTGCAGAAGAAAAGGTGATCAACAGTAAGAATAAAATGGATGAAAATATGGTCATTGACGAGACTCTGGATGTTAAG
GlyGlyLeuIleGlyAlaGluGluLysValIleAsnSerLysAsnLysMetAspGluAsnMetValIleAspGluThrLeuAspValLys

     280       290       300       310       320       330       340       350       360
GAAATGATTTTCAATGCTGAGAGAGTTGGGGGTCTGGAGGAAGAGCCGGACTCTGTGGGGCCTCTACGGGAGGACTTCAGTTTGAGCAGC
GluMetIlePheAsnAlaGluArgValGlyGlyLeuGluGluGluProAspSerValGlyProLeuArgGluAspPheSerLeuSerSer

     370       380       390       400       410       420       430       440       450
AGTGCCCTTATTGGCTTGCTGGTCATCCCGGTGGCCATTGCTACAGTCATCGTCATCAGCCTGGTGATGCTGAGGAAGAGGCAATACCGC
SerAlaLeuIleGlyLeuLeuValIleAlaValAlaIleAlaThrValIleValIleSerLeuValMetLeuArgLysArgGlnTyrGly

     460       470       480       490       500       510       520       530       540
ACCATCAGCCATGGCATTGTGGAGGTTCATCCAATGCTCACCCCAGAAGAGCGTCACTTGAACAAGATGCAGAACCACGGCTATGAAAAC
ThrIleSerHisGlyIleValGluValAspProMetLeuThrProGluGluArgHisLeuAsnLysMetGlnAsnHisGlyTyrGluAsn

     550       560       570       580       590       600       610       620       630
CCAACCTACAAATACCTGGAGCAGATGCAGATTTAAGGACAGCAGCGTGTGCGACACCCTGGCTGAGGCTGCTGCAGGTGGGCTGGAAGA
ProThrTyrLysTyrLeuGluGlnMetGlnIleEND

     640       650       660       670       680       690       700       710       720
GCCTCAGCGTTTGTGCTTGACTGCTGACCACCAGCGGTGCCAGAGGCCTCATCCTACATCCTGCTCTCCTGGATTGTTAAGACTATAAAG

     730       740       750       760       770       780       790       800       810
TACTACTGTAGGATTGCAATTTCCATTCTTTTAAATGGGTTTAAAAGATGTTAATATAACAATATATGATATATAAACCTTAAGTGAAAA

     820       830       840       850       860       870       880       890       900
AAAGATCTATTGCAGATATCTGATGGATGTAGTTTTCTTTTTTTAAATTAGAAATGCCACTTCTATTGTATTGTCTCACACATGCTCTAT

     910       920       930       940       950       960       970       980       990
ATAAATGGAAAATGTTGATTTTTCAATGATAGACTATATACACAGGCTGTTCCCGTTATGTAAGTCTGTTCTTTAGGCTCGTTTGCTGGC

    1000      1010      1020      1030      1040      1050      1060      1070      1080
CTGGTTTTGTCGTCATTTGTTTTAATGTATAAAGGCAGTATTCCCCTTTTCAGGTTGCTGAGAAATGTAAGTGGAACTGAAGTACATTGT

    1090      1100      1110      1120      1130      1140      1150      1160      1170
ATGCAGTTACTGACTGTTTTAGGCATAGTCTCCTTGGAAGCCTAGAGCTTCCAGTGCCGGGTGTCCAGTGCCTGTCACCAAAGCAAGGGC

    1180      1190      1200      1210      1220      1230      1240      1250      1260
TAAGTCACCTTGAGCTAGCTGGATGCAAACTAGATCCACTGTGCTTTCCTTCAAATCCAGTTCTTCCACAGCAACCAGCCCATAGTTGTT

    1270      1280      1290      1300      1310      1320      1330      1340      1350
CTGTGTTCTTCCACAGCTGTTTACGGTAGCCTCCTAGCCACTCTCCTCAGCAAGTGCATCCAAGAGTGCACCACCCCCTTCTTTGGACGT

    1360      1370      1380      1390      1400      1410      1420      1430      1440
CTCCGTCCCATGCACTGACCCTCTGCTTGCCTTCGTACCTCACTTCCTCCACCGCTCTTCAGCCCCTTTGATGTCCCCTCAGAGAATACC

    1450      1460      1470      1480      1490      1500      1510      1520      1530
GATATACACATGGCTAAGGACCCAGGAGACTTCACGGGAGGCCTCATTAGGTGAAAGGACGATGTTCTGGGCTGTACATGAAATTGGATC

    1540      1550      1560      1570      1580      1590      1600      1610      1620
TGTAGACACTGTGTTTCCTTCACTGACTTGTAATGTCACGCAGCTGGAGTTGATGCCACAACCCTTAGTGCTTTGTTGCTGTTTTGTTTT

    1630      1640      1650      1660      1670      1680      1690      1700      1710
TCAGGGTTCTGGTAACCTGCTACTGTTTTTGTTTTGGTTTTGGTTTGGTTTTTTTTGTATTTTTCTGTGATTTCCCTCCCCTTCCCCCCC

    1720      1730      1740      1750      1760      1770      1780      1790      1800
ATGCCTCTTCCCACTATGCACAGATGGAAACTTTACCTACAAACTCCTTCGTATGATCTGTGGAGAATGTACAGAACTTATTACATCAAT

    1810      1820      1830
AAAACACTTTAACTTCCCCCGAAAAAAAAAAGAATTC
```

FIG. 3. Nucleotide sequence and amino acid sequence of the ORF of the 1.8-kb cDNA fragment. Numbers refer to nucleotide positions. The deduced polypeptide is composed of 191 amino acids.

```
113   Arg  GluAsp  PheSer            Leu        SerSerSer
601   ArgHisAspSerGlyTyrGluValHisHisGlnLysLeuValPhePheAlaGluAspVal

122   Ala            LeuIleGlyLeuLeuValIleAlaValAlaIleAlaThrVal
621   GlySerAsnLysGlyAlaIleIleGlyLeuMetValGlyGlyValValIleAlaThrVal

137   IleValIleSerLeuValMetLeuArgLysArgGlnTyrGlyThrIleSerHisGlyIle
641   IleValIleThrLeuValMetLeuLysLysLysGlnTyrThrSerIleHisHisGlyVal

157   ValGluValAspProMetLeuThrProGluGluArgHisLeuAsnLysMetGlnAsnHis
661   ValGluValAspAlaAlaValThrProGluGluArgHisLeuSerLysMetGlnGlnAsn

177   GlyTyrGluAsnProThrTyrLysTyrLeuGluGlnMetGlnIleEnd
681   GlyTyrGluAsnProThrTyrLysPhePheGluGlnMetGln  Asn
```

FIG. 4. Abbreviated map of the amino acid residues of the deduced polypeptide and A4 amyloid protein. Note significant homology of the amino acid residues of the deduced polypeptide from positions 123 to 191 and the transmembrane–cytoplasmic domains (positions 627 to 695) of the A4 amyloid protein. ¦, Identical amino acids; :, related amino acids.

complete composite nucleotide sequence consisted of 1837 base pairs (bp) containing an open reading frame (ORF) of 573 bp that encodes 191 amino acid residues (Fig. 3).

Comparison of Homology of Amino Acid Sequence with A4 Protein. The amino acid sequence of the deduced polypeptide was compared with that of the A4 amyloid protein found in brain plaques of Alzheimer disease patients (Fig. 4) (7, 8). The overall matching of amino acid residues was 107 out of 696 of A4 protein, or about 15.4%. At positions 123–191 of the deduced polypeptide the matching of amino acid residues was 48 out of 68, or a homology of about 70.6%.

At amino acid residue positions 145–147, the positively charged sequence Arg-Lys-Arg was found in place of Lys-Lys-Lys, and at position 180–183, the sequence Asn-Pro-Thr was found to occur at the corresponding positions in the cytoplasmic domain of A4 protein (11, 12).

The N-terminal or extracellular domain did not contain the potential N-glycosylation sites of Asn-Xaa-Ser/Thr found in A4 protein (8, 11, 12); however, the sequence Asn-Ser-Lys-Asn at positions 71–74 might be considered as a possible glycosylation site. The N-terminal domain lacked the unusual sequence of seven threonines of A4 protein (8, 11); instead, a triple serine residue sequence occurred at positions 119–121, which might be potential phosphorylation sites.

Properties of the Coded YWK-II Protein. A hydropathy plot of the deduced amino acid sequence obtained by the method of Kyte and Doolittle (13) is depicted in Fig. 5. This protein is largely hydrophilic with a single highly hydrophobic region near the C terminus located at positions 120–145, corresponding to the transmembrane domain of A4 protein. High hydropathy value indicates probable association with lipids in the sperm membrane in an α-helical configuration (13).

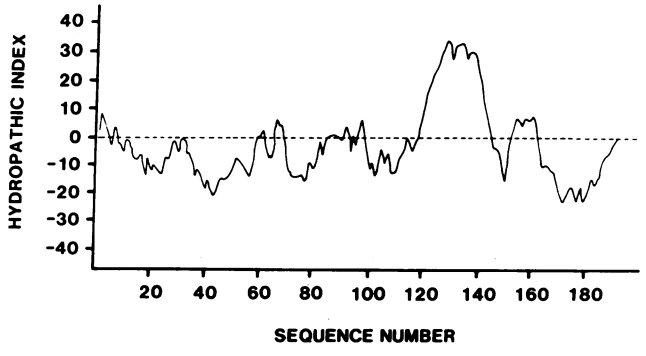

FIG. 5. Hydropathy plot of the amino acid sequence of the deduced polypeptide. Values were calculated according to Kyte and Doolittle (13). The hydrophobic region extends from positions 120 to 140.

DISCUSSION

The deduced polypeptide possesses partial sequence homology with the A4 amyloid protein found in brain plaques of Alzheimer disease patients. The region of high homology extends from position 123, includes the C terminus, and corresponds to the transmembrane–cytoplasmic domains, including the hydrophobic segment (Figs. 3 and 4). The hydrophobic region of the transmembrane domain is followed by the sequence Arg-Lys-Arg (Figs. 3 and 4) corresponding to Lys-Lys-Lys in A4 protein. This positively charged sequence usually occurs at the cytoplasmic–membrane junction of many membrane proteins and may interact with the membrane phospholipids or may act as a stop-transfer signal (14, 15). The extracellular region or N-terminal segment, however, shows no homology with A4 amyloid protein. This region in A4 protein contains a sequence corresponding to the serine protease inhibitor domain (16, 17). Although the functional role of the N-terminal domain of the sperm protein under study has not been established, this extracellular region might be involved in the fusion of the sperm and egg membranes during fertilization since this protein is located at the equatorial zone of the sperm head (1, 18), which is considered to be the site of sperm–egg interaction (3, 4).

Manning *et al.* (19) identified mRNA species encoding the A4 amyloid protein in nonneuronal tissues, including the testis, by hybridization with the A4 cDNA probe isolated from a brain expression library. Positive reaction indicates high homology of the cDNA encoding the A4 protein and the mRNA prepared from nonneuronal tissues. In the light of the present results it should be pointed out that polypeptides having partial homology to the A4 amyloid protein, especially in the transmembrane–cytoplasmic domains, may occur as structural components of various cells and tissues.

This study was supported by grants from the Rockefeller Foundation and Mellon Foundation.

1. Yan, Y. C., Wang, L. F. & Koide, S. S. (1987) *Arch. Androl.* **18**, 245–254.
2. Yan, Y. C., Wang, L. F., Sato, E. & Koide, S. S. (1983) *Am. J. Reprod. Immunol.* **4**, 111–115.
3. Shalgi, R. & Phillips, D. (1980) *J. Ultrastruct. Res.* **71**, 154–161.
4. Bedford, J. M., Moore, H. D. M. & Franklin, L. E. (1979) *Exp. Cell Res.* **119**, 119–126.
5. Wang, L. F., Yan, Y. C., Miao, S. Y. & Koide, S. S. (1987) in

New Horizon in Sperm Cell Research, ed. Mohri, H. (Jpn. Sci. Soc., Tokyo), pp. 409–420.
6. Haneji, T. & Koide, S. S. (1987) *Biol. Reprod.* **37,** 467–477.
7. Glenner, G. G. & Wong, C. W. (1984) *Biochem. Biophys. Res. Commun.* **120,** 855–890.
8. Kang, J., Lemaire, H. G., Unterbeck, A., Salbaum, J. M., Masters, C. L., Grzeschik, K. H., Multhaup, G., Beyreuther, K. & Muller-Hill, B. (1987) *Nature (London)* **325,** 733–736.
9. Hyman, B. T., Van Hoesen, G. W., Beyreuther, K. & Masters, C. L. (1989) *Neurosci. Lett.* **101,** 352–355.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
11. Robakis, N. K., Ramakrishna, N., Wolfe, G. & Wisniewski, H. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4190–4194.
12. Yarden, Y., Escobedo, J. A., Kuang, W.-J., Yang-Feng, T. L., Daniel, T. O., Tremble, P. M., Chen, E. Y., Ando, M. E., Harkins, R. N., Franke, V., Fried, V. A., Ullrich, A. & Williams, L. T. (1986) *Nature (London)* **323,** 226–232.
13. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
14. Blobel, G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1496–1500.
15. von Heijne, G. (1980) *Eur. J. Biochem.* **120,** 275–278.
16. Tanzi, R. E., McClatchey, A. I., Lamperti, E. D., Villa-Komaroff, L., Gusella, J. F. & Neve, R. L. (1988) *Nature (London)* **331,** 528–530.
17. Kitaguchi, N., Takahashi, Y., Tokushima, Y., Shiojiri, S. & Ito, H. (1988) *Nature (London)* **331,** 530–532.
18. Yan, Y. C., Wang, L. F., Mitsudo, S. M. & Koide, S. S. (1986) in *Immunological Approach to Contraception and Promotion of Fertility*, ed. Talwar, G. P. (Plenum, New York), pp. 231–240.
19. Manning, R. W., Reid, C. M., Lampe, R. A. & Davis, L. G. (1988) *Mol. Brain Res.* **3,** 293–298.