

# Identification and differential expression of yeast *SEC23*-related gene (*Msec23*) in mouse tissues\*

Renu Wadhwa<sup>a</sup>, Sunil C. Kaul<sup>b</sup>, Yasuhiko Komatsu<sup>b</sup>, Yoji Ikawa<sup>a</sup>, Akinori Sarai<sup>c</sup>  
and Yoshikazu Sugimoto<sup>a</sup>

<sup>a</sup>Laboratory of Molecular Oncology and <sup>c</sup>Gene Bank, Tsukuba Life Science Center, The Institute of Physical and Chemical Research, Koyadai, Tsukuba, Ibaraki 305, Japan and <sup>b</sup>Fermentation Research Institute, Agency of Industrial Science and Technology, Tsukuba, Ibaraki 305, Japan

Received 26 October 1992; revised version received 17 November 1992

We have isolated a yeast *SEC23*-related clone (*Msec23*) from mouse fibroblast cDNA library. It has an open reading frame of 1721 bp (64% homologous to *SEC23*–2.3 kb) which can potentially encode a 64.7 kDa protein (61% homologous to 85.4 kDa product of *SEC23*). The deduced *Msec23* protein (*Msec23p*) sequence contains three successive Ig-like domains at the N-terminus followed by amphipathic  $\alpha$ -helical regions, suggesting the potential of *Msec23p* to interact with other protein components. Further, *Msec23* is differentially expressed in mouse tissues with its high level in brain and fibroblasts.

Yeast *SEC23*-related gene; cDNA sequence; Differential tissue expression

## 1. INTRODUCTION

Molecular understanding of various biological phenomena demands the isolation and characterization of relevant cDNA clones. These are targeted by screening of cDNA libraries with specific antibodies or DNA probes, subtractive or differential hybridization. Using one of these strategies to isolate genes responsible for mortality and immortality in mouse fibroblasts we got hold of a novel clone which showed 64% homology to a member of yeast secretory pathway, *SEC23*. The latter has been demonstrated to play a role in intracellular transport from endoplasmic reticulum (ER) to Golgi both in vivo and in vitro [1]. Intracellular transport of proteins from cytoplasm to various organelles is the essential requirement for cellular structural and functional integrity. The molecules mediating such mechanism hence are the subject of great interest. In yeast more than 27 genes have been defined to have roles in secretory pathway by mutational complementation approach [2–5]. Mammalian homologue of *Sec23p* has been identified immunologically as a 84–85 kDa protein [6]. It is restricted to ER transitional cytoplasm as observed by immuno-electron microscopy in liver and

pancreas of Sprague–Dawley rats. However, the information on its DNA structure is lacking. We accidentally isolated a yeast *SEC23*-related clone from mouse fibroblast cDNA library. We report here the nucleotide sequence of the novel cDNA clone (*Msec23*) along with its tissue specific expression in mouse.

## 2. MATERIALS AND METHODS

### 2.1. Construction and screening of cDNA library

Poly(A)<sup>+</sup> RNA prepared from CD1 ICR mouse embryonic fibroblasts (MEF) by using Fast track mRNA isolation kit (Invitrogen) was used to construct cDNA library in the  $\lambda$ ZAPII vector (Stratagene) at *EcoRI* site. The clone dealt with in the present report was isolated while screening of the MEF cDNA library with p66 cDNA probe isolated from the library with anti-p66 antibody (see Wadhwa et al. [7] for p66 protein).

### 2.2. DNA sequencing

cDNA clone (1.9 kb) in pBluescript SK (Stratagene) was sequenced by the dideoxy chain termination method using T7 and T3 primers (Sequenase 2.0 kit, US Biochemicals). The clones were deleted sequentially from 3' end by exonuclease III using deletion kit (Takara). The complete nucleotide sequence was derived by ligating the sequence of deletion mutants and was confirmed by reverse direction cloning and sequencing. The DNA sequence was analyzed on a  $\mu$ Vax computer.

### 2.3. Northern analysis

Total cell and tissue RNA were extracted and separated on denaturing agarose gel [8]. The ultraviolet-cross-linked membranes were hybridized with <sup>32</sup>P-labeled *Msec23* cDNA probe (1.9 kb).

## 3. RESULTS AND DISCUSSION

It would rather be fair to state that the present report is the fortuitous off-shoot of our main project which

Correspondence address: Y. Sugimoto, Laboratory of Molecular Oncology, Tsukuba Life Science Center, The Institute of Physical and Chemical Research, Koyadai, Tsukuba, Ibaraki 305, Japan. Fax: (81) (298) 36 9020.

\*The nucleotide sequence reported in this paper has been submitted to the DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number D12713.

```

10          30          50          70          90          110
CGGCAGAAATAAGAAATCAAGTGCTACCATGACAACTATTGGAATTATCCAAACAAATGAAGAACGAGATGGGTCGGGTTCAGTTGGAATGTTGGCCATCGAGTGCCTGAAGCT
      M T T Y L E F I Q Q N E E R D G V R F S W N V W P S S R L E A
130          150          170          190          210          230
ACAAGAATGGTGTCCAGTGGCAGCCTGTTTACACCTCTGAAGGAGAGACCCGACCTACCAACCCATCCAGTAGAGCCCTGTCTGTGCAGTAGGACCACTTCCCGTGCAGTTTGAAT
T R M V V P V A A L F T P L K E R P D L P P I Q Y E P V L C S R T T C R A V L N
250          270          290          310          330          350
CCTTTATGCCAAGTGATTATCGAGCAACTCGGGCTTGCAACTTTGTACCAAGATCAGTTCCACCTACATATTCTGGGATATCTGAATTGAACAGCCCTCGAGAGTTTACCTAGTTTCA
P L C Q V D Y R A T R A C N F V P R S V P P T Y S G I S E L N Q P R E F Y L V S
370          390          410          430          450          470
GATGAATGTAGTTCTGCTGGTCCCGAGATGCCCTTTCATATTCTCTATGTGGTTGATATTCATAGAAAGATGAGGATTTGAGGCTCTGAAGGAGTCCATGCAGACGACGTTTCAGC
D E Y V V L R G P Q M P F I F L Y V V D T C I E D E D L Q A L K E S M Q T T F S
490          510          530          550          570          590
CTTTTCCACCTACAGCTTTGGTTGGACTTATTACCTTTGGGAGAACTGTGCAAGTTTCATGAGCTTGGATGCGAGCATTCAAGAGCTATGTCTTCAGAGGAACAAAGGATCTGCTGCC
L F P P T A L V G L I T F G R I V Q V H E L G C E H S K S Y V F R G T K D L S A
610          630          650          670          690          710
AAACAATGCGAGAAATGCTGGGACTCTTAAGTACCTGTACTCAAGCCACATCGCGGCTCTCAGGTACAGCAACCGCCACTTCAATAGATTCTACAGCAGTACAGAAATAGAC
K Q L Q E M L G L S K V P V T Q A T S R S S G T A T A T F N R F L Q P V Q K I D
730          750          770          790          810          830
ATGAATCTCAGAGATCTCTGGGAGAACTTCAGCGAGACCCCTGGGCTCTACCAAGCAAGAGAGACCGCTGCCGCCCTCAGGAGTGGCCCTTCCATAGCTGTGGAGTCTGCTGAGTGT
M N L T D L L G E L Q R D P W P V P Q G K R P L R P S G V A L S I A V G L L E C
850          870          890          910          930          950
ACTCCCAACACTGGTCTCGGATCATGATGTTTACCTGCTGCTGCTACCAAGGCTCTGCGATGCTGGGAGATGAGCTAAAGACACTATGAGATCTGGCAGCAGATTGAGAGGAC
T P Q H W C S D H D V H R C L L P G P G M V V G D E L K T P M R S W H D I E K D
970          990          1010          1030          1050          1070
AATCCAAATATGTTAAAGGGAATTAAGCATTTTGAAGCGTTGGCTAATCGAGTCTTACAACGGGCGATGTCATTGATATCTAGCCCTGTGCTGGACAGACAGGCTCTGCGAGATG
N P N M L K R E L S I L K R W L I E L L Q R G M S L I S T P V L D Q T G L L E M
1090          1110          1130          1150          1170          1190
AAGTCTGCTCTAACCTTACCTGGAGGATACATGGTAATGGGTGACTCTTCAATACCTCTTATTCAAGCAAACTTTTCAAGAGCTTTCACCAAGATATACATGCCAGTTTAAATG
K C C P N L T G G Y M V M G D S F N T S L F K Q T F Q R V F T K D I H G Q F K M
1210          1230          1250          1270          1290          1310
GGCTTTGGTGGCATTAGAAATAAGACTTCAAGGGAATAAGATTTCAGGAGCTATTGGACCTGTGCTCTTAATCAAAAGGACCTTGGCTGTCTGAAATGAGATTGGAACAG
G F G G T L E I K T S R E I K I S G A I G P C V L L I Q K D L A C L K M R L E Q
1330          1350          1370          1390          1410          1430
GAGGCACTGTGCTGGAATACTGTGGCTCACCCACCACAACCTTAGCCATATATTTGAAGTTGTTAATCAGCATAATGCTCAATTCCTCAAGGAGGACGAGGTGCGTCCCAATT
E A L V S G K S V A S P T T T L A I Y F E V V N Q H N A P I P Q G G R G A V Q F
1450          1470          1490          1510          1530          1550
GTGACCCAGTATCAGCACTCAAGTGGCCAGAGACGCTCCGAGTGACCAATGCTAGGAACCTGGGAGATGCTCAAACTCAAAATCAAAATATTGCTGATCTTTTGAACGAGGAGCA
V T Q Y Q H S S G Q R R I R V T T I A R N W A D A Q T Q I Q N I A R S F D Q E A
1570          1590          1610          1630          1650          1670
GCTCGATTCTCATGGCAGGCTGGCAATATACAGAGCAGAAACAGAGGAAGGGCCAGACGTGCTAAGATGTTGGACAGACAACCTATTGACTGTCTCAAAATTTGGAGAGTATCAC
A F I L M A R L A I Y R A E T E E G P D V L R W L D R Q L I R L C Q K F G E Y H
1690          1710          1730          1750          1770          1790
AAGATGATGCAAAATCTTCGGTTCAGAAACATTTCTCTTATCTCAGTTTATGTTTCATTGAAGAAGATCTCTTCTTCAAGATTTTAAACAATAGTCTGATGAGAGTTTCAT
K D D P N S S G F Q K H F L F I L S L C F I *
1810          1830          1850          1870
ACTATCGTCACCATTTCTGCTGAGGATCTGACCCAGTCTCTGATCATGATTACGCCCATCTCTGTACGCTTACTCTTCAAGTGGCCGG

```

Fig. 1. *Msec23* sequence. Complete nucleotide sequence of *Msec23* isolated from MEF cDNA library along with the deduced amino acid sequence.

aims to isolate genes involved in mortal and immortal phenotypes in mouse fibroblasts. We have identified and characterized a cytosolic 66-kDa protein which is associated with the mortal phenotype of mouse fibroblasts [7]. During the screening of MEF cDNA library with 1.6 kb *EcoRI* fragment of *p66* cDNA which was aimed to obtain the 5' upstream sequence of *p66* cDNA clone, clone 18 was isolated. However, upon partial sequencing it showed minor homology to *p66* cDNA but was found to be 64% homologous to yeast *SEC23* cDNA sequence which functions in intracellular transport from ER to Golgi [1]. Thus we have called it *Msec23*. Complete nucleotide sequence along with the predicted amino acid sequence is given in Fig. 1. It has an open reading frame of 1721 bp (28–1749). That ATG at 28 bp position acts as the initiation codon is supported by the presence of 5' upstream in-frame stop codon TAA at 10 bp position as well as by the Kozak rule [9]. The open reading frame has the potential to encode a leucine-rich protein (*pI* 7.9) consisting of 572

amino acids corresponding to the molecular weight of 64.7 kDa, in contrast to yeast *SEC23* protein (*YSEC23p*) with 768 amino acids with the molecular weight of 84 kDa [1] and immunologically detected 84–85 kDa mammalian *Sec23p* [6].

Comparison of the amino acid sequence of *Msec23p* with that of *YSEC23p* revealed a strong homology in the entire region, with 40% identical amino acids and 61% homology (Fig. 2). These sequences have no strong homology with other sequences in the database. However, close inspection of the amino acid sequences indicates that there are three homologous domains in the N-terminal halves of the proteins. These domains consist of 90–110 amino acids (D1, 15–120; D2, 120–215; D3, 215–315 in the *Msec23p*, and 10–115, 115–220, 220–320 in *YSEC23p*). In each domain, most of the hydrophobic amino acids and turn-forming amino acids such as Pro, Gly and polar amino acids are conserved between the two sequences. This conservation of structurally important amino acids suggests that the corre-

```

Msec  8  IQQNEERDGVRFESNNVWPSRLLEATRMVVPVAALFTPLKERPDLPPIQYE 57
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
sec23  3  FEINEDINGVRFTNNVFPSTRSDANSNVVPVGCCLYTPLKKEYDELNVAPYN 52

      58  FVLCSSRTTCRAVLNPLCOVDYRATRACNFVPRS...VPPTYSGISELNQF 104
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      53  PVVCSGPHCKSILNPFYCVIDPRNSSWSCPICNSRNHLPPQYTNLSQENMP 102

      105  REFYLVSDFYVVLRGQPMPFIPLYVVDTCIEDLQAIKESMOTTFSLFP 154
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      103  LELQSTTIEYITNKPVTPWPIFFVVDLTSETENLDSLKESIITSLSLP 152

      155  PTALVGLITFGRIVQVHELGCCH.SKSYVFRGTDLSAKOLOEMLGLSKV 203
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      153  PNALIGLITYGNVQLHDLSSSETIDRCNVFRGDREYQLEALTEMLTGQKP 202

      204  PVTQATSRSSGTATA.....TFNRFLQPVQKIDMNLTDLLGELQDPWPV 248
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      203  TGPGGAASHLPNAMNKVTPFSLNRFPLPLEQVEFKLNQLENLSPDQWSV 252

      249  PQGRPLRPSGVALSIAGVLLLECTPOHWCSDHVDHRCLLP.....GPGM 292
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      253  PAGHRPLRATGSALNIASLLQGC...YKNIPARIILFASGPGTVAPGL 298

      293  VVGDELKTPMRSWHDIEKDNPNMLKRELSILKRWLIEFLQRCMSL.ISTP 341
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      299  IVNSELKDFLRSHHDIDSDHAQHYKACKFYNQIAQRVAANGHTVDIFAG 348

      342  VLDQTGLLEMKCCRNLTGGYVMVGDSFNTSLFKQTFQVFTKDIHGQFKM 391
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      349  CYDQIGMSEMKQLTDSGGVLLLTDAFSTAIFKQSYLRLFAKDEEGYLKM 398

      392  GFGGTLEIKTSREIKISGAIGPCVLLIQKDLACLKMRLEQAEALVSGKSA 441
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      399  AFNGNMAVKTSKDLKVQGLIGHASAVKKTDAANNISESEIGIGATSTWKMA 448

      442  ..SPTTTLAIYFEVNN...QHNAPIQGGRGA.....VQFVTQYQHSSG 480
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      449  SLSPYHSAIFFEIANTAANSNPMMSAPGSADRPHLAYTQFITTYQHSSG 498

      481  QRRIRVTTIARNWADAQTQIQNIARSFDQEAAPILMARLAIYRAETEEGP 530
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      499  TNRIRVTTVANQLLPGFT..PAIAASFDQEAALVMARIAVHKAETDDGA 546

      531  DVLRLWDRQLRLCQKFGYHKDDPNSSGFQKHFLFILSLCFI 573
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      547  DVIRWLDRTLKLCQKYADYNKDDPQSFR LAPNFSLYPQFTYY 589

```

Fig. 2. Alignments of amino acid sequences of mouse *Msec23p* (Msec) with *YSEC23p* (sec23). Matches are indicated by vertical bars (identical amino acids) and dots (related amino acids). Underlining indicates three successive Ig-like domains (D1, D2 and D3) of *Msec23p*. Dotted line indicates helical region of *Msec23*.

sponding domains of the two proteins have similar structure. According to the locations of hydrophobic and turn-forming amino acids, these domains are likely to form Ig-like structure with seven to nine  $\beta$ -strands (Fig. 3). The Ig-like structures have been found in pro-

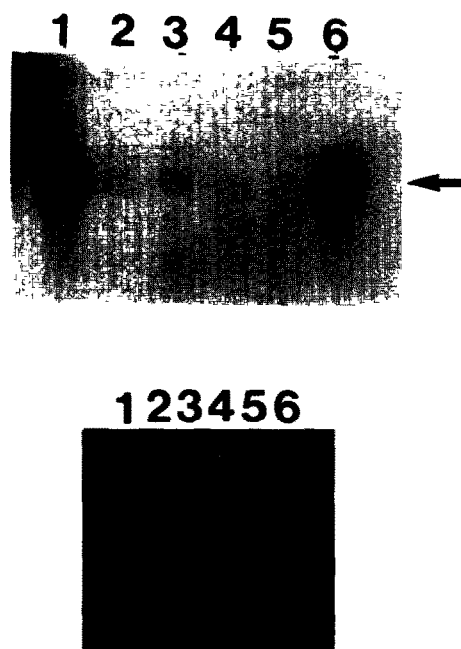


Fig. 4. Northern blot analysis of *Msec23* transcript in various mouse tissues. Lanes 1–6 are loaded with 20  $\mu$ g of total RNA from brain, lung, heart, liver, kidney and fibroblasts, respectively. Lower panel shows the ethidium bromide staining of corresponding RNA lanes.

teins which bind to peptides and other proteins, i.e. in T cell receptor, MHC, CD4, CD8, interleukin 6 receptor and colony-stimulating factor receptor. Thus the Ig-like structure of *Msec23p* as well as *YSEC23p* accords well their role in the association with other molecules.

In addition, the C-terminal halves of *Msec23p* and *YSEC23p* have rather amphipathic  $\alpha$ -helical characteristics in several places. In particular the mouse sequence has repeated Leu's 6–8 amino acid apart in the region between 315 and 375, next to the Ig-like domains (Fig. 2). This structure could potentially mediate its interaction with other proteins similar to the 'leucine-zipper' motif.

*Msec23* is the first *SEC23*-related cDNA isolated from mammalian system though it remains to be seen whether it represents the yeast *SEC23*-corresponding gene. *Msec23p* (64.7 kDa) lacks the C-terminal 200 amino acids contained in *YSEC23p* (84 kDa) [1]. In addition, anti-*YSEC23p* antibody recognized 84–85 kDa protein in mouse liver and pancreas [6]. However, as can be seen in Fig. 4, brain (lane 1) and fibroblast (lane 6) expressed the high level of 3.8 kb *Msec23* transcript as compared to the others like liver where 84–85 kDa mammalian *Sec23p* was found to be abundant [6]. Thus *Msec23* may not be a gene encoding 84–85 kDa mammalian *Sec23p*. It is rather intriguing to delineate the biological significance of its differential expression and also if it can complement yeast *sec23*.

	0	A	B	C	49
Msec-D1	RDGVRF	SWNV W..PSSR.LE	ATRMVVPVAA LF....TP..	...LKER...	
Msec-D2	FIFLYV	VDTC I..EDED...	LQALKESMOT TF..SLFPPT	.ALVGLIT..	
Msec-D3	TATATF	NRFL ....QP....	VQKIDMNLTD LL.GELQRDP	...WFPVQGK	
CD4C	LVFGLT	ANS D..THL..LQ	GQSLTLTLES ...PPGSSP.	..SVQCRS..	
CD4N	KKVVLG	K..KGD....	..TVELTCTA ..SQKSIQ.	...FHWKN..	
Ig-C	PTVSIF	PPS ..SEQLTS..	.GGASVVCFL NN.FYPKD..	.INVKWKID.	
Ig-V	<u>EIVLTQS.</u>	<u>.PAITAAS.L</u>	<u>GQKVTITCSA</u>	<u>...SSSVSS.</u>	<u>...LHWYQQK</u>
		# #	# # #	# #	
	50	C'	C''	D	99
Msec-D1	..PDLPP..I	QYE.....	P....VLCS.	RTT..CRAVL	N....PL..C
Msec-D2	..FGR.....	IVQVHEL..G	CE.....	.....	.....H
Msec-D3	.RPLRPSG.V	ALSIA.....	VG...LLEC.	.TP....QHW	C....SD...
CD4C	..PRGK....	NIQG.....G	K.....	.....	.....
CD4N	..SNQ....I	KILGNQ...G	S....FLTK.	GPSK.LNDRA	DSR.RSLWDQ
Ig-C	GSERQNG..V	LNSWTDQD..	SKDST.....	.....	.....
Ig-V	<u>.SGTSPKP.W</u>	<u>IYEI.....</u>	<u>SK...LASGV</u>	<u>.PAR.FSGSG</u>	<u>S...GTSY..</u>
		# # #	##	# #	
	100	E	F	G	149
Msec-D1	QVDYRATRAC	NFV.PRSVPP	....TYSGIS	ELNQ..PRE.	..FYLVSDEY
Msec-D2	SKSYVFR...	....GTKD..	....LSAQ	LQEM.LGLSK	VP.VTQATSR
Msec-D3	...HDVHRCL	L....PGPG.	....MVVD	ELKT..PMRS	...WHDIEKD
CD4C	..TSLVSQ	L....ELQD..	..SGTWTCTV	LQ....NQK.	..KVEFKIDI
CD4N	.GNFPLIK.	...NLKIEDS	...DTYICEV	....EDQK.	....EEVQL
Ig-C	...YSMSSTL	TLT..KDEYE	R.HNSYTCEA	TH..KTSTSP	..IVKSFNRN
Ig-V	<u>..SLTINTME</u>	<u>A....ED..</u>	<u>..AAIYYCQQ</u>	<u>WT....YP..</u>	<u>..LITFGAGT</u>
		# #	# #	#	
	150				
Msec-D1	VVLRGP				
Msec-D2	SSGT				
Msec-D3	NPNM				
CD4C	VVL				
CD4N	LVF				
Ig-C	EC				
Ig-V	<u>KLELKR</u>				

Fig. 3. Multiple sequence alignment of the Ig-like domains of Msec23p (Msec-D1 to Msec-D3), CD4 C and N domains (CD4C and CD4N), and immunoglobulin constant and variable domains (Ig-C and Ig-V). The  $\beta$ -strand portions are indicated by underlinings and designated by letters. # denotes the conserved hydrophobic amino acids.

## REFERENCES

- [1] Hicke, L. and Schekman, R. (1989) EMBO J. 8, 1677-1684.
- [2] Deshaies, R. and Schekman, R. (1987) J. Cell Biol. 105, 633-645.
- [3] Ferro-Novick, S., Hansen, W., Schauer, I. and Schekman, R. (1984) J. Cell Biol. 98, 44-53.
- [4] Novick, P. and Schekman, R. (1979) Proc. Natl. Acad. Sci. USA 76, 1858-1862.
- [5] Novick, P., Field, C. and Schekman, R. (1980) Cell 21, 205-215.
- [6] Orci, L., Ravazzola, M., Meda, P., Holcomb, C., Moore, H., Hicke, L. and Schekman, R. (1991) Proc. Natl. Acad. Sci. USA 88, 8611-8615.
- [7] Wadhwa, R., Kaul, S.C., Ikawa, Y. and Sugimoto, Y. (1991) Mut. Res. 256, 243-254.
- [8] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) in: Molecular Cloning: a Laboratory Manual, Cold Spring Harbor, Cold Spring Harbor Laboratory, NY.
- [9] Kozak, M. (1984) Nucleic Acids Res. 12, 857-872.