Cloning and Expression of Human Homolog HSMT3 to Yeast SMT3 Suppressor of MIF2 Mutations in a Centromere Protein Gene

Hideyuki Mannen,* Hui-Min Tseng,† Chung-lung Cho,† and Steven S.-L. Li*,†,1

*Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709; and †Institute of Life Sciences, National Sun Yat-Sen University, Kaohsiung, Taiwan 80424, R.O.C.

Received April 3, 1996

A human HSMT3 cDNA encoding a homolog of the yeast SMT3, a suppressor of MIF2 mutations in a centromere protein gene, was identified and sequenced. The sequence of 95 amino acids deduced from the human HSMT3 cDNA exhibited 51.1% identity and 69.6% similarity to the yeast Smt3p sequence. The HSMT3 transcripts of 1.35Kb were found to be abundantly expressed in various human tissues. © 1996 Academic Press, Inc.

The proper segregation of eukaryotic chromosomes during mitosis and meiosis is mediated by centromere/kinetochore. Centromeres facilitate sister chromatid association. Kinetochores contain spindle attachment sites and molecular motors that move chromosomes on the microtubules during chromosome segregation (1–2). In human, centromere proteins (i.e. CENP-C) were first isolated using autoantibodies from patients of scleroderma, and their sequences have been determined (3). In yeast, *Saccharomyces cerevisiae*, the MIF2 gene has been shown to encode an essential centromere protein which binds A+T-rich CDEII region of centromere DNA (4). The yeast Mif2 protein shares at least two regions of similarity with mammalian centromere protein CENP-C, suggesting an evolutionary conservation of centromere proteins from yeast to mammals (5). Studies using temperature-sensitive mutants showed that the loss of yeast Mif2p function results in chromosome missegregation, mitotic delay, and aberrant microtubule morphologies (6). The yeast SMT3 suppressor was isolated from the suppression of the MIF2 mutations. The SMT3 gene is located on chromosome IV and its nucleotide sequence was recently deposited in the GenBank (accession no. U27233 and U33057).

A porcine cDNA of 230 nucleotides, isolated serendipitously (Cho, unpublished), has a sequence highly similar to 19 human EST sequences. Partial amino acid sequences deduced from this porcine cDNA and the human EST sequences were unexpectedly found to exhibit strong homology with the yeast Smt3 protein sequence. As the first step to identify and elucidate the function, here we report the cloning and expression of a human HSMT3 cDNA using the porcine cDNA as a probe.

MATERIALS AND METHODS

Cloning and sequencing of human HSMT3 cDNA. Human cDNA library in λ ZAPII (Stratagene) was screened using the porcine cDNA probe labeled with digoxigenin system. The positive clones were identified with chemiluminescent detection according to the procedure recommended by the manufacturer (Boehringer Manneheim). The purified cDNA insert was labeled with the Dye Terminator kit (Perkin Elmer), and its nucleotide sequence was determined using an automatic DNA sequencer (Applied Biosystems model 373A). Both strands of the inserted cDNA were completely sequenced.

Northern blot analysis. Human northern blot containing $poly(A^+)$ -RNAs from adult tissues, including heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas, was obtained from Clontech and hybridized for 16–18h at 42C in formamide, 10× Denhardt's solution, 5× buffer A (0.75M sodium chloride, 50mM sodium phosphate, 5mM EDTA, pH 7.4), 1% SDS, salmon sperm DNA (100ug/ml) with ³²P-labeled cDNA probe. The blot was rinsed twice in 2× NaCl/Cit., 0.05% SDS at room temperature for 10min, and washed twice in 0.1× NaCl/Cit., 0.1% SDS at 50C for 20min. The x-ray film was exposed overnight at –70C.

¹ To whom all correspondence should be addressed in Taiwan. Fax: 886-7-532-1213.

															-					11
ATGGCCGACGAAAAGCCCAAGGAAGGAGTCAAGACTGAGAACAACGATCATATTAATTTG													74							
M	A	D	Е	K	₽	ĸ	Е	G	v	ĸ	т	Е	N	N	D	H	I	N	L	20
77	000	200		201	00.1	maai	TIMO	m¢m	COT	003	രനന		<u>съ</u> т	ת תרחת	0.20	003	m	1200	13 O MM	124
K	V	A SGC	G	0 0	םםם מ	G	s s	v	v	OCA O	F F	TAA K	T	K	R	RJD. H	TAC T	ACC P	.ACTT T.	134
	•		-	-	-	-	-	•	•	*	-		_				-	-	-	
AG	TAA	ACT	AATO	GAA	AGC	CTA'	TTG	TGA	ACG	ACA	GGG.	ATT	GTC	AAT	'GAG	GCA	GA1	CAC	ATTC	194
S	ĸ	L	М	ĸ	A	Y	С	Е	R	Q	G	L	S	М	R	Q	I	R	F	60
CC	<u>አ</u> ጥጥי	TCA	raad		ACC	ል አጥሰ	יממיז	ጥር እ	2 2 C		CAC	200	നവന	202	CTT	יררא	אאת	0007	00.00	254
R	F	D	G	0	P	I	N	E	T	D	T T	P	A	0	L	E	M	E	D	204
	-	_	-	~	-	_		_	-	-	-	-		-	_	-		-	-	•••
GA	AGA'	TAC.	AAT	FGA	TGT	GTT	CCA	ACA	GCA	GAC	GGG.	AGG	TGT	CTA	CTG	AAA	AGG	GAA	CCTG	314
E	D	т	I	D	v	F	Q	Q	Q	т	G	G	v	Y	*					95
E	ם דרידי	ד דיד <i>י</i> ם מ	I CTCC	D מר	ע סבב	F ጥርጥ	Q Стт	Q Ռաթար	Q תאש	Т 202	G	G	ע דידים	Y CAT	* יידיריתי	יריא	ጥጥ አ		አእርጥ	95
E CT GC	ם דכדי מאדי	T TTA TTG	I CTC(GTT(D CAG	V AAC	F TCT CAT	Q GTT CCT	Q CTT GAC	Q TAA TAC	T AGA	G CCA	G AGA	V TTA GTT	Y CAT TTC	* TCT	CAA	TTA CTT	GAA	AACT	95 374 434
E CT GC CC	D TCT' AAT' CCT'	T TTA TTG TCC	I CTC(GTT(CCA	D CAG CCA	V AAC' CCA' CTT'	F TCT(CAT(TAT	Q GTT CCT CCT	Q CTT GAC ACA	Q TAA TAC TAA	T AGA TAC AGT	G CCA CGT. AAC	G AGA ATA TGG	V TTA GTT TAT	Y CAT TTC ATG	* TCT TCT	'CAA 'ATT 'ACA	TTA CTT AGC	GAA TCA	AACT TTTC	95 374 434 494
E GC CC AT	D TCT' AAT' CCT' TTT'	T TTA TTG TCC TTT	I CTC(GTT(CCA' FTT'	D CAG CCA TTC TTC	V AAC CCA CTT TAA	F TCT CAT TAT CTA	Q GTT CCT TGT. AAC.	Q CTT GAC ACA AGC	Q TAA TAC TAA CAA	T AGA TAC AGT TGG	G CCA CGT AAC	G AGA ATA TGG GTT	V TTA GTT TAT TTG	Y CAT TTC ATG ATT	* TCT TGC GAC	'CAA 'ATT 'ACA 'ATC	TTA CTT AGC AAG	GAA TCA ATA	AACT ATTTC ATTGC AGAC	95 374 434 494 554
E GC CC AT GG	D TCT' AAT' CCT' TTT' GAT(T ITG(ITG(ICC(ITT' GGG(I GTTC GTTC CCA TTT GAA	D CAG CCA PTC PTT AAA	V AAC CCA CTT TAA TAC	F TCT CAT TAT CTA TGA	Q GTT CCT TGT AAC. TTC	Q GAC ACA AGC TGT	Q TAA TAC TAA CAA GAA	T AGA TAC AGT TGG AAT	G CCA CGT AAC TAT	G AGA ATA TGG GTT CCC	V TTA GTT TAT TTG TTG	Y CAT TTC ATG ATT CTC	TCT TCT TGC GAC CAT	'CAA 'ATT 'ACA 'ATC 'TAG	TTA CTT AGC AAG TGG	GAA TCA ATA TGO CA1	AACT ATTTC ATTGC AGAC GCTC	95 374 434 494 554 614
E GC CC AT GG AT	D TCT AAT CCT TTT GAT TCA	T TTA TTG TCC TTT GGG GCT	I GTTC GTTC CCA TTT GAA CTT	D CAG CCA ITC ITC AAA ATC	V AAC CCA CTT TAA TAC TTT	F TCT CAT TAT CTA TGA ATA	Q GTT CCT TGT. AAC. TTC TTC	Q GAC ACA AGC TGT CAG	Q TAA TAC TAA CAA GAA TAA	T AGA TAC AGT TGG AAT GTT	G CCA CGT AAC TAT ACC ATT	G AGA ATA TGG GTT CCC TTG	V TTA GTT TAT TTG TTT CTC	Y CAT TTC ATG ATT CTC TCA	* TCT TGC GAC CAT CTG	'CAA 'ATT 'ACA 'ATC 'TAG	TTA CTI AGC AAG TGG TAA	GAA TCA ATA TGO CAT	AACT ATTTC ATTGC GAGAC GCTC AAAA	95 374 434 494 554 614 674
E CT GC CC AT GG AT AA	D TCT AAT CCT TTT GAT GAT TCA AAA	T TTA TTG TCC TTT GGG GCT ACA	I GTTC GTTC CCA TTT GAA CTT ACA	D CAG CCA FTC FTT AAA ATC ACA	V AAC CCA CTT TAA TAC TTT. TAA	F TCT(CAT(TAT' CTA TGA' ATA' AAA'	Q GTT CCT TGT. AAC. TTC TTC TCC	Q GAC ACA AGC TGT CAG TTG	Q TAA TAC TAA CAA GAA TAA CAT	T AGA TAC AGT TGG AAT GTT	G CCA CGT AAC TAT ACC ATT TTG	G AGA ATA IGG GTT CCC ITG ITC	V TTA GTT TAT TTG TTT CTC AAT	Y CAT TTC ATG ATT CTC TCA TGG	* TCT TGC GAC CAT CAT CTG AGA	CAA ATT ACA ATC TAG TTAG	TTA CTI AGC AAG TGG TAA TTA	GAA TCA TGG CAT CAA ATG	AACT ATTTC ATTGC AGAC GCTC AAAA TTTT	95 374 434 494 554 614 674 734
E CT GC CC AT GG AT AA TC	D TCT' AAT' CCT' TTT' GATG TCAG AAAJ ATT'	T TTA TTG TCC TTT GGG GCT ACA TAT	I GTTC GTTC CCA TTT GAA GAA CTT ACA CAT	D CAG CCA ITC ITT AAA ATC ACA IGT	V AAC CCA TAA TAC TAC TAA AAA	F TCT CAT TAT CTA TGA ATA AAA ACC	Q GTT CCT TGT. AAC. TTC TTC TCC AAG	Q GAC ACA AGC TGT CAG TTG GAC	Q TAA TAC TAA CAA GAA TAA CAT AAT	T AGA TAC AGT TGG AAT ACC TTT	G CCA CGT AAC TAT ACC ATT TTG ATA	G AGA TGG GTT CCC TTG TTC ACT	V TTA GTT TAT TTG TTT CTC AAT TTT	Y CAT TTC ATG ATT CTC TCA TGG TTG	TCT TCT GAC CAT CTG AGA TAC	CAA ATT ACA TAG TTT ATT GTA	TTA CTT AGC AAG TGG TAA TTA GCT	GAA TCA TGC CAT CAA ATC GTT	AACT ATTTC ATTGC AGAC GCTC AAAA TTTT ACAT	95 374 434 494 554 614 674 734 794
E CT GC CC AT GG AT AA TC GT	D TCT AAT CCT TTT GAT GAT TCA AAA ATT AGG	T TTA TTG TCC TTT GGG GCT ACA TATA GCA	I GTC(GTT(CCA' GAA) GAA) CTTJ ACA) CAT' ATC'	D CAG CCA TTC TTT AAA ATC ACA TGT TGT	V AAC CCA CTT TAA TAC TTT. TAA AAA CTT	F TCT(CAT(TAT' CTA) TGA' ATA' AAA' ACC) TAA(ACC)	Q GTT CCT TGT, AAC, TTC TTC TCC AAG GTA	Q CTT GAC ACA AGC TGT CAG TTG GAC GGG	Q TAA TAC CAA GAA TAA CAT AATA	T AGA TAC AGT TGG AAT ACC TTT AAT	G CCA CGT AAC TAT ACC ATT TTG ATA TAC	G AGA TGG GTT CCC TTG TTC ACT ICT	TTA GTT TAT TTG TTT CTC AAT TTT AAA	Y CAT TTC ATG ATT CTC TCA TGG TGG ACA	TCT TCT TGC GAC CAT CAT CAT CAT AGA TAC	CAA ATT ACA TAG TTAG TTT ATT GTA AAG	TTA CTI AGC AAG TGG TAA TTA GCI AAI	GAA TCA TGG CAT CAT CAA ATC GTT CCT	AACT ATTC AGAC GCTC AAAA ATTTT ACAT	95 374 434 494 554 614 674 734 794 854
E CT GC CC AT GG AT AA TC GT AG	D TCT AAT CCT TTT GAT GAT TCA AAA ATT AGG TTT AAA	TTA TTA TTG TCC TTT GGGG GCT ACA TAT GCA TCC AAA	I GTCC GTTC CCA TTTT GAA CTTT ACA CAT CAT CAT	D CAG CCA ITC ITC AAA ATC AACA IGT CAA	V AAC CCA CTT TAA TAC TAA AAA CTT GTC	F TCT(CAT(TAT' CTA) TGA' ATA' AAA' AAA' AAC) TAA(AAG(Q GTT CCT TGT AAC TTC TTC TCC AAG GTA CGT	Q GAC ACA AGC TGT CAG TTG GAC GGG CTT	Q TAA TAC TAA GAA TAA CAT AAT GTT	T AGA TAC AGT TGG AAT ACC TTT AAT GTT	G CCA CGT AAC TAT ACC ATT TTG ATA TAC TA <u>A</u>	G AGA ATA IGG GTT CCC ITG ITC ACT ACT ATA	V TTA GTT TAT TTG TTT CTC AAT TTT AAA <u>AA</u> C	Y CAT TTC ATG ATT CTC TCA TGG TTG ACA TTC	TCT TCT TGC GAC CAT CAT CAT AGA TAC AAA TTG	CAA ATT ACA TAG TAG TTT ATT GTA AAG	TTA CTI AGC AAG TGG TAA TTA GCI AAI AAA	GAA TCA TGG CAT CAA ATG GTI CCI ATG	AACT ATTC AGAC GCTC AAAA ATTTT ACAT AGAT	95 374 434 494 554 614 674 734 794 854 914

FIG. 1. The nucleotide and deduced amino acid sequences of human HSMT3 cDNA. The stop codon TGA is indicated by an asterisk and the polyadenylation signal of AATAAA is underlined.

RESULTS

Human HSMT3 cDNA and Amino Acid Sequences

A human HSMT3 cDNA clone was isolated and its sequence of 922 nucleotides was determined (GenBank accession no. L76416). This cDNA contains a protein-coding sequence of 285 nucleotides, 5' (14 nucleotides) and 3' (611 nucleotides) noncoding regions, and poly(A) tail of 12 nucleotides (Fig. 1). The deduced sequence of 95 amino acids from the human HSMT3 cDNA was found to exhibit 51.1% identity and 69.6% similarity with yeast Smt3 protein (Genbank U27233 and U33057) (Fig. 2). This result indicates the strong evolutionary conservation between human HSMT3 and yeast Smt3 proteins.

Northern Blot Analysis

The Northern blot analysis of human poly (A)-RNAs indicated that the HSMT3 transcripts of 1.35Kb were abundantly expressed in the various tissues, including heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (Fig. 3). The 19 human EST sequences exhibiting strong homology with the HSMT3 cDNA sequence were obtained from brain, liver, spleen, white blood cells, endothelial cells, and testis. These results are consistent with the ubiquitous expression of the HSMT3 gene in dividing tissues.

Human	MADE-KPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLM 	44
Yeast	MSDSEVNQEAKPEVKPEVKPETHINLKV-SDGSSEIFFKIKKTTPLRRLM	49
Human	KAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG-VY	95
Yeast	EAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGATY	101

FIG. 2. Comparison of amino acid sequences between human HSMT3 and yeast Smt3 proteins.



FIG. 3. Northern blot analysis of human HSMT3 transcripts from various tissues.

DISCUSSION

The size of the human HSMT3 transcript was found to be 1.35Kb using Northern blot analysis, while this HSMT3 cDNA clone contains only 922 nucleotides. The remaining sequence may be present in the 5' non-coding region. It may be noted that a porcine PSMT3 cDNA clone was found to contain a 5' non-coding region of 166 nucleotides (Cho, unpublished).

The deduced amino acid sequence of the human HSMT3 cDNA exhibits 51.1% identity and 69.6% similarity with that of yeast Smt3 protein. The strong evolutionary conservation between human HSMT3 and yeast Smt3 proteins indicates its functional significance. However, the exact function of these proteins remain to be elucidated.

ACKNOWLEDGMENTS

We thank Dr. F. Johnson and A. Enomoto for reading the manuscript and helpful suggestions.

REFERENCES

- 1. Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995) Science 270, 1591–1594.
- 2. Bloom, K. (1993) Cell 73, 621-624.
- Saitoh, H., Tomkiel, J., Cooke, C. A., Ratrie, H. III, Maurer, M., Rothfield, N. F., and Earnshaw, W. C. (1992) Cell 70, 115–125.
- 4. Meluh, P. B., and Koshland, D. (1995) Mol. Biol. Cell 6, 794-807.
- 5. Brown, M. T. (1995) GENE 160, 111-116.
- 6. Brown, M. T., Goetsch, L., and Hartwell, L. H. (1993) J. Cell Biol. 9, 2544-2550.