

Isolation and Characterization of a cDNA for Human, Mouse, and Rat Full-Length Stem Cell Growth Factor, a New Member of C-Type Lectin Superfamily

Hiroyuki Mio,¹ Naofumi Kagami, Sachiko Yokokawa, Hironori Kawai, Satoshi Nakagawa, Kyoko Takeuchi, Susumu Sekine, and Atsunobu Hiraoka*

Tokyo Research Laboratories, Kyowa Hakko Kogyo Company, Ltd., 3-6-6 Asahi-cho, Machida, Tokyo 194-8533, Japan; and
*Department of Internal Medicine, Osaka Dental University, 1-5-17 Ohtemae, Chuo-ku, Osaka 540-0008, Japan

Received June 26, 1998

cDNA encoding stem cell growth factor (SCGF; 245 aa), a novel human growth factor for primitive hematopoietic progenitor cells, has been previously reported (Hiraoka, A., Sugimura, A., Seki, T., Nagasawa, T., Ohta, N., Shimonishi, M., Hagiya, M. and Shimizu, S. *Proc. Natl. Acad. Sci. USA* 94, 7577–7582, 1997). Here we report the cloning and characterization of a full-length SCGF cDNA. This protein consists of 323, 328 and 328 aa in the human, murine and rat forms, the latter two of which share 85.1% and 83.3% aa identity, and 90.4% and 90.4% aa similarity to the human protein, respectively. Because the newly identified human clone encodes the protein longer by 78 aa than that previously identified, we term the longer clone as hSCGF- α and the shorter one as hSCGF- β . The computer-assisted homology search reveals that SCGF is a new member of the C-type lectin superfamily, and that SCGF shows the greatest homology to tetranectin among the members of the family (27.2–33.7% aa identity and 46.0–53.6% aa similarity). SCGF transcripts are detected in spleen, thymus, appendix, bone marrow and fetal liver. Fluorescent *in situ* hybridization mapping indicates that the SCGF gene is located on chromosome 19 at position q13.3 for human form and on chromosome 7 at position B3-B5 for murine form, which are close to *flk-2/flt3* ligand and interleukin-11 genes of both human and murine species. © 1998 Academic Press

A cDNA for novel hematopoietic cytokine, stem cell growth factor (SCGF), is originally isolated from cDNA library prepared from a human myeloid cell line, KPB-

M15, derived from chronic myelogenous leukemia in blast crisis (1,2). SCGF cDNA encodes 245 aa protein without N-linked glycosylation. No significant homology with the database in the EMBL, GenBank and Swiss-Prot is found for the SCGF nucleotide and amino acid sequence. SCGF mRNA is expressed by myeloid and stromal cells, but not by lymphoid cells.

Primitive hematopoietic progenitor cells are stimulated by synergistic action of multiple growth factors, including colony-stimulating factors (CSFs) and interleukins (ILs); in particular, such early acting cytokines as stem cell factor (SCF or *c-kit* ligand) (3–5) and *flk-2/flt3* ligand (6) play a major role. Human SCGF alone does not exhibit colony-stimulating activity, but burst-promoting activity and granulocyte/macrophage (GM)-promoting activity for erythroid and GM progenitor cells in primary agar culture with erythropoietin and GM-CSF, respectively. It further supports survival or growth of hematopoietic progenitor cells through a short-term liquid culture of human bone marrow cells. Consequently, SCGF could mediate an interaction between primitive hematopoietic progenitor and stromal cells within the hematopoietic microenvironment, in conjunction with other growth factors.

In this paper, we isolate and characterize full-length human (h), murine (m) and rat (r) SCGF cDNAs, indicating that SCGF is a new member of the C-type lectin superfamily, and that its mRNA is exclusively expressed within the hematopoietic tissues.

MATERIALS AND METHODS

DDBJ/EMBL/GenBank accession numbers. The accession numbers for the sequences reported in this paper are AB009244 for hSCGF- α , AB009245 for mSCGF and AB009246 for rSCGF.

Cloning of the SCGF cDNA. Oligonucleotides 5'-GAGTCCAGCTTAATGCAGGC A -3' and 5'-CTAGAAGGGG AACTCGCAGA C-3' were carried on PCR to amplify the entire length of hSCGF with the

¹ To whom correspondence should be addressed at Pharmaceutical Research Institute, Kyowa Hakko Kogyo Co., Ltd., 1188 Shimotogari, Nagaizumi-cho, Sunto-gun, Shizuoka 411-8731, Japan. Fax: +81-559-86-7430. E-mail: hmio@kyowa.co.jp.

template of first strand cDNA synthesized from human bone marrow poly(A) RNA (Clontech, Palo Alto, CA). The thermal cycling reaction was performed in the presence of 10% DMSO with ExTaq polymerase (Toyobo, Tokyo, Japan); denaturation at 94°C for 1 min, annealing at 55°C for 2 min, and extension at 72°C for 2 min. The cDNA library derived from murine pre-adipocyte cell line, MC3T3-G2/PA6 (Riken, RCB1127) (7), was screened to isolate the clone containing the sequence that could be amplified under the same PCR condition as for hSCGF. rSCGF cDNA was amplified with the primers containing oligonucleotides 5'-ATTGTTGGTGC TGGGAAGCCC AGCT -3' and 5'-TCCTGGGCAG AGACCGTTTC TCTA -3', sequences of which were based upon 5' and 3' untranslated region of mSCGF, respectively. The template for PCR was synthesized from total RNA of rat osteosarcoma cell line, ROS-17/2.8-5 (Riken, RCB0462). The cloned DNA was subcloned into pBluescript II SK(-) vector to create pBluescript-SCGF and sequenced using DNA sequencer model 4000 (LI-COR Inc., Lincoln, Neb) (8).

Homology search. The deduced amino acid sequence of SCGF was subjected to homology search with the database of Swiss-Prot and PIR, using a program of BLASTP1.4.9 (9). Motif sequence analysis was performed with the database of PROSITE and BLOCKS using MacPattern 3.4 (10). A multiple alignment was operated for amino acid sequences using the CLUSTAL W Multiple Sequence Alignment Program 1.6 (11), and an unrooted phylogenetic tree was visualized by TreeView 1.4. Signal sequence cleavage sites were analyzed by AnalyzeSignalase, ver. 2.03 (12), and transmembrane regions were by TopPred II, ver. 1.3 (13).

RNA hybridization. Northern blot analysis of RNA samples was performed using Clontech human multiple tissue Northern blots immune systems. Antisense RNA probe was generated using T3 RNA polymerase and pBluescript-SCGF linearized with Hind III.

Fluorescent in Situ hybridization (FISH) mapping. The procedure for FISH detection was performed according to the method previously reported (14,15). Briefly, slides were baked at 55°C for 1 hour. After RNase treatment, the slides were denatured with 70% formamide in $2 \times$ SSC at 70°C for 2 min, followed by dehydration with ethanol. Probes were denatured at 75°C for 5 min, in a hybridization mixture containing 50% formamide and 10% dextran sulfate. Probes were loaded onto the denatured chromosomal slides. After overnight hybridization, slides were washed and detected as well as amplified. FISH signals and the DAPI banding pattern was recorded separately by taking photographs, and the assignment of the FISH mapping data with chromosomal bands was achieved by superimposing FISH signals with DAPI banded chromosomes (15).

RESULTS

Isolation of Human, Mouse and Rat cDNAs

Encoding SCGF

We have previously found that a human myeloid cell line, KPB-M15, produces SCGF, and cloned a sole cDNA encoding 245 aa protein using expression cloning method (2). To obtain another transcripts of SCGF, RT-PCR analysis was performed with a template of single strand cDNA prepared from human bone marrow poly(A) RNA. Five clones were analyzed independently, and appeared the same nucleotide sequence. The resulting new cDNA had an open reading frame capable of encoding a protein of 323 aa, which was longer by 78 aa than the previously isolated clone (2). Here we named the longer clone as hSCGF- α and the shorter one as hSCGF- β .

```
GAAGCTGGCAGAAGAAGGTCAAGGGGCTTGTGAGCTGCCACACAGACTGGGACACTTGTCT 60
AGGCTATACAGCAGTCTACCCCTGGCACTTCTGACCTCTCTACTATTGGTGCTGGGA 120
AGCCACAGCTGGATGCAGGCGAGCTGGCTCTTGGGGGCCCTAGTGGTCCCTCAGCTTTTGA 180
      M Q A A W L L G A L V V P Q L L
GTTTGTGTCATGGAGCCGAGGTCCTGGGAGGAGTGGGAGGAGGCTGGGAGGTTGCC 240
S F G H G A R G P G R E W E G G W G G A 36
TGGAGGAGGAGAGAGAGCGGGAGTCAGAGATGTTGAAGATCTCCAGGAGGCCCTAGGGC 300
L E E E R E R E S Q M L K N L Q E A L G 56
TGGCCACTGGGGTGGGAAATGAGGATAATCTTCTGAAACCGTGGAGCAAGAGGCTCT 360
L P T G V G N E D N L A E N P E D K E V 76
GGGAGACCAGAGACTCAAGGGGAGAGAGGAGAGGAAATCACCACAGCAGCTTCTT 420
W E T T E T Q G E E E E E E I T T A P S 96
CTAGTCCCAACCTTCCCGAGCCTTCTCCACACAGGAGGACACTGTCACTTACATCT 480
S S P N P F P S P S T P T G E D T V T Y I 116
TGGGCCGCTTGGCCAGCCTCGATGCAGGCTACACCAATGACAGTCCGCTGCGACGTTT 540
L G R L A S L D A G L H Q L H V R L H V 136
TGGACACCCGTGTGGTGGAGTGCACCGAGGGCTGCGGAGCTGCGGGATGCTGCGAGTG 600
L D T R V V E L T Q G L R Q L R D A A S 156
ACACCCGCGACTCAGTGAAGCCCTGAAGAGGTCAGGAGCGTGCTGAGCAGGAGCAGC 660
D T R D S V Q A L K E V Q D R A E Q E H 176
GCCGCTTGGAGGGCTGCCTGAAGGGCTGCGCCTTGGCCACAAGTGCTTCTGCTCTCGC 720
G R L E G C L K G L R Q L R K F L S 196
GAGACTTCGAGACCCAGGCGCGCGCGAGCGCGGTGCAAGGCGCGAGGTGGGAGCTTAG 780
R D F E T Q A A A Q A R C K A R G G S L 216
CACAGCCTGCGGACCGCCAGCAATGGATGCGCTAAGCGGTACTTACGCGCCGCTCTCG 840
A C Q P A D R Q Q M D A L S R Y L R A A L 236
C C C C T A C A A C T G G C G G T G G C T G G A G T G C A C G A T C G G C G C T C G A 900
A P Y N W P V W L G V H D R R S E G L Y 256
TTTTGAGAAACGGCCAGGCGGTGTCTTCTTCCGCTGGCAGCCGCGATTACAGCTGGAAT 960
L F E N G Q R V S F F A W H R A F S L E 276
CGGCGCCGACCTAGTGCAGCAACACATCCATCAGCCGCGATCAGCCCAATGGCGGCG 1020
S G A Q P S A A T H P L S P D Q P N G G 296
TCCTGGAGAACTGCGTGGCCAGGCTCAGACGACGCTTCTGTTGGGAGCATGACTGTG 1080
V L E N C V A Q A S D T G A G S W W D H D C 316
AGCGCGCCCTACTTCTGCTGCGAGTTCCCTCTGAGAGAACCGGTCTCTGCCAGGAG 1140
E R R L Y F V C E F P * 328
CTCTAGTGACATTTTGACACCGTACACCGCGCACCTATTGTTAGGGGCTGGGAGTCGC 1200
TCAGAGATTAAGCGTGACCATGAATATATTTAATCAGAAGAGGTTTTTATTTAGATA 1260
TGGCAGCCAGACTGATTGGGCGAGGTGCTCTGAGATTGCTTCCAGATGCAATTAT 1320
CAGCCAGGGATTTTAAAGGCAACCCACAAGATTGCATGTAGCTGCTTACATGTAGG 1380
CGGAGCATAAAAATTTAAATATAAAAAA 1411
```

FIG. 1. Nucleotide and deduced amino acid sequence of the murine SCGF cDNA. Numbering is from the 5' end of the nucleotide sequence (top) and from initiation methionine of the amino acid sequence (bottom). An asterisk is a stop codon.

Homologous mSCGF cDNA (1411 bp; Figure 1) was amplified from a cDNA library prepared from the murine preadipocyte cell line, MC3T3-G2/PA6, using PCR with hSCGF primers. The predicted coding region of mSCGF was 328 aa in length, and showed 85.1% aa identity and 90.4% aa similarity to the human homologue (hSCGF- α). Homologous rSCGF cDNA was amplified from a single strand cDNA originated from rat osteosarcoma cell line, ROS-17/2.8-5, by the method of RT-PCR. Five clones were analyzed independently, and appeared the same nucleotide sequence. The predicted coding region of rSCGF was 328 aa in length, and showed 93.3% and 83.3% aa identity and 96.6% and 90.4% aa similarity to the murine homologue and human homologue, respectively.

Hydrophilicity analysis of the human, murine and rat SCGF proteins indicated the presence of signal se-

SC6F(h)	1	--MQAAWLLGALVVPQLLGFHGARGAEREWE ^{•••} GGWGAQEER-EREALMLKHLQALGLPAG ^{•••} RGDE-NPAGTV--EGKEWMEEDQGEETEE-EATPT	93
SC6F(m)	1	--MQAAWLLGALVVPQLLGFHGARGPGREWE ^{•••} GGWGALEER-ERESQMLKNLQALGLPTGVGNEDNLAENP--EDKEVWTTETQGEETEE-EITTA	94
SC6F(r)	1	--MQAAWLLGALLVPLLSFGHGARGHGKEWGVWGALEER-DRESLMKLNQALGLPTGVGNKNDLAENS--EGKEVWTEATQGEETEE-EITTA	94
TN	1	-----MELWGAAYLLLCFLSLLTQVYTE	22
CPCP	2001	APEASREDSGSPDLSETTSFAHEANLERS ^{•••} SGLSSTLTFQGEASAAPEVSGESTTTSDVGTAPGLPSATP T-ASGDRTEISGDLSGHTSGLGVVIS	2099
ASGR-1	1	-----MTKEYQDLQHLNDESDHQLRKGP ^{•••} PPPPQLLR-LCSGPRLLLLSLGLSLLL ^{•••} VVVCVIGSQNSQ-LQELRGLRETFSNFTASTEAAQVK	89
IgE-FcR	1	-----MEEQYSEIEELPR-RRCCRRGTQIVLLGLVTAALWAGLLTLL ^{•••} LHHWDTTQSLKQLEERAARNSVQ-VSKNLESHHGDMQAQKSQ-SQTIS	89
MBP-C	1	-----MSLFPSLPLLLLSMAVASYSETVTGCEAQKTCPAV ^{•••} IACSSPGINGF-PGKDRGDTGKGEKPEGQ-GLRGL	69
PSP-D	35	CTLVMCSSESGLPGRDGRDGRGPRGKGPGLPAAAGAGMPGAG ^{•••} VPVKGDNGSVGEPGPKGDTGSPGPPGPPGVPAGREGALGKGNIGPQK	134
L-Selectin			
NKG2-A&B	1	-----MDNQGVISYDLNLPNPKRQKRPKGNKSSILATEQEITYAELNLQKASQDFQGN	55
CD72	26	GQDPGADDDGEITYENVQPAVLGVPSLASSVLGDAAVKSEPTASWRAVTS ^{•••} PAVGRI ^{•••} LCPTTCLRYLLGLLLTCLLLGVTAICLGVRYLQVSQQL	125
MMR	1005	FSAWTGLNDVNSEHTFLWTDGRGVHYTNWKGYPGGRSSLSYEDADC ^{•••} VWIIIGGASNEAGKWMDDTCD ^{•••} SKRGYICQTSRDPSTLNPPATIQDGFVKYK	1104
PAP-1			
MBP	1	-----MKPLLLALLFGAVSALH-LRSETSTFETPLGAKTLP-DEET	41
SC6F(h)	94	PSG-----P-SPSPTPEDIVTYILGRLAGDAG--LHQLHVRH--ALDTRVVELTQGLRQLRNAAGDTRDAVQALQEAQ-RAE-----R	169
SC6F(m)	95	PSSSPN--PFP-SPSPTPEDIVTYILGRLASLDAG--LHQLHVRH--VLDTRVVELTQGLRQLRDAASDTRDSVQALKEVQD-RAE-----Q	174
SC6F(r)	95	SSSPT--PFP-SPSPTPEDIVTYILGRLASLDAG--LHQLHVRH--VLDTRVVELTQGLRRLRDAASDTRDSVQALKEVQD-RSE-----Q	174
TN	23	P-----PTQPKKI VN--AKKDVNTKM--FEELKSRLD--TLAQEVALKE-----Q-QALQTV-----	70
CPCP	2100	TSIPES--EWT--QQTQRP ^{•••} AETHLEI ^{•••} ESSLLYSGEF--THVETATSS--PTDASIPASPEWKRESESTAAPARSACEEPGCGAGTCKE-----T	2181
ASGR-1	90	GLS-----TGGGNVGRMKSLSESLEKQKQD--LSEDHSSLL--LHVQFVS ^{•••} DLRSL--CQMAALQGN ^{•••} SGER-----	151
IgE-FcR	90	QLEEL--R--AEQQLKSDQL ^{•••} ELSWNLGLQAD--LSFSQSE--NERNEASD ^{•••} LLERL--EEVTKL ^{•••} RMELQVSSG-----	157
MBP-C	70	QGPP-----GKLGPPNPGSPGSPGKQKGD--PGKSPDGS--SLAASERKALQT-----EMARIKKWL T-----	127
PSP-D	135	PGPKE--AGPKGEVGA ^{•••} PMGQSGAGRLAGPKGERGVP ^{•••} GERGVPN--TGAAGSAGAMPQSPGARGPPGLGKDGKIPGDKGAKGESGLPDVASLRQ	229
L-Selectin	1	-----MIFPWKCQ--STQRDLWNIFK-----LWGWTML-----	26
NKG2-A&B	56	D-----KTYHCKDLPSAPEKLIVGILGII--CLILMASVY--TIVVIPSTLIQRHN--NSSLNTRTQKARHCG-----	117
CD72	126	QQTNRVLEVTNSSRLQLRLK ^{•••} ITQLGQSAEDLQGRRLAQSQEALQVEQRAHQAAGQLQACQADRQKTETLQSEEQRRALQKLSNMENRLKPFET	225
MMR	1105	SSYSLMRQKFWHEAETCYCKLHNSLIASILDPSYNAFANLQWES ^{•••} NERWVIALNSLNTDNQYTTDKWRVRYTNWAAD ^{•••} EPKLKACVYLDLQYKTAHC	1204
PAP-1	1	-----MLPPMALPSV--SWMLLSCLM--LLSQVQGE ^{•••} EPQ-----RELPSARI-R-----	39
MBP	42	P-----EQEMEETPCRELEEEECWGS--EDASKDGA--VESISVPDMVDKN-----LTCPEED-----	94
SC6F(h)	170	EHGRLEGQLKGLRLGHK-----CFLSRDFE--AQAAQARCTAR--GGSLAQPADRQDMALTRYLRAALAPYNWVWL ^{•••} G	241
SC6F(m)	175	EHGRLEGQLKGLRLGHK-----CFLSRDFE--TQAAQARCKAR--GGSLAQPADRQDMALTRYLRAALAPYNWVWL ^{•••} G	246
SC6F(r)	175	EHGRLEGQLKGLRLGHK-----CFLSRDFE--TQAAQARCKAR--GGSLAQPADRQDMALTRYLRAALAPYNWVWL ^{•••} G	246
TN	71	-----CLKGTQVHKM-----CFLAFTQTK--TFHEASED ^{•••} ISR--GGTLSTPQTGSENDALYEYLRQSGV-NEAEIWL ^{•••} G	134
CPCP	2182	EGHVICLOPPGYTGEHCNDQEVCEE-----GWNKYQGHQYRHPDRE--TWVDAERRDREQ--QSHLSIVTPEEQ-EFVNNAQDYQ-----WIG	2263
ASGR-1	152	-----TCQPVNWEHERS-----CYWFSRSGK--AWADADNYCRLE--DAHLVVVTSWEEQ-KFVQH ^{•••} HGPVN--TWMG	213
IgE-FcR	158	-----FVNC ^{•••} TP ^{•••} PEKWINFQRK-----QYFYGKGTK--QWVHARYACDDM--EGQLVSIHSP ^{•••} EEQ-DFLTKHASHTG--SWIG	222
MBP-C	128	-----FSLGKQVHKM-----CFLTNGEIM--TFEKVKALCVKF--QASVATPRNAEN-GAIGNLIK ^{•••} EEA--FLG	185
PSP-D	230	QVEALQGVQHLQAAFSQYK ^{•••} VELFP-----NGQSVGEIKF ^{•••} TAGFVK--PFT ^{•••} EAQLLCTQA--GGQLASPRSAEN-AALQQLV ^{•••} VAKN--EAAFLS	314
L-Selectin	27	-----CCDFLAHHTD-----CWTYHYSEKPMNWRARRFCRDN--YTDLVAIQNKAEI ^{•••} EYLEKTL ^{•••} PFSS--YYWIG	90
NKG2-A&B	118	-----HOPEEWI ^{•••} TYSNS-----CYYIGKERR--TWEESSLACTSK--NSSLLSIDNEEM-KFLS-IISPS--WIG	176
CD72	226	CGSADTCPSGWI ^{•••} MHKS-----CFYILSTSK--NWQESQKQCE ^{•••} TLSSKATFSEIYPQSHSYFYFLNSLLP ^{•••} N6GS-GNSYWTG	300
MMR	1205	NESFYFLKRSDEIPATEPQLPGRCPESDHTAWIPFHGHYI ^{•••} ESSYTR-NWQASLECLRM--GSSLVSI ^{•••} ESAAES-SFLSYRVEPLKS-KTNFWIG	1298
PAP-1	40	-----CPKGSKAYGSH-----CYALFLSPK--SWTDADLACQKRP--SGNLVSVLSGAEG-SFVSSLVKSIGNSY ^{•••} YVWIG	105
MBP	95	-----TVKVGIPGCQT-----CRYLLVRSLG--TF ^{•••} QAWFTCR ^{•••} RCY--RGNLVS ^{•••} IHNFIN-YRIQCSVSALN--GQGVWIG	160
SC6F(h)	242	VHD--RRAE--G-LYLFENGQR-VSFFAWHRSRPELGAQPSA ^{•••} PHPLSPDQPN ^{•••} GGILENCAQASDD--GSWHDHCERRLY-VYCEFFP--	323
SC6F(m)	247	VHD--RRSE--G-LYLFENGQR-VSFFAWHRAFSLESQAQPSAATHPLSPDQPN ^{•••} GGVLENCVAQASDD--GSWHDHCERRLY-VYCEFFP--	328
SC6F(r)	247	VHD--RRSE--G-LYLFENGQR-VSFFAWHRAFSLESQAQPSAASHPLSPDQPN ^{•••} GGILENCAQASDD--GSWHDHCERRLY-VYCEFFP--	328
TN	135	LND--MAAE--G-TWVDMTGAR-IAYKNWET--EITAQ-----PD--GGKTENDAVLSGAAN--GKMF ^{•••} DKRQDQLP-YIQFGIV-	202
CPCP	2264	LND--RTIEG--DFRWSGDHP-MQFENWRP--NQPDN-----FFAAG-----EDCVMIWHEK--GEWDVPCNYHLP-FTCKKGTVA	2331
ASGR-1	214	LHD--QNGP--W-KWVDGTYE-TGKFNWRP--EQPDD-----WYGHGLG-GG-EDCAHFTDDG--R ^{•••} NDV ^{•••} VQRPYR-VYCETELDC	283
IgE-FcR	223	LRN--LDLK--G-EF ^{•••} IWVDGSH-VDSYNWAPG--EPTSR-----SQG-----EDCVMMRSGG--R ^{•••} NDAF ^{•••} CDRLKGA ^{•••} WCDRLATC	288
MBP-C	186	ITD--EKTE--G-QFVDLTGNR-LTYTNWNE--GEPNN-----AGSD-----EDCVLLKNG--Q ^{•••} NDV ^{•••} POSTSHL-AVCEFFI--	248
PSP-D	315	MTD--SKTEG--KFTYPTGES-LVYSNWP--GEPND-----DGG--S-EDC ^{•••} VEIFTNG--K ^{•••} NDRAC ^{•••} G ^{•••} EKRL-VYCEFF--	375
L-Selectin	91	IRK-IGGVI--TWVGTKSLTEEAENWD--GEPNN-----KKNK-----EDC ^{•••} VEI ^{•••} YIKR ^{•••} NDAG ^{•••} K ^{•••} ND ^{•••} ACH ^{•••} KLKA-ALCYTASQ	161
NKG2-A&B	177	VFR--NSS--HHPWVTMG-LAFKHEIK--DSN-----AEL-----NCAVLQVNR--L ^{•••} KAQ ^{•••} GSII-YHCKHL-	233
CD72	301	LSS--NK--DWKLTDDTQ--RTRTYAQ-----SSK-----CNKVHKTWS--W ^{•••} MTLES ^{•••} SCRSLP-YICEMTAFR	356
MMR	1299	LFR--NVGE--TWLWINSNP-VSFWNNT--GDPSG-----ERN-----DCVALHASS--G ^{•••} FSNI ^{•••} HCSYK ^{•••} G-YICKRPKII	1361
PAP-1	106	LHPTQGTGTEPNGEWSSDV-MNYFAWERN--PSTIS-----SPG-----HCASLSRSTAF--L ^{•••} R ^{•••} NDY ^{•••} NCWRLP-YVCKFTD-	175
MBP	161	GRIT-GSRG-CRRFQWVDSR-WNFAYWAA--HQPWS-----RG-----GHCVALCTR6--G ^{•••} Y ^{•••} NR ^{•••} RAH ^{•••} CLRLP-IC ^{•••} SY--	222

FIG. 2. Alignment of SC6F with members of the C-type lectin superfamily. The putative signal sequence is underlined. The different amino acids among hSC6F, mSC6F and rSC6F (solid circles), the amino acids absent in hSC6F- β (a line with arrows at both ends) and 5 conserved Trps upstream to the second Cys from the C-terminal (cross mark) are denoted on the top. Polyglutamic acidic region is boxed and Pro/Ser/Thr-rich PT box is boxed with dots. RGD sequence is lettered in white with black background. A framework of conserved Cys or Trp is arranged in columns, and 4 conserved amino acids (VDYV) are circled in the consensus pattern of the CRD of C-type lectin motif sequence (doubly underlined). Abbreviations used are as follows: h; human, m; murine, r; rat, TN; tetranectin, CP; cartilage proteoglycan core protein, ASGR; asialoglycoprotein receptor, MBP-C; mannose-binding protein-C, PSP; pulmonary surfactant associated protein, NKG; natural killer group, MMR; macrophage mannose receptor, PAP; pancreatitis-associated protein, MBP; major basic protein.

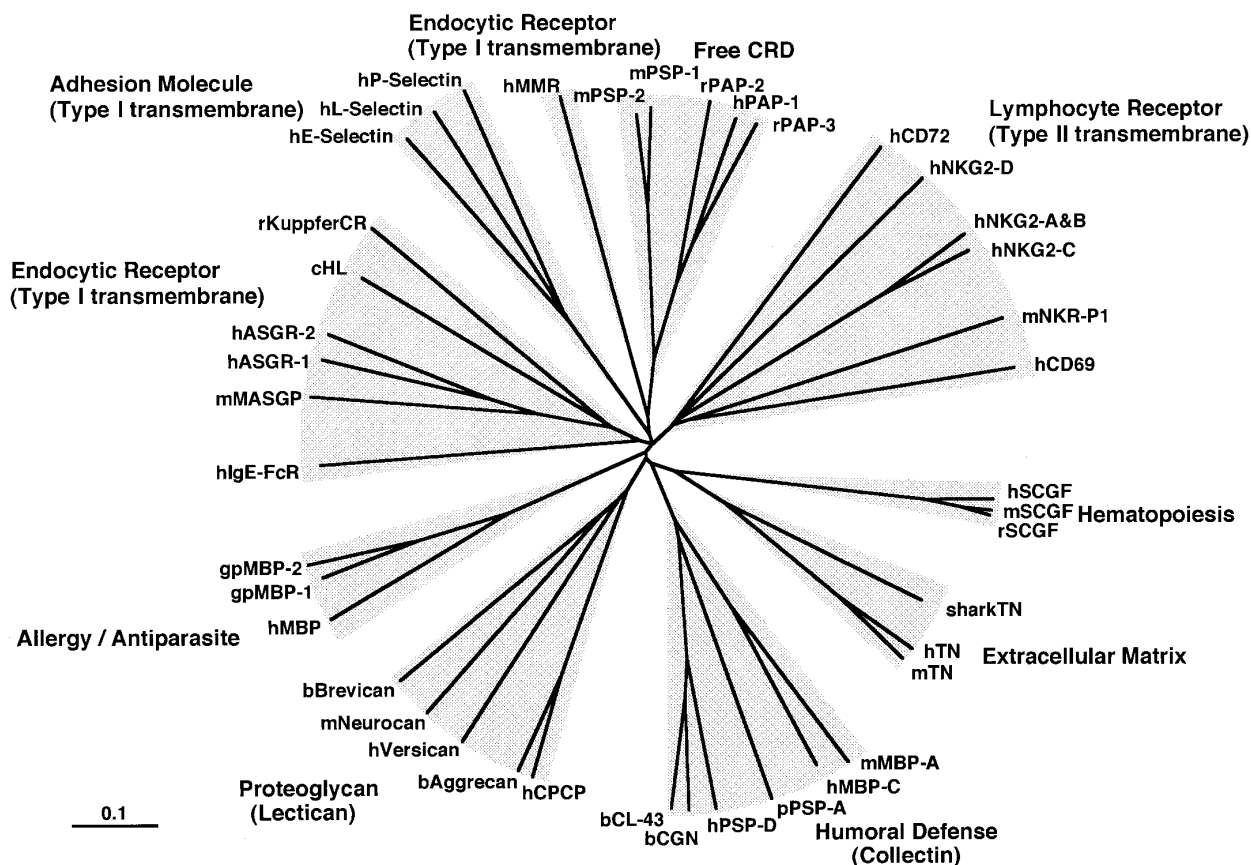


FIG. 3. Unrooted phylogenetic tree of the C-type lectin superfamily. Scale bar represents branch length corresponding with 0.1 amino acid substitution per site. Abbreviations used are as follows: p; pig; b; bovine; gp; guinea pig; c; chicken; CGN; conglutinin, CL; collectin, MASGP; macrophage asialoglycoprotein-binding protein, HL; hepatic lectin, PSP; pancreatic stone protein, NKR; natural killer cell receptor. Refer to the legend of Figure 2 for other abbreviations.

quence and the absence of putative transmembrane region. There was no potential N-linked glycosylation sites and only hSCGF contained RGD sequence at amino acid position 61 (Figure 2).

Characterization of SCGF *in Silico*

The computer-assisted homology search revealed that the primary structure of hSCGF- α , mSCGF and rSCGF shared a motif sequence designated as C-type lectin (Figure 2). The conserved domain, which seems to function as a calcium-dependent carbohydrate-recognition domain (CRD) (16), is shown below. This domain consists of about 110 to 130 amino acid residues containing six cysteines as well as other optional amino acids. Four cysteines are perfectly conserved and involved in two disulfide bonds, and two more cysteines are optionally conserved. A striking consensus pattern is C-[LIVMFATG]-x(5,12)-[WL]-x-[DNSR]-x(2)-C-x(5,6)-[FYWLIVSTA]-[LIVSTA]-C. Among the members of C-type lectin superfamily identified thus far, SCGFs

except hSCGF- β , showed the greatest homology with tetranectin (TN), a plasminogen kringle 4-binding protein, especially with shark TN at a slightly higher score of homology than with TN of other species. hSCGF showed 32.2%, 27.2% and 33.7% aa identity, and 48.0%, 46.5% and 53.6% aa similarity to hTN (17), mTN (18) and shark TN (19), respectively. SCGFs appeared close to a member of tetranectin, but formed a cluster separate from the other members of C-type lectin superfamily (Figure 3).

Northern Analysis of Human SCGF mRNA

Figure 4 shows Northern blot analysis of the expression of hSCGF transcripts in hematopoietic tissues. The predominant RNA band in the blots had a size of 1.4-1.6 kb. SCGF transcripts were seen in many human tissues, including spleen, thymus, appendix, bone marrow and fetal liver, whereas the expression in peripheral blood leukocytes was low. An additional transcript of 2.2-2.4 kb was detected in thymus and bone marrow.

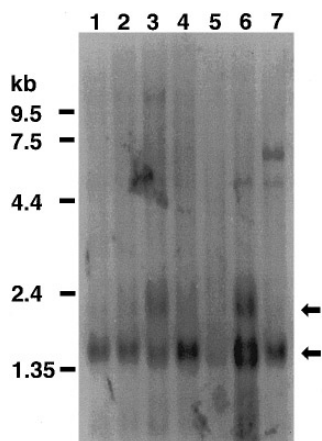


FIG. 4. Northern blot analysis of SCGF transcripts. RNAs isolated from the human tissues were probed with labeled antisense RNA encoding human SCGF. Lane 1; spleen, lane 2; lymphnode, lane 3; thymus, lane 4; appendix, lane 5; peripheral blood leukocytes, lane 6; bone marrow and lane 7; fetal liver. Each position of RNA size markers is given in kb on the left. Arrows indicate SCGF transcripts of 1.4-1.6 kb and 2.2-2.4 kb.

Chromosomal Mapping of SCGF Gene

Chromosomes from peripheral blood and spleen lymphocytes were FISH analyzed in order to determine where the SCGF gene resided in the human and murine genome, respectively. Under the condition used, the hybridization efficiency was approximately 61% for hSCGF probe. The DAPI banding was used to identify the specific chromosome, and the assignment between the signal from probe and the long arm of chromosome 19 was achieved. There was no additional locus picked by FISH detection, therefore hSCGF gene was located on human chromosome 19, region q13.3 (Figure 5A and B). mSCGF gene was mapped on chromosome 7, region B3-B5 with the hybridization efficiency of 64% for mSCGF probe (Figure 5C and D).

DISCUSSION

In this paper, we have described the cloning and characterization of a full-length SCGF cDNA. Unlike SCF and *flk-2/flt3* ligand, which contain transmembrane region, any hydrophobic region other than N-terminal signal sequence is neither seen in hSCGF- α nor in hSCGF- β , indicating that this protein exists in a soluble form. The newly isolated hSCGF- α is significantly different from the previously cloned hSCGF- β in that 78 aa is inserted into the latter to form the former. Analysis of RT-PCR was carried out using mRNA prepared from human bone marrow and KPB-M15 cells, from the latter of which SCGF- β cDNA was originally isolated. As far based on the same condition of amplification as used for cloning of SCGF- α cDNA,

a band corresponding to SCGF- β was undetectable (data not shown). Currently the reason of the rare expression of SCGF- β mRNA in KPB-M15 and human bone marrow cells is unclear. Further experiments to identify the genome DNA and how to regulate the gene expression of the two forms should be progressed.

hSCGF- β lacks a conserved CRD, whereas hSCGF- α , mSCGF and rSCGF have well conserved CRD, and are supposed to belong to C-type lectin superfamily; that is, the inserted 78 aa sequence occupies a most half of the CRD (Figure 2). According to the phylogenetic tree of C-type lectin superfamily, shark TN is closer to SCGFs than human or murine TN. Since C-type lectins generally bind through CRD as a ligand or counter-receptor in a calcium-dependent manner to carbohydrates of their specific receptors (16), SCGF could interact with carbohydrates. While CD94/NGK2A,B,C are another member of C-type lectin superfamily (Figure 3), carbohydrates found on HLA-E as their receptor are not required for binding, but form additional points of interaction (20). Another exceptional example is a lectican; chondroitin sulfate proteoglycans including aggrecan, versican, neurocan and brevican (Figure 3) bind tenascin-R by protein-protein interactions independent of carbohydrates (21). Since SCGF- β lacks a half of CRD, the affinity in interaction with the receptor might be decreased. Although the activity of SCGF- α on colony-formation has not been measured yet, the existence of CRD can play a role to

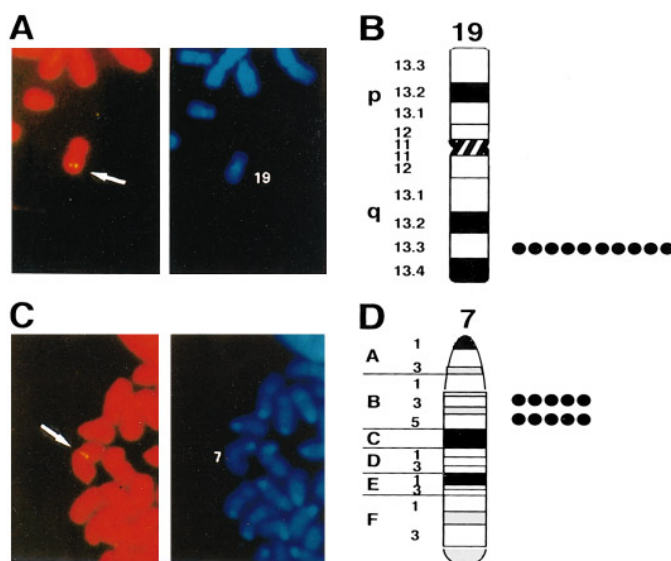


FIG. 5. FISH mapping of SCGF gene. An example of FISH mapping of hSCGF (A) and mSCGF (C) is shown; the FISH signals on chromosome (left panel) and the same mitotic figure stained with DAPI (right panel) to identify chromosome 19 (A) and chromosome 7 (C). Diagram of FISH mapping results is shown for hSCGF (B) and mSCGF (D). Each dot represents the double FISH signals detected on chromosome.

enhance the activity. hSCGFs, both α and β forms, contain RGD sequence at amino acid position 61, whereas mSCGF and rSCGF do not, indicating that little, or if any, interaction with adhesion molecules would occur at this site.

Northern blot analysis carried out previously using tissues and cell lines indicates that SCGF functions within the hematopoietic microenvironment (2). The present data are compatible with the previous ones; SCGF mRNA is expressed at a high level within the hematopoietic tissues, particularly bone marrow. Peripheral blood leukocytes show a low content of SCGF mRNA, which could be ascribed not to lymphocytes but to contaminated monocytes in the cell population. In fact, the expression of SCGF mRNA was determined using monocytoid cell lines (2). Therefore, it is likely that the receptor for SCGF is expressed on primitive hematopoietic progenitor cells, which, when identified, should provide valuable insight into the biological function of SCGF.

SCGF gene is located most likely on chromosome 19q13.3 for hSCGF and chromosome 7B3-B5 for mSCGF. Among hematopoietic growth factors, *flt-2*/*flt3* ligand (22) and IL-11 (23) genes are located on the same chromosomal region for both human and murine genes, while hTN gene is found on chromosome 3q22-p21.3 (24). Multiple human genes encoding hematopoiesis-related growth factors or receptors, e.g. GM-CSF, M-CSF receptor, IL-3, IL-4, IL-5, IL-9, IL-12B, IL-13, fibroblast growth factor (FGF) 1, FGF receptor 4, platelet-derived growth factor receptor and *flt4* are clustered on the long arm of chromosome 5, implicating that they are involved in the pathogenesis of acute lymphoblastic leukemia with karyotype t(5;14)(q31;q32) (25), malignant histiocytosis with karyotype t(2;5)(p23;q35) (26) and 5q- syndrome (27, 28); in the last case, clonal evolution with karyotype del(5)(q13q35) is associated with acute nonlymphoblastic leukemia or myelodysplastic syndrome. Two causative genes on chromosome 19q13 have been reported for hematological diseases; congenital hypoplastic (Diamond-Blackfan) anemia (DBA) (29) and *bcl-3* (30) on 19q13.2. Translocation occurs between immunoglobulin heavy chain gene and *bcl-3* in a certain case of chronic lymphocytic leukemia with karyotype t(14;19)(q32;q13). SCGF gene could be one of the candidate genes involved in certain types of hematopoietic disease, and it would be intriguing to examine karyotype abnormalities in 19q13.3 or SCGF gene mutation in such diseases.

ACKNOWLEDGMENTS

We thank Mayumi Ibonai and Naoko Saito for technical assistance. Chromosomal mapping was supported by SeeDNA Biotech Inc. (Ontario, Canada) and synthesis of oligonucleotides was supported by Sawaday Technology Co., Ltd. (Tokyo, Japan). This work was sup-

ported in part by Grant-in-Aid for Scientific Research (B) from the Ministry of Education, Science, Sports and Culture, Japan (Grant No. 10557094).

REFERENCES

1. Hiraoka, A., Ohkubo, T., and Fukuda, M. (1987) *Cancer Res.* **47**, 5025–5030.
2. Hiraoka, A., Sugimura, A., Seki, T., Nagasawa, T., Ohta, N., Shimonishi, M., Hagiya, M., and Shimizu, S. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 7577–7582.
3. Zsebo, K. M., Wypych, J., McNiece, I. K., Lu, H. S., Smith, K. A., Karkare, S. B., Sachdev, R. K., Yuschenko, V. N., Birkett, N. C., Williams, L. R., Satyagal, V. N., Tung, W., Bosselman, R. A., Mendiaz, E. A., and Langley, K. E. (1990) *Cell* **63**, 195–201.
4. Huang, E., Nocka, K., Beier, D. R., Chu, T.-Y., Buck, J., Lahm, H.-W., Wellner, D., Leder, P., and Besmer, P. (1990) *Cell* **63**, 225–233.
5. Anderson, D. M., Lyman, S. D., Baird, A., Wignall, J. M., Eisenman, J., Rauch, C., March, C. J., Boswell, H. S., Gimpel, S. D., Cosman, D., and Williams, D. E. (1990) *Cell* **63**, 235–243.
6. Lyman, S. D., James, L., Bos, T. V., de Vries, P., Brasel, K., Gliniak, B., Hollingsworth, L. T., Picha, K. S., McKenna, H. J., Splett, R. R., Fletcher, F. A., Maraskovsky, E., Farrar, T., Foxworth, D., Williams, D. E., and Beckmann, M. P. (1993) *Cell* **75**, 1157–1167.
7. Satoh, M., Mioh, H., Shiotsu, Y., Ogawa, Y., and Tamaoki, T. (1997) *Exp. Hematol.* **25**, 972–979.
8. Middendorf, L. R., Bruce, J. C., Bruce, R. C., Eckles, R. D., Grone, D. L., Roemer, S. C., Sloniker, G. D., Steffens, D. L., Sutter, S. L., Brumbaugh, J. A., and Patonay, G. (1992) *Electrophoresis* **13**, 487–494.
9. Karlin, S., and Altschul, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2264–2268.
10. Fuchs, R. (1994) *Comput. Appl. Biosci.* **10**, 171–178.
11. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
12. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4863–4890.
13. Claros, M. G., and von Heijne, G. (1994) *Comput. Appl. Biosci.* **10**, 685–686.
14. Heng, H. H. Q., Squire, J., and Tsui, L.-C. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9509–9513.
15. Heng, H. H. Q., and Tsui, L.-C. (1993) *Chromosoma* **102**, 325–332.
16. Drickamer, K. (1993) *Annu. Rev. Cell Biol.* **9**, 237–264.
17. Wewer, U. M., and Albrechtsen, R. (1992) *Lab. Invest.* **67**, 253–262.
18. Serensen, C. B., Berglund, L., and Peterson, T. E. (1995) *Gene* **152**, 243–245.
19. Neame, P. J., Young, C. N., and Treep, J. T. (1992) *Protein Sci.* **1**, 161–168.
20. Braud, V. M., Allan, D. S. J., O'Callaghan, C. A., Söderström, K., D'Andrea, A., Ogg, G. S., Lazetic, S., Young, N. T., Bell, J. I., Phillips, J. H., Lanier, L. L., and McMichael, A. J. (1998) *Nature* **391**, 795–799.
21. Aspberg, A., Miura, R., Bourdoulous, S., Shimonaka, M., Heinegard, D., Schachner, M., Ruoslahti, E., and Yamaguchi, Y. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10116–10121.
22. Lyman, S. D., James, L., Escobar, S., Downey, H., de Vries, P., Brasel, K., Stocking, K., Beckmann, M. P., Copeland, N. G.,

- Cleveland, L. S., Jenkins, N. A., Belmont, J. W., and Davison, B. L. (1995) *Oncogene* **10**, 149–157.
23. Yang-Feng, T. L., Gibson, L., and Yang, Y. C. (1991) *Cytogenet. Cell Genet.* **58**, 20–27.
24. Durkin, M. E., Naylor, S. L., Albrechtsen, R., and Wewer, U. M. (1997) *Cytogenet. Cell Genet.* **76**, 39–40.
25. Grimaldi, J. C., and Meeker, T. C. (1989) *Blood* **73**, 2081–2085.
26. Benz-Lemoine, E., Brizard, A., Huret, J.-L., Babin, P., Guilhot, F., Couet, D., and Tanzer, J. (1988) *Blood* **72**, 1045–1047.
27. Boultonwood, J., Lewis, S., and Wainscoat, J. S. (1994) *Blood* **84**, 3253–3260.
28. Le Beau, M. M., Espinosa, R., III, Neuman, W. L., Stock, W., Roulston, D., Larson, R. A., Keinänen, M., and Westbrook, C. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5484–5488.
29. Gustavsson, P., Willing, T. N., van Haeringen, A., Tchernia, G., Dianzani, I., Donner, M., Elinder, G., Henter, J. I., Nilsson, P. G., Gordon, L., Skeppner, G., van't Veer-Korthof, L., Kreuger, A., and Dahl, N. (1997) *Nature Genet.* **16**, 368–371.
30. McKeithan, T. W., Rowley, J. D., Shows, T. B., and Diaz, M. O. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9257–9260.