Vol. 144, No. 1, 1987 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1987 Pages 74-80

PARTIAL PRIMARY STRUCTURES OF HUMAN AND MURINE MACROPHAGE COLONY STIMULATING FACTOR (CSF-1)

Albert Boosman*, James E. Strickler+, Kenneth J. Wilson++

Departments of Analytical Development and Protein Chemistry, Cetus Corporation, 1400 Fifty-Third Street, Emeryville, California 94608

and

E. Richard Stanley

Department of Microbiology and Immunology and Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461

Received February 25, 1987

SUMMARY: Approximately 40 amino-terminal residues and 20 internal residues of CSF-1 purified from the media of cultured human pancreatic carcinoma (MIA PaCa) cells and of cultured murine L cells have been identified. Results indicated that the two subunits in each molecule of biologically active CSF-1 are identical in their amino-terminal portions. The twelve amino-terminal residues of MIA PaCa CSF-1 were found to be identical to those of human-urinary CSF-1, suggesting that the polypeptide portions of the two human proteins may be identical. Approximately 75% of the amino acids identified in both MIA PaCa CSF-1 and murine CSF-1 were found to be common to both. No homology to other proteins was observed. This study suggests a subunit polypeptide Mr nearer to 17K than to 26K predicted from cDNA.

Colony stimulating factors (CSFs) regulate the proliferation and differentiation of cells of the granulocytic and mononuclear phagocytic lineages. CSF-1 selectively stimulates cell survival, proliferation and differentiation of mononuclear phagocytic cells via a specific cell surface receptor (reviewed in 1). It has been purified from serum-free

* To whom correspondence should be addressed; + Present address: Department of Molecular Genetics, Smith, Kline and French, Swedeland, Pennsylvania 19406; ++ Present address: Applied Biosystems, Inc., Foster City, California 94404

Abbreviations used: LCM CSF-1, macrophage colony stimulating factor isolated from serum-free culture medium conditioned by murine L cells; MIA PaCa CSF-1, CSF-1 isolated from serum-free culture medium conditioned by MIA PaCa cells; HU CSF-1, CSF-1 purified from human urine; cDNA, complementary desoxyribonucleic acid; PTH-amino acids, phenylthiohydantoin derivatives of amino acids; HPLC, high-performance liquid chromatography; CNBr, cyanogen bromide; OPA, ortho-phthalaldehyde.

0006-291X/87 $1.50

media conditioned by murine L cells (2) or by human pancreatic carcinoma cells (3) and from human urine (4,5). The Mr of LCM CSF-1 has been reported to be between 66K and 80K (6) whereas the Mr of the purified MIA PaCa CSF-1 used in this study and of HU CSF-1 is ∿46K (3-5). Treatment with reducing agents halves the apparent molecular weights of all CSF-1 preparations and destroys their biological activities (2-6), suggesting that the biologically active CSF-1 molecule consists of two subunits of similar size maintained in a dimeric state by disulfide bonds. Analysis of 66K LCM CSF-1 and 46K HU CSF-1 following treatment with endoglycosidases revealed that these proteins are glycosylated with asparagine-linked oligosaccharides of the "complex type" which, when removed, leave a polypeptide subunit of Mr ∿14.5K (6). Thus, the molecular weight differences between several forms of CSF-1 may be due solely to differences in glycosylation.

In this paper, we describe partial sequences of LCM CSF-1 and MIA PaCa CSF-1 and compare them with the sequence predicted from a MIA PaCa CSF-1 cDNA clone which encodes for a protein of Mr ∿26K (7).

MATERIALS AND METHODS

Protein isolation: LCM CSF-1 was isolated as previously described (2,8). Alternatively, batch calcium phosphate gel chromatography (8) followed by immunoaffinity chromatography on anti-murine CSF-1 monoclonal antibody columns (9) was used. MIA PaCa CSF-1 was purified as described earlier (3). All preparations had specific activities of 4×10^7 to 8×10^7 units/mg protein. (A unit is defined by comparison of the activity of the sample preparation with the activity of a stable reference preparation to which an activity value has been arbitrarily assigned. One unit is equivalent to approximately 0.44 femtomoles of CSF-1 (9).)

Primary structure determination: Amino acid sequences were determined using either an Applied Biosystems Model 470A gas-phase sequencer or a Beckman System 890M spinning-cup sequencer. PTH-amino acids recovered from either instrument were identified using reversed-phase HPLC employing isocratic elution from an Altex Ultrasphere ODS(PTH) column. Some samples were reduced and (^{14}C)carboxymethylated as described by Hirs (10). Methionyl bonds were cleaved with CNBr as described by Gross and Witkop (11). Edman degredation of selected peptides was blocked by a modification of the method of Bhown et al. (12) in which OPA in the presence of 2-mercaptoethanol was substituted for fluorescamine.

Amino acid composition determination: Amino acid compositions were determined using a Beckman 6300 amino acid analyzer following acid hydrolysis of samples for 24 h at 108 degrees.

RESULTS

A preliminary sequence analysis, using the gas-phase instrument, of ∿100 pmol of LCM CSF-1 containing both 66K and 80K species, allowed the identification of the first

Table 1. Amino acid sequences of purified MIA PaCa CSF-1 and L cell CSF-1 aligned with the amino-acid sequence deduced from the nucleotide sequence of MIA PaCa cDNA (7)

```
                                   5                  10                  15
cDNA:          Glu-Glu-Val-Ser-Glu-Tyr-Cys-Ser-His-Met-Ile-Gly-Ser-Gly-His-
MiaPaCa:       Glu-Glu-Val-Ser-Glu-Tyr- X -Ser-His-Met-Ile-Gly-Ser-Gly-His-
L cell:        Lys-Glu-Val-Ser-Glu-His-Cys-Ser-His-Met-Ile-Gly-Asn-Gly-His-

          20                  25                  30                  35
Leu-Gln-Ser-Leu-Gln-Arg-Leu-Ile-Asp-Ser-Gln-Met-Glu-Thr-Ser-Cys-Gln-Ile-Thr-Phe-
Leu-Gln-Ser-Leu-Gln-Arg-Leu-Ile-Asp-Ser-Gln-Met-Glu-Thr-Ser- X -Gln-Ile-Thr-Phe-
Leu-Lys-Val-Leu-Gln-Gln-Leu-Ile-Asp-Ser-Gln-Met-Glu-Thr-Ser- X -Gln-Ile-Ala-Phe-

          40                  45                  50                  55
Glu-Phe-Val-Asp-Gln-Glu-Gln-Leu-Lys-Asp-Pro-Val-Cys-Tyr-Leu-Lys-Lys-Ala-Phe-Leu-
Glu-Phe-Val-Asp-Gln-Glu-Gln-Leu...
Glu- X- Val...

          60                  65                  70                  75
Leu-Val-Gln-Tyr-Ile-Met-Glu-Asp-Thr-Met-Arg-Phe-Arg-Asp-Asn-Thr-Pro-Asn-Ala-Ile-
                                (Met-Arg)Phe-Arg-Asp-Asn-Thr-Pro-Asn-Ala-Ile-
                                (Met-Arg)Phe-Lys-Asp-Asn-Thr-Pro-Asn-Ala- X -

          80                  85                  90                  95
Ala-Ile-Val-Gln-Leu-Gln-Glu-Leu-Ser-Leu-Arg-Leu-Lys-Ser-Cys-Phe-Thr-Lys-Asp-Tyr-
Ala- X -Val-Gln-Leu-Gln-Glu- X -Ser...
Ala(Thr)Glu(Arg)Leu-Gln-Glu- X -Ser...

         100                 105                 110                 115
Glu-Glu-His-Asp-Lys-Ala-Cys-Val-Arg-Thr-Phe-Tyr-Glu-Thr-Pro-Leu-Gln-Leu-Leu-Glu-

         120                 125                 130                 135
Lys-Val-Lys-Asn-Val-Phe-Asn-Glu-Thr-Lys-Asn-Leu-Leu-Asp-Lys-Asp-Trp-Asn-Ile-Phe-

         140                 145                 150                 155
Ser-Lys-Asn-Cys-Asn-Asn-Ser-Phe-Ala-Glu-Cys-Ser-Ser-Gln-Gly-His-Glu-Arg-Gln-Ser-
```

```
                160                 165                 170                 175
Glu-Gly-Ser-Ser-Ser-Pro-Gln-Leu-Gln-Glu-Ser-Val-Phe-His-Leu-Leu-Val-Pro-Ser-Val-

                180                 185                 190                 195
Ile-Leu-Val-Leu-Leu-Ala-Val-Gly-Gly-Leu-Leu-Phe-Tyr-Arg-Trp-Arg-Arg-Arg-Ser-His-

                200                 205                 210                 215
Gln-Glu-Pro-Gln-Arg-Ala-Asp-Ser-Pro-Leu-Glu-Gln-Pro-Glu-Gly-Ser-Pro-Leu-Thr-Gln-

                220
Asp-Asp-Arg-Gln-Val-Glu-Leu-Pro-Val
```

13 amino acids, except for residue seven (Table 1). Two similar analyses (not shown) of reduced and (^{14}C)carboxymethylated LCM CSF-1 revealed the seventh residue to be cysteine. Since analysis using the gas-phase instrument failed to yield information beyond the thirteenth residue, the spinning-cup sequencer was used to examine a larger (~1 nmol) sample of LCM CSF-1. This experiment extended the known sequence to the 28th residue and defined the locations of two of approximately four methionyl residues known as a result of compositional analysis (not shown) to be present in this molecule. Treatment of LCM CSF-1 with CNBr left no uncleaved material, and in both the presence and absence of 2-mercaptoethanol yielded a fragment only ~5K smaller than the untreated subunits, indicating asymmetric distribution of methionyl residues and suggesting that these large fragments are not linked by interchain disulfide bonds in the intact dimer. Attempts to separate the CNBr fragments using size-exlcusion HPLC, reversed-phase HPLC, or preparative SDS-PAGE were only partially successful. Sequence analyses using the gas-phase instrument revealed that all fractions contained varying amounts of the same three major fragments. Because the amounts varied, the PTH-amino acids recovered from each sequencer cycle in analyses of several fractions could be tentatively assigned to three sequences: The first corresponding to the amino terminus; the second beginning at position 28 (overlapping the previously-known amino-terminal sequence and extending the known sequence to position 39); and a third, more carboxyl-terminal sequence which included a prolyl residue at position seven. The last

Table 2. Experimentally determined amino acid composition of purified MIA PaCa CSF-1 compared to the theoretical composition of the amino-terminal 153 residues of the sequence predicted from MIA PaCa CSF-1 cDNA and to the theoretical composition of the entire 224 residue hypothetical protein

Amino Acid	Experimental	Theoretical	
		153 Residues	224 Residues
Asp+Asn	13.18	11.84	9.46
Thr	5.99	5.26	4.05
Ser	6.51	7.89	9.46
Glu+Gln	16.97	17.11	18.02
Pro	4.77	1.97	4.50
Gly	2.25	1.97	3.15
Ala	4.10	3.29	3.15
Cys	2.14	4.61	3.15
Val	6.90	5.26	6.76
Met	3.41	2.63	1.80
Ile	3.63	4.61	3.60
Leu	9.56	10.53	12.16
Tyr	3.12	3.29	2.70
Phe	3.35	5.92	4.95
His	2.19	2.63	2.70
Lys	8.02	7.24	4.95
Trp	N.D.*	--	--
Arg	3.90	3.95	5.41
Totals	99.99	100.00	99.97

Results are expressed in mole-percent.

* N.D. = Not determined and not included in calculations.

sequence was extended in a subsequent analysis in which contaminating sequences were blocked by OPA treatment after six cycles of Edman degradation.

In the case of MIA PaCa CSF-1, ~ 1 nmol of protein was cleaved with CNBr and the unfractionated peptide mixture was analyzed using the spinning-cup instrument. On the basis of homology, amino acids were assigned to three segments corresponding to the major CNBr fragments of LCM CSF-1. The amino-acid composition of MIA PaCa CSF-1 was determined and is compared in Table 2 to the theoretical composition of the entire 224 residue protein predicted from the corresponding cDNA and to the theoretical composition of the 153 amino-terminal residues of that hypothetical protein.

DISCUSSION

In six separate sequence analyses of four different preparations of CSF-1, only a single sequence was observed, indicating that all non-blocked polypeptides in the sample had identical amino termini. Recoveries of amino-terminal residues accounted for 50-70% of the applied samples, precluding the existence of an occult subunit; i.e., one of similar size possessing a blocked amino terminus. Such a subunit would have caused recoveries of less than 40% and would probably have become evident during sequence analysis of unfractionated CNBr-treated MIA PaCa CSF-1. Furthermore, the observed expression of CSF-1 biological activity by primate (Cos-7) cells transformed by MIA PaCa CSF-1 cDNA (7) could not have occurred if a second type of subunit were required for activity.

The LCM CSF-1 amino-terminal sequence shown in Table 1 differs from the partial sequence of this protein originally reported by Ben-Avram et al. (13), in which residues eight and nine, as well as residue seven, were identified as cysteines. These authors have since altered their results (14) to be consistent with the sequence reported here. The observation that 75% of the amino acids identified in both LCM CSF-1 and MIA PaCa CSF-1 are identical, together with their similar subunit structures and biological activities, illustrate the close relationship between these two proteins. That the similarity extends even farther can be inferred by comparing the present work to the results of Ben-Avram et al. (13). Trypsin treatment of LCM CSF-1 by these investigators liberated two peptides containing carbohydrate, one of which also contained cysteine. Precisely the same result would be predicted if MIA PaCa CSF-1 of the sequence shown in Table 1 were so treated, if it were glycosylated at both potential N-linked sites (positions 122 and 140). A comparison of the CSF-1 sequences in Table 1 to the data banks of Dayhoff (15) and Doolittle (16) failed to reveal relationships to any proteins of known primary structures.

Hydropathic analysis of cDNA-predicted amino-acid sequences of several growth factors suggest that they may exist in vivo on the surface of the cells of their origin, anchored by a hydrophobic transmembrane segment from which they are later cleaved by proteolysis. The 224 residue hypothetical MIA PaCa CSF-1 molecule predicted from the cDNA sequence includes a segment (positions 167 to 188) consisting almost entirely of

hydrophobic residues and may be an example of such a membrane-bound precursor. One or more proteolytic events could release a CSF-1 molecule of fewer than 167 residues. Although the length of the active MIA PaCa CSF-1 molecule analyzed in this study remains undefined, the experimentally determined composition suggests that it is nearer to 153 residues than to 224 (Table 2). Consistent with this hypothesis are the observations that L cells exhibit cell surface immunofluorescence following incubation with fluorescently-labeled F(ab')2 fragments of anti-CSF-1 monoclonal antibodies (K. Price and E. R. Stanley, unpublished observations) and that treatment of L cells with trypsin causes release of biologically active CSF-1 (17,18).

ACKNOWLEDGMENTS

This work was supported in part by NIH grant CA32551 and the AECOM Cancer Center Core grant (to ERS). ERS is the recipient of an Irma T. Hirschl Career Scientist Award. We thank Judit Csejtey, Jeremiah Desmond, Fran Pellerito and Claudia Saez for technical assistance.

REFERENCES

1. Stanley, E. R., Guilbert, L. J., Tushinski, R. J., and Bartelmez, S. H. (1983) **J. Cell. Biochem. 21,** 151-159.
2. Stanley, E. R., and Heard, P. M., (1977) **J. Biol. Chem. 252,** 4305-4312.
3. Csejtey, J., and Boosman, A. (1986) **Biochem. Biophys. Res. Comm. 138,** 238-245.
4. Das, S. K., Stanley, E. R., Guilbert, L. J., and Forman, L. W. (1981) **Blood 58,** 630-641.
5. Wong, F. F., and Goldwasser, E. (1983) **J. Cell Biochem. 21,** 263-276.
6. Das, S. K., and Stanley, E. R. (1982) **J. Biol. Chem. 257,** 13679-13684.
7. Kawasaki, E. S., Ladner, M. B., Wang, A. M., Van Arsdell, J., Warren, M. K., Coyne, M. Y., Schweikert, V. L., Lee, M.-T., Wilson, K. J., Boosman, A., Stanley, E. R., Ralph, P., and Mark, D. F. (1985) **Science 230,** 291-296.
8. Stanley, E. R., and Guilbert, L. J. (1981) **J. Immmunol. Methods 42,** 253-284.
9. Stanley, E. R. (1985) **Methods Enzymol. 116,** 564-587.
10. Hirs, C. H. W. (1967) **Methods Enzymol. 11,** 199-203.
11. Gross, E., and Witkop, B. (1961) **J. Biol. Chem. 237,** 1856-1860.
12. Bhown, A. S., Bennett, J. C., Morgan, P. H., and Mole, J. E. (1981) **Analyt. Biochem. 112,** 158-162.
13. Ben-Avram, C. M., Shively, J. E., Shadduck, R. K., Waheed, A., Rajavashisth, T., and Lusis, A. J. (1985) **Proc. Natl. Acad. Sci. USA 82,** 4486-4489.
14. Ben-Avram, C. M., Shively, J. E., Shadduck, R. K., Waheed, A., Rajavashisth, T., and Lusis, A. J. (1985) **Proc. Natl. Acad. Sci. USA 82,** 7801.
15. Dayhoff, M. O. (1972) **The Atlas of Protein Sequence and Structure.** National Biomedical Research Foundation, Georgetown University, Washington, D. C.
16. Doolittle, R. F. (1981) **Science 214,** 149-159.
17. Cifone, M., and Defendi, V. (1974) **Nature 252,** 151-153.
18. Stanley, E. R., Cifone, M., Heard, P. M., and Defendi, V. (1976) **J. Exp. Med. 143,** 631-647.