

CLONING AND CHARACTERIZATION OF A cDNA FOR A
NEW MOUSE T CELL GROWTH FACTOR (P40)

JACQUES VAN SNICK,* ALINE GOETHALS,* JEAN-CHRISTOPHE RENAULD,*
EMIEL VAN ROOST,* CATHERINE UYTENHOVE,* MICHAEL R. RUBIRA,†
ROBERT L. MORITZ,‡ AND RICHARD J. SIMPSON‡

*From the *Ludwig Institute for Cancer Research, Brussels Branch, B-1200, Brussels, Belgium; and
†the Joint Protein Structure Laboratory, Ludwig Institute for Cancer Research, Melbourne Branch,
and The Walter and Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital,
Parkville, 3050 Victoria, Australia*

Helper T cell lines are readily initiated from lymph nodes of antigen-primed mice by culture in the presence of antigen and irradiated syngeneic APCs (1). Supernatants collected from these cell lines after stimulation with Con A or phorbol esters contain a variety of factors including IL-2 through IL-6, granulocyte/macrophage colony-stimulating factor (GM-CSF) and IFN- γ (2, 3). Recently, we observed that some of these supernatants supported the growth of certain helper T cell clones in the absence of antigen or APCs. This activity could not be inhibited by antibodies blocking the action of IL-2 or IL-4, and co-purified with a 32-39-kD single-chain glycoprotein, which was designated P40 (4). The NH₂-terminal residue of the protein appeared to be blocked but internal amino acid sequence data derived from peptides obtained after cyanogen bromide or enzymatic cleavage of the molecule suggested that it had no obvious sequence homology with known proteins. The cloning and sequencing of a cDNA for P40, which is described here, unequivocally identifies this molecule as a new entity in the family of factors capable of stimulating T cell growth.

Materials and Methods

Protein Sequencing. Native P40 (200 pmol) was purified as described (4). Amino acid sequence information was obtained following in situ cyanogen bromide cleavage of the molecule (8 μ g, 200 pmol) on the sample disk of the sequencer (5) and from peptides generated by treatment of 10 μ g of reduced and *S*-carboxymethylated P40 (250 pmol) with *L*-(1-tosyl-amino-2-phenyl) ethyl chloromethyl ketone-treated trypsin.

Tryptic peptides were purified by microbore column reversed-phase HPLC using procedures described elsewhere (6). Automated Edman degradation of peptides was performed using a sequencer (No. 477A; Applied Biosystems, Inc., Foster City, CA) equipped with an on-line phenylthiohydantoin amino acid analyzer (No. 120A).

Screening of cDNA Library. Double-stranded cDNA was prepared according to Gubler and Hoffman (7) using polyadenylated RNA isolated from P40-producing helper T cells TUC7.51 after a 24-h stimulation with Con A (2.5 μ g/ml). The cDNA was cloned into the Bam HI site of a pUC8 vector and transformed into *Escherichia coli* strain DH5 α . Transformants were screened by in situ hybridization with two end-labeled 20-mer oligonucleotide probes. For the initial screening, we used a 64-fold degenerate probe [5'TGCAT(C+T)TC(X)GT(C+T)-TT(C+T)TG(G+A)AA3'] corresponding to amino acid sequence FQKTEMQ (positions 114-120, see Fig. 3). Positive clones were subsequently tested with a 128-fold degenerate probe

[5'GG(A+G)TC(A+G)TC(T+C)TT(X)AG(A+G)TT(C+T)TC3'] corresponding to sequence ENLKDDP (positions 17–23, see Fig. 3).

DNA Sequencing. DNA was sequenced by the dideoxynucleotide procedure (8) after subcloning into an M13 vector (9). Appropriate fragments were generated by digestion with Pst I and Nco I restriction endonucleases (see Fig. 3).

P40 Growth Factor Assay. P40 activity was tested using P40-dependent TS1 line as previously described (4). Briefly, 2,000 TS1 cells were incubated for 3 d in the presence of serial sample dilutions and cell growth was assessed by colorimetric determination of hexosaminidase levels (10).

Data Bases. Data bases searched for sequences with homology to P40 included: nucleic acid data bases Genbank (release June 1988) and EMBL (release May 1988), and protein data bases NBRF (release June 1988), GBtrans Protein Data Base (release 3.0 August 1988), and Swiss-Prot (release April 1988).

Results and Discussion

A cDNA library was prepared, in a pUC8 vector, from a helper T cell clone that produces large amounts of P40 after stimulation with Con A. This library was screened with two oligonucleotide probes synthesized on the basis of selected amino acid sequence data obtained by analysis of P40 peptides. Of 20,000 independent transformants, 112 hybridized with the two probes. Most of these clones contained cDNA inserts of ~500 bp. Using one of these cDNAs as a probe, a strong signal was obtained with a transcript of ~700 nucleotides in Northern blots of poly(A)⁺ RNA isolated from P40-producing helper T cell clone TUC7.51. Poly(A)⁺ RNA from P815 mastocytoma, which does not produce detectable P40 activity, gave no signal at all (Fig. 1).

To prove that the selected clones contained authentic P40 cDNA, we tried to express one of them in fibroblasts. Insert P40.2B4 was cloned into the Bam HI site of plasmid pZIPneoSV(X)1 (11) and transfected into Clone-1d fibroblasts (12). Cell supernatants collected 48 h after transfection were tested for their growth factor activity on P40-dependent TS1 cells. As shown in Fig. 2, supernatants from cells transfected with P40.2B4 cDNA, but not from mock-transfected cells, supported the growth of TS1. This result also indicated that P40.2B4 cDNA presumably contains the entire coding region of P40.

The complete nucleotide sequence of the cDNA insert of clone P40.2B4 was determined. As shown in Fig. 3, it consists of 554 nucleotides with a 5' untranslated sequence of 15 nucleotides, an open reading frame of 432 nucleotides, and a 3' untranslated region of 107 nucleotides. The 3' end terminates with a string of 18 adenine residues located 12 nucleotides downstream from an AATAAA polyadenylation signal consensus sequence (13). The 3' untranslated region contains 3 copies of the sequence ATTTA which is characteristic of transiently expressed genes such as GM-CSF, G-CSF, interferons, several interleukins, TNF, and oncogenes *c-fos* and *c-myc* (14); two of these repeats at nucleotide positions 461–468 and 470–477 are part of an eight-nucleotide motif TATTTATT, defined by Caput et al. (15), which is also present in many of these molecules.

The predicted polypeptide encoded by the cDNA insert of clone P40.2B4 consists of 144 residues. This size estimation is based on the presumption that the first ATG in the sequence (nucleotide position 16–18) is the initiator codon, a view supported by the efficient expression of the cDNA in fibroblasts and by the presence of an adenine at nucleotide position 13, in concordance with the consensus sequence defined

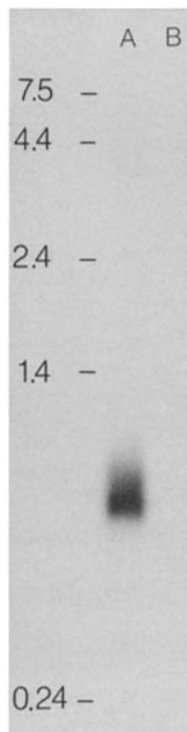


FIGURE 1. Northern blot analysis of P40 mRNA. Poly(A)⁺ RNA (5 μ g) from Con A-stimulated TUC7.51 helper T cells (A) and from P815 mastocytoma (B) was hybridized with a ³²P-labeled P40 cDNA insert. Size of RNA standards is indicated in kilobases.

by Kozak for the initiator ATG codon (16); an in-frame TGA translation termination codon occurs at nucleotides 448–450. The deduced P40 sequence is characterized by a hydrophobic NH₂-terminal sequence typical of a signal peptide and contains all amino acid sequences known from the analysis of P40 peptides (see underlined segments in Fig.3). Because of the presence of a blocked NH₂ terminus in the native protein, there is some uncertainty concerning the NH₂-terminal residue. Based on the probability weight-matrix described by von Heijne (17), the most likely NH₂-terminal sequence of the mature protein would be Gln-Arg-Cys. This would be con-

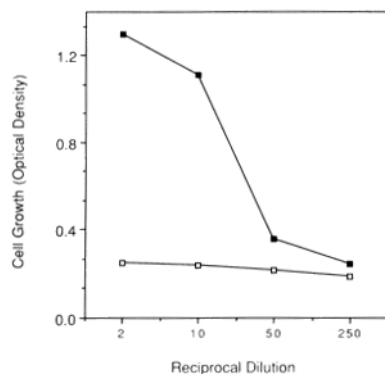


FIGURE 2. Expression of P40 cDNA in fibroblasts. P40-dependent cell line TS1 was cultured for 3 d in the presence of medium conditioned by fibroblasts transfected with P40.2B4 cDNA inserted in expression vector pZIPneoSV(X)1 (closed symbols). Open symbols give the results obtained with mock-transfected cells.

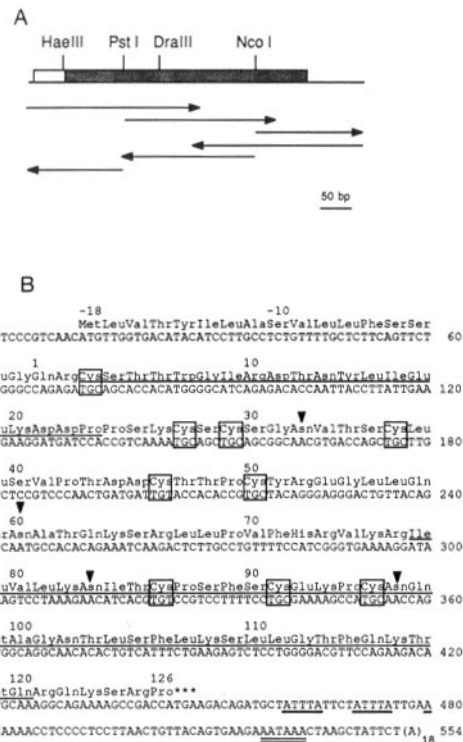


FIGURE 3. (A) Schematic representation of P40.2B4 clone and of sequence strategy. Open box corresponds to the leader sequence and shaded box to the coding region of P40. Arrows indicate the direction and extent of sequencing. (B) Nucleotide sequence and deduced amino acid sequence of clone P40.2B4. Numbers above the sequence refer to amino acid positions and those on the right to nucleotide positions. Arrows indicate potential *N*-glycosylation sites, cysteine residues are boxed, and amino acids matching those determined by protein sequencing are underlined. (Double underline) Presumptive polyadenylation signal; (heavy underline) sequences potentially involved in mRNA degradation.

sistent with preliminary evidence obtained by biochemical analysis of P40 peptides. Mature P40 would then consist of 126 amino acids with a predicted relative molecular mass of 14,150. The difference with the M_r measured for native P40 appears to be due to glycosylation as suggested by the presence of four potential *N*-linked glycosylation sites and confirmed by the $\sim 15 \times 10^3 M_r$ of the native protein after *N*-glycanase treatment (Simpson, R. J., unpublished). The sequence of P40 is further characterized by the presence of 10 cysteines and a strong predominance of cationic residues, which explains the elevated pI (~ 10) of the native protein (4).

Using the FASTN program (18) (κ -tuple 1-6), we found no significant homology between P40 cDNA and recorded nucleic acid sequences. A search of several protein databanks (listed under Materials and Methods) was equally unsuccessful in detecting a strong homology with known proteins.

Previous experiments based on the use of cell lines have demonstrated a strict specificity of P40 for helper T cells, in sharp contrast with the much broader spectrum of activities of the two other known T cell growth factors, IL-2 and IL-4. In particular, purified native P40 completely failed to stimulate cytolytic T cell clones as well as a variety of factor-dependent cell lines of myeloid and B cell origin (4). The availability of recombinant P40 should now make it possible to examine whether this specificity holds true with freshly isolated cells, which, unlike cell lines, have not gone through the selection associated with long term in vitro cultures. Finally, it must be mentioned that a growth factor activity with a specificity similar to that of P40 was recently described by Ogata et al. (19) in the supernatant of a thymic

stroma cell line. The structural relation between this factor and P40 is presently not known.

Summary

Recently, we described a murine helper T cell-derived molecule with T cell growth factor activity that is functionally and structurally distinct from IL-2, IL-4, and other known growth factors (4). This molecule, designated P40, was identified as a glycoprotein capable of supporting antigen-independent growth of certain helper T cell clones. Here, we report the cloning and expression of a cDNA for this new growth factor. The predicted mature protein is a cationic cysteine-rich polypeptide of 14 kD without significant homology to previously sequenced proteins.

Received for publication 14 September 1988.

References

1. Corradin, G., H. M. Etlinger, and J. M. Chiller. 1977. Lymphocyte specificity to protein antigen. I. Characterization of the antigen-induced in vitro T cell-dependent proliferative response with lymph node cells from primed mice. *J. Immunol.* 119:1048.
2. Cherwinski, H. M., J. H. Schumacher, K. D. Brown, and T. R. Mosmann. 1987. Two types of mouse helper T cell clone. III. Further differences in lymphokine synthesis between Th1 and Th2 clones revealed by RNA hybridization, functionally monospecific bioassays, and monoclonal antibodies. *J. Exp. Med.* 166:1229.
3. Van Snick, J., S. Cayphas, A. Vink, C. Uyttenhove, P. G. Coulie, M. R. Rubira, and R. J. Simpson. 1986. Purification and NH₂-terminal amino acid sequence of a new T-cell derived lymphokine with growth factor activity for B-cell hybridomas. *Proc. Natl. Acad. Sci. USA.* 83:9679.
4. Uyttenhove, C., R. Simpson, and J. Van Snick. 1988. Functional and structural characterization of P40, a mouse glycoprotein with T cell growth factor activity. *Proc. Natl. Acad. Sci. USA.* 85:6934.
5. Simpson, R. J., and E. C. Nice. 1984. In situ cyanogen bromide cleavage of N-terminally blocked proteins in a gas-phase sequencer. *Biochem. Intl.* 8:787.
6. Simpson, R. J., R. L. Moritz, M. R. Rubira, and J. Van Snick. 1988. Murine hybridoma/plasmacytoma growth factor: complete amino acid sequence and relation to human interleukin-6. *Eur. J. Biochem.* In press.
7. Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene (Amst.)* 25:263.
8. Sanger, F. 1981. Determination of nucleotide sequences in DNA. *Science (Wash. DC)* 214:1205.
9. Kieny, M. P., R. Lathe, and J. P. Lecoq. 1983. New versatile cloning and sequencing vectors based on bacteriophage M13. *Gene (Amst.)* 26:91.
10. Landegren, U. 1984. Measurement of cell numbers by means of the endogenous enzyme hexosaminidase. Applications to detection of lymphokines and cell surface antigens. *J. Immunol. Methods.* 67:379.
11. Cepko, C. L., B. E. Roberts, and R. C. Mulligan. 1984. Construction and applications of a highly transmissible murine retrovirus shuttle vector. *Cell.* 37:1053.
12. Kit, S., D. Dubbs, L. Pickarski, and T. Hsu. 1963. Deletion of thymidine kinase activity from L cells resistant to bromodeoxyuridine. *Exp. Cell. Res.* 31:297.
13. Birnstiel, M. L., M. Busslinger, and K. Strub. 1985. Transcription termination and 3' processing: the end is in site! *Cell.* 41:349.

14. Shaw, G., and R. Kamen. 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*. 46:659.
15. Caput, D., B. Beutler, K. Hartog, R. Thayer, S. Brown-Shimer, and A. Cerami. 1986. Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc. Natl. Acad. Sci. USA*. 83:1670.
16. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15:8125.
17. Von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683.
18. Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science (Wash. DC)*. 227:1435.
19. Ogata, M., S. Sato, H. Sano, T. Hamaoka, H. Doi, K. Nakanishi, Y. Asano, T. Itoh, and H. Fujiwara. 1987. Thymic stroma-derived T cell growth factor (TSTGF): functional distinction of TSTGF from interleukins 2 and 4 and its preferential growth promoting effect on helper T cell clones. *J. Immunol.* 139:2675.