11-14-1984

T-cell Growth Factor: Complete Nucleotide Sequence and Organization of the Gene in Normal and Malignant Cells

Nikki Holbrook
National Cancer Institute

Kendall Smith
Dartmouth College

Albert Fornace
National Cancer Institute

Claudette Comeau
National Cancer Institute

Robert L. Wiskocil
University of California, San Francisco

See next page for additional authors

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1634–1638, March 1984
Biochemistry

T-cell growth factor: Complete nucleotide sequence and organization of the gene in normal and malignant cells

(gene structure/human DNA library screening/lymphokines/interleukin 2)

NIKKI J. HOLBROOK*, KENDALL A. SMITH†, ALBERT J. FORNACE, JR.*, CLAUDETTE M. COMEAU*, ROBERT L. WISKOCIL‡, AND GERALD R. CRABTREE*

*Hematopathology Section, Building 10, Room 2N-113, Laboratory of Pathology, National Cancer Institute, Bethesda, MD 20205; †Immunology Program, Dartmouth Medical School, Norris Cotton Cancer Center, Hanover, NH 03756; and ‡Howard Hughes Medical Institute, University of California, San Francisco, CA 94143

Communicated by David W. Talmage, November 14, 1983

ABSTRACT **Using a cloned cDNA copy of T-cell growth factor (TCGF) mRNA from the Jurkat leukemic T-cell line, we have isolated three overlapping *TCGF* genomic clones from a human DNA library. The entire *TCGF* gene is contained within two adjacent *Eco*RI fragments spanning about 8 kilobases. The complete nucleic acid sequence was determined. The gene is divided into four exons. The 5′ untranslated region and the first 49 amino acids of the protein, 20 of which constitute a signal polypeptide and are not present in the secreted protein, are encoded by the first exon. Exons 2 and 3, separated from each other by a long intervening sequence, contain coding information for the next 20 and 48 amino acids, respectively. The remaining 36 amino acids and the 3′ untranslated region are contained in the fourth exon. A promoter sequence T-A-T-A-A-A is present 77 base pairs (bp) upstream from the translation initiation site, and a CAT homology region occurs 104 bp upstream from the initiation site. A putative site for initiation of mRNA transcription was identified 53 bp 5′ of the translation initiation codon. The organization of the gene was shown by Southern blot analysis to be identical in normal peripheral blood lymphocytes and in a variety of malignant lymphoid cell types. Restriction analysis of these cellular DNAs produced results exactly as predicted by the map for the cloned genomic *TCGF*, indicating that there is only a single copy of the human *TCGF* gene.**

T-cell growth factor (TCGF), also known as interleukin 2 (IL-2), is a protein produced by T lymphocytes that is capable of initiating and maintaining long term *in vitro* growth of activated T cells (1). TCGF was first shown to be released into media from lectin-stimulated human peripheral blood and bone marrow T-cell cultures, but it has since been shown to be present in other mammalian systems (1, 2). Human TCGF has been purified to homogeneity and has a molecular weight of about 15,000 (3–5). Taniguchi *et al.* (6) recently reported the first cloning and nucleotide sequence of the cDNA coding for TCGF from the human Jurkat leukemic T-cell line. The cDNA codes for a protein of 153 amino acids, the first 20 of which appear to constitute a signal polypeptide and do not appear in the secreted protein (6).

Because it plays a key regulatory role in mediating T-cell proliferation, TCGF has been suggested as a possible treatment for immunodeficiency syndromes as well as a means of obtaining large numbers of tumor-specific cytotoxic cells for immunotherapy of malignancy. Mitogens such as phytohemagglutinin and concanavalin A induce TCGF production in human T lymphocytes (3, 7). Phorbol myristic acetate by itself does not induce TCGF production, but it greatly enhances the induction by mitogen (3, 7). An understanding of the processes involved in the mitogenic stimulation of TCGF activity would surely contribute to our overall understanding of T-cell differentiation. To investigate the molecular mechanisms through which mitogens induce TCGF we have cloned the gene from a human DNA library. We report the complete nucleotide sequence of the gene for TCGF and its organization in normal and malignant cells.

MATERIALS AND METHODS

Preparation and Screening of a cDNA Library for TCGF. Polyadenylylated TCGF RNA was prepared (8) from the human Jurkat leukemic T-cell line after stimulating the cells for 6 hr with phytohemagglutinin (1.5 μg/ml) and phorbol myristic acetate (50 ng/ml). The RNA was fractionated on a sucrose density gradient and TCGF mRNA activity was monitored by translation in a rabbit reticulocyte lysate and immunoprecipitation of TCGF using a monoclonal antibody to TCGF (9). Double-stranded cDNA was prepared and the cDNA was inserted into pBR322 at the *Pst* I site using the standard G-C tailing method (10). The library was screened for TCGF sequences after transformation into *Escherichia coli* HB101 using an 18-mer synthetic oligonucleotide probe (G-C-A-C-C-T-A-C-T-T-C-A-A-G-T-T-C-T) from positions 108–125 of the reported sequence of TCGF cDNA (6). Colonies grown overnight on nitrocellulose filters overlying LB agar containing tetracycline (20 μg/ml) were lysed as described by Thayer (11). Filters were hybridized overnight at 50°C with the probe end-labeled with [γ-^{32}P]ATP in 6× NET (1× NET is 0.15 NaCl/0.015 M Tris·HCl, pH 7.5/0.001 M EDTA), 5× Denhardt's solution (1× Denhardt's solution is 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone)/0.5% Nonidet P-40/10% (vol/vol) dextran sulfate/salmon sperm DNA (250 μg/ml). The filters were washed at 0°C with four changes each in 6× standard saline citrate (1× standard saline citrate is 0.15 M NaCl/0.015 M sodium citrate, pH 7.2) for a total of 20 min, then at 50°C for 1 min in 6× standard saline citrate.

Isolation and Mapping of the Gene. Approximately 1×10^6 plaques from three human genomic libraries were screened by the Benton–Davis procedure (12) using the ^{32}P-labeled cDNA probe for TCGF. DNA was prepared from plaque-purified clones according to Maniatis *et al.* (13). Restriction endonuclease digestions were carried out on the resulting samples according to the manufacturers' instructions and samples were analyzed on 0.8–1.2% agarose gels. The structural organization of the *TCGF* gene in normal peripheral blood lymphocytes and various malignant lymphoid cell types was examined by hybridization of cellular DNA to ^{32}P-labeled TCGF cDNA using the standard Southern blot procedure (14).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: TCGF, T-cell growth factor; bp, base pair(s); kb, kilobase(s).

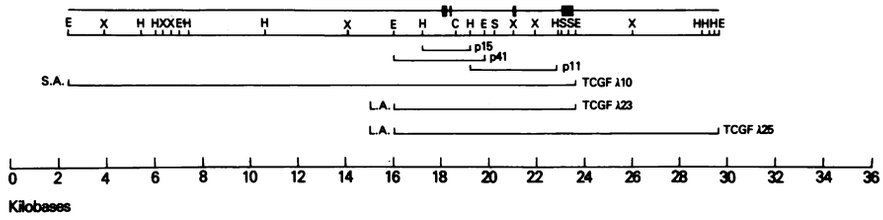

FIG. 1. Restriction map of human *TCGF*. TCGF λ10, λ23, and λ25 were isolated from a human DNA library in Charon 4A. The 2.1-kilobase (kb) *Hin*dIII fragment (p15), 3.7-kb *Eco*RI fragment (p41), and 3.6-kb *Hin*dIII fragment (p11) were subcloned into pBR322 for DNA sequence analysis. The locations of exons are indicated by black rectangles. E, *Eco*RI; H, *Hin*dIII; C, *Cla* I; S, *Stu* I; X, *Xba* I. The sequence is represented from 5′ to 3′ going left to right.

DNA Sequence Analysis. DNA nucleotide sequences were determined by the dideoxynucleotide termination method of Sanger *et al.* (15) after subcloning restriction endonuclease fragments into M13-mp8, mp9, mp10, and mp11 phage vectors. Sequences were analyzed and compared on an IBM system 370 computer using the program described by Queen and Korn (16).

mRNA Sequence Analysis and Identification of the Transcription Initiation Site. A 5′ ^{32}P-end-labeled 31-base-pair (bp) fragment from the TCGF cDNA beginning just after the translation initiation codon and extending 3′ was isolated by cleavage of the cDNA with *Dde* I and *Rsa* I. This fragment was used as a primer for mRNA sequence analysis following the procedure of Bina-Stein *et al.* (17).

Materials. Human spleen, placental, and peripheral blood lymphocyte genomic libraries were graciously provided by P. Leder, T. Maniatis, and E. Benz, respectively. DNA polymerase and DNase I were from Boehringer Mannheim; restriction endonucleases were from Bethesda Research Laboratories, New England BioLabs and Boehringer Mannheim; T_4 DNA ligase, M13 phage vectors, T_4-polynucleotide kinase, and unlabeled deoxynucleotide triphosphates and dideoxynucleotide triphosphates were from P-L Biochemicals; and [γ- and α-^{32}P]deoxynucleotide triphosphates were from Amersham.

RESULTS AND DISCUSSION

Isolation and Sequence of cDNA Clones for TCGF. Of >6000 cDNA clones tested for hybridization to the synthetic oligonucleotide probe based on the published TCGF cDNA sequence only 2 positives were detected. Neither of these clones, which were 420 and 450 bp long, contained the complete cDNA sequence. Nucleic acid sequence analysis showed them both to contain the 5′ portion of the message with sequence identical to that reported by Taniguchi *et al.* (6). The 420-bp clone started at the same place as the clone reported by Taniguchi and extended to position 420 of that clone. The 450-bp clone contained a few more residues 5′ to the 420-bp clone and extended farther 3′. The fact that both of these clones contained the 5′ end rather than 3′ portion of the cDNA, is likely due to selection of this region, because the synthetic probe we used was directed toward the 5′ end of the message. Rescreening of the 6000 clones with these partial cDNAs did not produce any other positive clones.

Isolation and Characterization of Genomic Clones. Approximately 10^6 plaques from three different human genomic libraries in bacteriophage Charon 4A were screened for *TCGF* using the cDNA probe. We were unsuccessful in isolating the gene from two of these libraries (spleen and placental). However, three overlapping clones containing the entire *TCGF* gene and 5′- and 3′-flanking regions were isolated from the peripheral blood lymphocyte library (Fig. 1). These are designated λ23, λ10, and λ25. Digestion of these clones with restriction endonucleases and hybridization with a ^{32}P-labeled cDNA probe showed the gene to be contained within two contiguous *Eco*RI fragments of ≈3700 bp each. Two *Hin*dIII sites were shown to occur in one of these, while the other contained a single *Hin*dIII site. These sites were used to subclone the regions containing *TCGF* into pBR322, generating the three overlapping subclones p15, p41, and p11. Further analysis of these subclones with a variety of restriction endonucleases and their known sites of cleavage in the cDNA allowed the proper orientation of the two *Eco*RI fragments and the construction of the map of the gene shown in Fig. 1.

Sequence and Structure of the Isolated *TCGF* Gene. A combination of directed and nondirected strategies was used to determine the sequence of the *TCGF* gene (Fig. 2). The entire sequence is shown in Fig. 3. The gene is 5040 bp long. The sequence includes 292 nucleotides upstream from the translation initiation site and extends 297 bp downstream from the polyadenylylation signal. Consistent with other genes, a promoter sequence T-A-T-A-A-A occurs 77 bp upstream from the translation initiation site. A CAT homology region, also implicated in the regulation of other genes, occurs 104 nucleotides upstream from the translation initiation site. To determine the site at which transcription is initiated, we analyzed the sequence of TCGF mRNA by extension of a 31-bp primer using reverse transcriptase. The primer, complementary to the beginning of the coding region just after the translation initiation site, was annealed to total Jurkat poly(A) RNA, and extended by incorporation of deoxy- and dideoxynucleotides. Analysis of the products indicates that

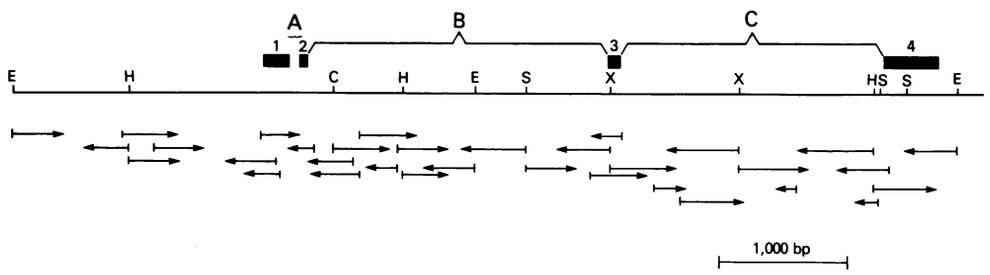

FIG. 2. Sequence analysis strategy for the human *TCGF* gene. The locations of the exons are indicated by black rectangles. Symbols for the endonuclease restriction sites are the same as for Fig. 1. The introns are labeled A, B, and C. Arrows indicate the direction and length of the clones used for sequence analysis.

```
        10         20         30         40         50         60         70         80         90        100        110
CGAATTCCCC TATCACCTAA GTGTGGGCTA ATGTAACAAA GAGGGATTTC ACCTACATCC ATTCAGTCAG TCTTTGGGGG TTTAAAGAAA TTCCAAAGAG TCATCAGAAG

       120        130        140        150        160        170        180        190        200        210
AGGAAAAATG AAGGTAATGT TTTTTCAGAC TGGTAAAGTC TTTGAAAATA TGTGTAATAT GTAAAACATT TTGACACCCC CATAATATTT TTCCAGAATT

       220        230        240        250        260        270        280        290                   307
AACAGTATAA ATTGCATCTC TTGTTCAAGA GTTCCCTATC ACTCTTTAAT CACTACTCAC AGTAACCTCA ACTCCTGCCA CA ATG TAC AGG ATG CAA CTC
                                                                                            MET Tyr Arg Met Gln Leu

           322                337                352                367                382                397
CTG TCT TGC ATT GCA CTA AGT CTT GCA CTT GTC ACA AAC AGT GCA CCT ACT TCA AGT TCT ACA AAG AAA ACA CAG CTA CAA CTG GAG
Leu Ser Cys Ile Ala Leu Ser Leu Ala Leu Val Thr Asn Ser Ala Pro Thr Ser Ser Ser Thr Lys Lys Thr Gln Leu Gln Leu Glu

               412                427                449        459        469        479        489        499
CAT TTA CTG CTG GAT TTA CAG ATG ATT TTG AAT GGA ATT AAT GTAAGTATAT TTCCTTTCTT ACTAAAATTA TTACATTTAG TAATCTAGCT GGAGATCATT
His Leu Leu Leu Asp Leu Gln Met Ile Leu Asn Gly Ile Asn

       509        519        529                544                559                574                589
TCTTAATAAC  AATGCATTAT ACTTTCTTAG AAT TAC AAG AAT CCC AAA CTC ACC AGG ATG CTC ACA TTT AAG TTT TAC ATG CCC AAG AAG
                                  Asn Tyr Lys Asn Pro Lys Leu Thr Arg Met Leu Thr Phe Lys Phe Tyr Met Pro Lys Lys

       599        609        619        629        639        649        659        669        679        689        699
GTAAGTACAA TATTTTATGT TCAATTTCTG TTTTAATAAA ATTCAAAGTA ATATGAAAAT TTGCACAGAT GGGACTAATA GCAGCTCATC TGAGGTAAAG AGTAACTTTA

       709        719        729        739        749        759        769        779        789        799        809
ATTTGTTTTT TTGAAAACCC AAGTTTGATA ATGAAGCCTC TATTAAAACA GTTTTACCTA TATTTTTAAT ATATATTTGT  GTGTTGGTGG GGGTGGGAGA AAACATAAAA

       819        829        839        849        859        869        879        889        899        909        919
ATAATATTCT CTCACTTTAT CGATAAGACA ATTCTAAACA AAAATGTTCA TTTATGGTTT CATTTAAAAA TGTAAAACTC TAAAATATTT GATTATGTCA TTTTAGTATG

       929        939        949        959        969        979        989        999       1009       1019       1029
TAAAATACCA AAATCTATTT CCAAGGAGCC CACTTTTAAA AATCTTTTCT TGTTTTAGGA AAGGTTTCTA AGTGAGAGGC AGCATAACAC TAATAGCACA GAGTCTGGGG

      1039       1049       1059       1069       1079       1089       1099       1109       1119       1129       1139
CCAGATATCT GAAGTGAAAT CTCAGCTCTG CCATGTCCTA GCTTTCATGA TCTTTGGCAA ATTACCTACT CTGTTTGTGA TTCAGTTTCA TGTCTACTTA AATGAATAAC

      1149       1159       1169       1179       1189       1199       1209       1219       1229       1239       1249
TGTATATACT TAATATGGCT TTGTGAGAAT TAGTAAGTAA ATGTAAAGCA CTCAGAACCG TGTCTGGCAT AAGGTAAATA CCATACAAGC ATTAGCTATT ATTAGTAGTA

      1259       1269       1279       1289       1299       1309       1319       1329       1339       1349       1359
TTAAAGATAA AATTTTCACT GAGAAATACA AAGTAAAATT TTGGACTTTA TCTTTTTACC AATAGAACTT GAGATTTATA ATGCTATATG ACTTATTTTC CAAGATTAAA

      1369       1379       1389       1399       1409       1419       1429       1439       1449       1459       1469
AGCTTCATTA GGTTGTTTTT GGATTCAGAT AGAGCATAAG CATAATCATC CAAGCTCCTA GGCTACATTA GGTGTGTAAA GCTACCTAGT AGCTGTGCCA GTTAAGAGAG

      1479       1489       1499       1509       1519       1529       1539       1549       1559       1569       1579
AATGAACAAA ATCTGGTGCC AGAAAGAGCT TGTGCCAGGG TGAATCCAAG CCCAGAAAAT AATAGGATTT AAGGGGACAC AGATGCAATC CCATTGACTC AAATTCTATT

      1589       1599       1609       1619       1629       1639       1649       1659       1669       1679       1689
AATTCAAGAG AAATCTGCTT CTAACTACCC TTCTGAAAGA TGTAAAGGAG ACAGCTTACA GATGTTACTC TAGTTTAATC AGAGCCACAT AATGCAACTC CAGCAACATA

      1699       1709       1719       1729       1739       1749       1759       1769       1779       1789       1799
AAGATACTAG ATGCTGTTTT CTGAAGAAAA TTTCTCCACA TTGTTCATGC CAAAAACTTA AACCCGAATT TGTAGAATTT GTAGTGGTGA ATTGAAAGCG CAATAGATGG

      1809       1819       1829       1839       1849       1859       1869       1879       1889       1899       1909
ACATATCAGG GGATTGGTAT TGTCTTGACC TACCTTTCCC ACTAAAGAGT GTTAGAAAGA TGAGATTATG TGCATAATTT AGGGGTGGTA GAATTCATGG AAATCTAAGT

      1919       1929       1939       1949       1959       1969       1979       1989       1999       2009       2019
TTGAAACCAA AAGTAATGAT AAACTCTATT CATTTGTTCA TTTAACCCTC ATTGCACATT TACAAAAGAT TTTAGAAACT AATAAAAATA TTTGATTCCA AGGATGCTAT

      2029       2039       2049       2059       2069       2079       2089       2099       2109       2119       2129
GTTAATGCTA TAATGAGAAA GAAATGAAAT CTAATTCTGG CTCTACCTAC TTATGTGGTC AAATTCTGAG ATTTAGTGTG CTTATTTATA AAGTGGAGAT GATACTTCAC

      2139       2149       2159       2169       2179       2189       2199       2209       2219       2229       2239
TGCCTACTTC AAAAGATGAC TGTGAGAAGT AAATGGGCCT ATTTTGGAGA AAATTCTTTT AAATTGTAAT ATACCATAGA AATATGAAAT ATTATATATA ATATAGAATC

      2249       2259       2269       2279       2289       2299       2309       2319       2329       2339       2349
AAGAGGCCTG TCCAAAAGTC CTCCCAAAGT ATTATAATCT TTTATTTCAC TGGGACAAAC ATTTTTAAAA TGCATCTTAA TGTAGTGATT GTAGAAAAGT AAAAATTTAA

      2359       2369       2379       2389       2399       2409       2419       2429       2439       2449       2459
GACATATTTA AAAATGTGTC TTGCTCAAGG CTATATTGAG AGCCACTACT ACATGATTAT TGTTACCTAG TGTAAAATGT TGGGATTGTG ATAGATGGCA TCCAAGAGTT

      2469       2479       2489       2499       2509       2519       2529       2539       2549       2559       2569
CCTTCTCTCT CAACATTCTG TGATTCTTAA CTCTTAGACT ATCAAATATT ATAATCATAG AATGTGATTT TTATGCCTTC CACATTCTAA TCTCATCTGG TTCTAATGAT

      2579       2589       2599       2609       2619       2629       2639       2649       2659       2669       2679
TTTCTATGCA GATTGGAAAA GTAATCAGCC TACATCTGTA ATAGGCATTT AGATGCAGAA AGTCTAACAT TTTGCAAAGC CAAATTAAGC TAAAACCAGT GAGTCAACTA

      2689       2699       2709       2719       2729       2739       2749       2759       2769       2779       2789
TCACTTAACG CTAGTCATAG GTACTTGAGC CCTAGTTTTT CCAGTTTTAT AATGTAAACT CTACTGGTCC ATCTTTACAG TGACATTGAG AACAGAGAGA ATGGTAAAAA

      2799       2809       2819       2829       2839       2849       2859       2869       2879                  2896
CTACATACTG CTACTCCAAA TAAAATAAAT TGGAAATTAA TTTCTGATTC TGACCTCTAT GTAAACTGAG CTGATGATAA TTATTATTCT AG GCC ACA GAA CTG AAA
                                                                                                  Ala Thr Glu Leu Lys
```

(*Fig. 3 continues on the next page.*)

```
            2911          2926                2941          2956          2971          2986
CAT CTT CAG TGT CTA GAA GAA GAA CTC AAA CCT CTG GAG GAA GTG CTA AAT TTA GCT CAA AGC AAA AAC TTT CAC TTA AGA CCC AGG GAC
His Leu Gln Cys Leu Glu Glu Glu Leu Lys Pro Leu Glu Glu Val Leu Asn Leu Ala Gln Ser Lys Asn Phe His Leu Arg Pro Arg Asp

            3001          3016                3035    3045    3055    3065    3075    3085
TTA ATC AGC AAT ATC AAC GTA ATA GTT CTG GAA CTA AAG GTAAGGCATT ACTTTATTTG CTCTCCTGGA AATAAAAAAA AAAAAGTAGG GGGAAAAGTA
Leu Ile Ser Asn Ile Asn Val Ile Val Leu Glu Leu Lys

     3095    3105    3115    3125    3135    3145    3155    3165    3175    3185    3195
CCACATTTTA AAGTGACATA ACATTTTTGG TATTTGTAAA GTACCCATGC ATGTAATTAG CCTACATTTT AAGTACACTG TGAACATGAA TCATTTCTAA TGTTAAATGA

     3205    3215    3225    3235    3245    3255    3265    3275    3285    3295    3305
TTAACTGGGG AGTATAAGCT ACTGAGTTTG CACCTACCAT CTACTAATGG ACAAGCCTCA TCCCAAACTC CATCACCTTT CATATTAACA CAAAACTGGG AGTGAGAGAG

     3315    3325    3335    3345    3355    3365    3375    3385    3395    3405    3415
AAGTGACTGA GTTGAGTTTC ACAGAAACGC AGGCAAGATT TTATTATATA TTTTTCAAGT TCCTTCACAG ATCATTTACT GGAATAGCCA ATACTGAGTT ACCTGAAAGG

     3425    3435    3445    3455    3465    3475    3485    3495    3505    3515    3525
CTTTTCAAAT GGTGTTTCCT TATCATTTGA TGGAAGGACT ACCCATAAGA GATTTGTCTT AAAAAAAAAA ACTGGAGCCA TTAAAATGGC CAGTGGACTA AACAAACAAC

     3535    3545    3555    3565    3575    3585    3595    3605    3615    3625    3635
AATCTTTTTA GAGGCAATCC CACTTTCAGA ATCTTAAGTA TTTTTAAATG CACAGGAAGC ATAAAATATG CAAGGGACTC AGGTGATGTA AAAGAGATTC ACTTTTGTCT

     3645    3655    3665    3675    3685    3695    3705    3715    3725    3735    3745
TTTTATATCC CGTCTCCTAA GGTATAAAAT TCATGAGTTA ATAGGTATCC TAAATAAGCA GCATAAGTAT AGTAGTAAAA GACATTCCTA AAAGTAACTC CAGTTGTGTC

     3755    3765    3775    3785    3795    3805    3815    3825    3835    3845    3855
CAAATGAATC ACTTATTAGT GGACTGTTTC AGTTGAATTA AAAAAATACA TTGAGATCAA TGTCATCTAG ACATTGACAG ATTCAGTTCC TTATCTATGG CAAGAGTTTT

     3865    3875    3885    3895    3905    3915    3925    3935    3945    3955    3965
ACTCTAAAAT AATTAACATC AGAAAACTCA TTCTTAACTC TTGATACAAA TTTAAGACAA AACCATGCAA AAATCTGAAA ACTGTGTTTC AAAAGCCAAA CACTTTTTAA
     3975    3985    3955    4005    4015    4025    4035    4045    4055    4065    4075
AATAAAAAAA TCCCAAGATA TGACAATATT TAAACAATTA TGCTTAAGAG GATACAGAAC ACTGCAACAG TTTTTTAAAA GAGAATACTT ATTTAAAGGG AACACTCTAT

     4085    4095    4105    4115    4125    4135    4145    4155    4165    4175    4185
CTCACCTGCT TTTGTTCCCA GGGTAGGAAT CACTTCAAAT TTGAAAAGCT CTCTTTTAAA TCTCACTATA TATCAAAATA GTTGCCTCCT TAGCTTATCA ACTAGAGGAA

     4195    4205    4215    4225    4235    4245    4255    4265    4275    4285    4295
GCGTTTAAAT AGCTCCTTTC AGCAGAGAAG CCTAATTTCT AAAAAGCCAG TCCACAGAAC AAAATTTCTA ATGTTTAAAG CTTTTAAAAG TTGGCAAATT CACCTGCATT

     4305    4315    4325    4335    4345    4355    4365    4375    4385    4395    4405
GATACTATGA TGGGGTAGGG ATAGGTGTAA GTATTTATGA AGATGTTCAT TCACACAAAT TTACCCAAAC AGGAAGCATG TCCTACCTAG CTTACTCTAG TGTAGCTCGT

     4415    4425    4435    4445    4455    4465    4475    4485    4495    4505    4515
TTCGTCTTTG GGGAAAATAT AAGGAGATTC ACTTAAGTAG AAAAATAGGA GACTCTAATC AAGATTTAGA AAAGAAGAAA GTATAATGTG CATATCAATT CATACATTTA

     4525    4535    4545    4555    4565    4575    4585    4595    4605    4615    4625
ACTTACACAA ATATAGGTGT ACATTCAGAG GAAAAGCGAT CAAGTTTATT TCACATCCAG CATTTAATAT TTGTCTAGAT CTATTTTTAT TTAAATCTTT ATTTGCACCC

     4635    4645    4655    4665    4675    4685    4695    4705    4715    4725    4735
AATTTAGGGA AAAAATTTTT GTGTTCATTG ACTGAATTAA CAAATGAGGA AAATCTCAGC TTCTGTGTTA CTATCATTTG GTATCATAAC AAAATACGCA ATTTTGGCAT

     4745    4755    4765    4775    4785    4795    4805    4815    4825    4835    4845
TCATTTTGAT CATTTCAAGA AAATGTGAAT AATTAATATG TTTGGTAAGC TTGAAAATAA AGGCAACAGG CCTATAAGAC TTCAATTGGG AATAACTGTA TATAAGGTAA

     4855    4865    4875    4885          4901          4916          4931
ACTACTCTGT ACTTTAAAAA ATTAACATTT TTCTTTTATA G GGA TCT GAA ACA ACA TTC ATG TGT GAA TAT GCT GAT GAG ACA GCA ACC ATT GTA GAA
                                             Gly Ser Glu Thr Thr Phe Met Cys Glu Tyr Ala Asp Glu Thr Ala Thr Ile Val Glu

4946          4961          4976          4991          5004    5014    5024    5034
 TTT CTG AAC AGA TGG  ATT ACC TTT TGT CAA AGC ATC ATC TCA ACA CTG ACT [TGA]TAATTAA GTGCTTCCCA CTTAAAACAT ATCAGGCCTT
 Phe Leu Asn Arg Trp  Ile Thr Phe Cys Gln Ser Ile Ile Ser Thr Leu Thr

     5044    5054    5064    5074    5084    5094    5104    5114    5124    5134    5144
CTATTTATTT AAATATTTAA ATTTTATATT TATTGTTGAA TGTATGGTTT GCTACCTATT GTAACTATTA TTCTTAATCT TAAAACTATA AATATGGATC TTTTATGATT

     5154    5164    5174    5184    5194    5204    5214    5224    5234    5244    5254
CTTTTTGTAA GCCCTAGGGG CTCTAAAATG GTTTCACTTA TTTATCCCAA AATATTTATT ATTATGTTGA ATGTTAAATA TAGTATCTAT GTAGATTGGT TAGTAAAACT

     5264    5274    5284    5294    5304    5314    5324    5334    5344    5354    5364
ATTT[AATAAA] TTTGATAAAT ATAAACAAGC CTGGATATTT GTTATTTTGG AAACAGCACA GAGTAAGCAT TTAAATATTT CTTAGTTACT TGTGTGAACT GTAGGATGGT

     5374    5384    5394    5404    5414    5424    5434    5444    5454    5464    5474
TAAAATGCTT ACAAAAGTCA CTCTTTCTCT GAAGAAATAT GTAGAACAGA GATGTAGACT TCTCAAAAGC CCTTGCTTTG TCCTTTCAAG GGCTGATCAG ACCCTTAGTT

     5484    5494    5504    5514    5524    5534    5544    5554
CTGGCATCTC TTAGCAGATT ATATTTTCCT TCTTCTTAAA ATGCCAAACA CAAACACTCT TGAAACTCTT CATAGATTTG GTGTGGC
```

FIG. 3. DNA sequence of the human *TCGF* gene. The amino acid sequences encoded by exons of *TCGF* are shown below the DNA sequences. The 3′- and 5′-untranslated regions are underlined. The T-A-T-A-A-A box, putative transcription initiation site, translation initiation codon, termination codon, and polyadenylylation signal are enclosed in rectangles.

transcription starts 53 nucleotides from the translation initiation site. Consistent with these findings, an A (indicated in Fig. 2), which occurs at most transcription initiation sites, is present at this position.

Comparison of the human genomic *TCGF* sequence with that of the cDNA revealed that the gene is divided into four exons separated by intervening sequences. Exon 1 contains the 5′-nontranslated region and codes for the first 49 amino

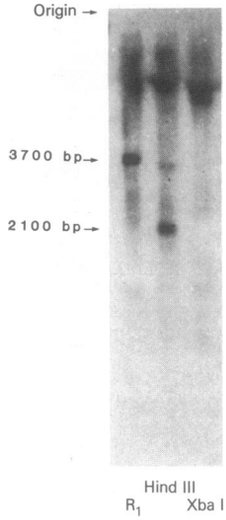

FIG. 4. Southern blot analysis of the *TCGF* gene in cellular DNA. Cellular DNA (10 μg) prepared from normal peripheral mononuclear cells of an individual and digested with *Eco*RI (R_1), *Hin*dIII, and *Xba* I was fractionated on an agarose gel, transferred to nitrocellulose paper, and hybridized with a ^{32}P-labeled TCGF cDNA probe.

acids of TCGF, 20 of which constitute the putative signal polypeptide (6). An intervening sequence of 91 bp separates this exon from exon 2, 60 bp long, which codes for the next 20 amino acids. The second and third exons are separated by a long intervening sequence of 2292 bp. Exon 3 (144 bp), which codes for the next 48 amino acids, is again followed by a long intervening sequence 1364 bp long. The fourth and final exon codes for the remaining 36 amino acid residues followed by the termination codon TGA. The polyadenylylation signal occurs 261 nucleotides after the termination codon. The consensus sequences, A-G and G-T, occur at the 5′ and 3′ exon–intron junctions, respectively, in all cases. A single difference in the nucleotide sequence of the gene coding for TCGF and that reported for the cDNA by Taniguchi *et al.* (6) was noted. Those workers reported an A at position 503 of their cDNA clone. This corresponds to position 4879 of our genomic sequence, where we find a G residue. Either nucleotide results in a codon for leucine. Computer analysis of the sequence revealed no segments of repetitive sequence within the introns.

It is worth noting that several regions of homology with a number of different potential enhancer core elements are present in the second intron. These include position 678–691 homologous with the Mo-MSV core sequence (18), position 1430–1440 homologous with the SV40 core element (19), position 1514–1527 homologous with the SNV element (20), and position 1702–1715 homologous with the mouse Ig heavy-chain core element (21).

The *TCGF* Gene in Normal and Malignant Cells. The organization of TCGF in cellular DNAs prepared from peripheral blood lymphocytes of 6 normal individuals, lymphoma tissue from 15 patient specimens, epithelial thymoma tissue from 1 patient specimen, and 7 cell lines was compared by Southern blot analysis after digestion of the DNA with the restriction endonucleases *Eco*RI, *Hin*dIII, and *Xba* I. The lymphoma tissues included 14 B- and 1 T-cell malignancies. The cell lines included the following: 4 MLA (gibbon ape T cell) subclones, 2 Jurkat (human T cell) subclones, and 1 HUT (human T-cell leukemia virus infected T-cell line) subclone. The findings were similar in all cases examined and a representative Southern blot is shown in Fig. 4. The restriction fragments hybridizing to the cDNA clone are precisely those predicted from the map in Fig. 1. The 900-bp *Xba* I fragment hybridizes weakly, because it has only 124 nucleotides homologous to the cDNA. The 4000-bp *Xba* I fragment and the 5000-bp *Hin*dIII fragment (Fig. 2) do not hybridize, because our cDNA has only 16 nucleotides from the fourth exon and therefore would not be expected to form a stable hybrid to this exon under the highly stringent conditions used. Since all of the restriction fragments hybridizing to the cDNA can be explained by the locus shown in Fig. 1, it is likely there is only a single copy of the human *TCGF* gene and in the limited number of malignant cells examined, the gene is not rearranged. This latter point must be qualified by the recognition that we have not examined a large number of malignant human T cells and that a more extensive survey is essential.

Although it is clear that TCGF plays a key role in mediating the T-cell proliferation that results from activation of T cells by mitogen, nothing is known of the molecular mechanisms associated with these processes. With the organization and sequence of the *TCGF* gene established, we are now in a position to study directly at the genomic level how mitogens regulate *TCGF*. Such studies should lead to an increased understanding of T-cell differentiation.

Note Added in Proof. Fujita *et al.* (22) also recently reported the complete nucleotide sequence for human *TCGF*. It is valuable to compare these two sequences obtained by different sequencing procedures as there are a number of small discrepancies, particularly within the third intron. There are no discrepancies in the coding or enhancer-like sequences.

We wish to thank Dr. Elaine Jaffe for support and thoughtful review of the manuscript and William Mohler for technical assistance. N.J.H. was supported by a fellowship from the Leukemia Society of America and a grant from the Eli Lilly Research Laboratories.

1. Morgan, D. A., Ruscetti, F. W. & Gallo, R. C. (1976) *Science* **193,** 1007–1008.
2. Smith, K. A. (1980) *Immunol. Rev.* **51,** 337–357.
3. Robb, R. J. & Smith, K. A. (1981) *Mol. Immunol.* **18,** 1087–1094.
4. Robb, R. J., Munck, A. & Smith, K. A. (1981) *J. Exp. Med.* **154,** 1455–1474.
5. Mier, J. W. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 6134–6138.
6. Taniguchi, T., Matsui, H., Fujita, T., Takaoka, C., Kashima, N., Yoshimoto, R. & Hamuro, J. (1983) *Nature (London)* **302,** 305–309.
7. Gillis, S. & Watson, J. (1980) *J. Exp. Med.* **152,** 1709–1719.
8. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
9. Smith, K. A., Favata, M. F. & Oroszlan, S. (1983) *J. Immunol.* **131,** 1808–1815.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 212–253.
11. Thayer, R. E. (1979) *Anal. Biochem.* **98,** 60–63.
12. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
13. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
14. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–507.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Queen, C. L. & Korn, L. J. (1980) *Methods Enzymol.* **65,** 595–609.
17. Bina-Stein, M., Thoren, M., Salzman, N. & Thompson, J. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4853–4857.
18. Dhar, R., McClements, N. L., Enquist, L. W. & Vande Woude, G. F. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3937–3941.
19. Weiner, H., Konig, M. & Gruss, P. (1983) *Science* **219,** 626–631.
20. Shimotohno, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285,** 550–554.
21. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. (1983) *Cell* **33,** 717–728.
22. Fujita, T., Takaoka, C., Matsui, H. & Taniguchi, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7437–7441.