Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 4806–4810, May 1994
Biochemistry

Stat3 and Stat4: Members of the family of signal transducers and activators of transcription

(transcription factors)

ZHONG ZHONG, ZILONG WEN, AND JAMES E. DARNELL, JR.*

Laboratory of Molecular Cell Biology, The Rockefeller University, New York, NY 10021

Contributed by James E. Darnell, Jr., January 24, 1994

ABSTRACT **The deduced amino acid sequence of two members of the *s*ignal *t*ransducers and *a*ctivators of *t*ranscription (STAT) family from the mouse are described. Comparison with the deduced protein sequence of the two previously described genes (Stat91 and Stat113), discovered because of their activation as transcription factors after interferon-induced tyrosine phosphorylation, shows several highly conserved regions, including the putative SH3 and SH2 domains. The conserved amino acid stretches likely point to conserved domains that enable these proteins to carry out the several required functions they are known and proposed to carry out. While Stat1 and Stat3 are widely expressed, Stat4 expression is restricted to testis, thymus, and spleen. Antiserum to Stat3 detects a major ≈92-kDa protein and a minor ≈89-kDa protein, while antiserum to Stat4 precipitates one major protein of ≈89 kDa.**

A family of proteins called STATs that serve a dual function as signal transducers and activators of transcription has been identified through the study of interferon (IFN) α and IFN-γ induction of transcription (28). In IFN-α treated cells, Stat113 and Stat91/84 (the Stat91 protein contains 38 carboxyl-terminal amino acids lacking in the Stat84) become phosphorylated on tyrosine, join a 48-kDa DNA-binding protein in the cell nucleus, and direct IFN-α-dependent transcription by binding to a specific response element, the IFN-α stimulated response element. In IFN-γ-treated cells, Stat91 (or Stat84) but not Stat113 becomes phosphorylated, forms a dimer (1), enters the nucleus, and binds to a different response element, the IFN-γ activated site (GAS) element, to direct IFN-γ-activated transcription. The Stat91 concentration is greater than Stat84 in most cells, and only Stat91 (not Stat84) activates IFN-γ-dependent transcription (2, 3). Recently several laboratories have found that a number of polypeptide ligands other than the IFNs can induce tyrosine phosphorylation on Stat91, and other such ligands induce DNA-binding factors that contain phosphorylated tyrosine on proteins similar in size to 91 kDa but not reactive with anti-Stat91 antiserum (4–10).

The initial results with IFN-α and IFN-γ, plus the newer results with other ligands, suggested that a family of STAT proteins might exist to specifically activate different genes in response to different ligands. With that possibility in mind, we have screened mouse cell cDNA libraries for additional STAT family members and report here the identification of two such proteins. For simplicity, we propose that the family members be named in the order in which they were characterized. Thus, the Stat91 and Stat84 would be designated Stat1α and Stat1β, and the Stat113 protein would become Stat2. The two additional proteins are therefore named Stat3 and Stat4. As described elsewhere, Stat3 has been found to be activated as a DNA-binding protein through phosphorylation on tyrosine in cells treated with epidermal growth factor (EGF) or interleukin (IL) 6 but not after IFN-γ treatment (11). No ligand has yet been found to activate Stat4. J. Ihle and colleagues independently have cloned the Stat4 cDNA (29).

MATERIALS AND METHODS

PCR and Library Screening. Primer 1 (5′-CCAGATCTC-TGGAATGATGG) and primer 2 (5′-GGGAATTCTCAGG-TATATTCTC), containing a *Bgl* II and a *Eco*RI site, respectively, at their 5′ termini (underlined), were synthesized based on the two conserved peptide sequences in the SH2 domain of human Stat1 (12). Cytoplasmic RNA obtained from BALB 3T3 cells by Nonidet P-40 lysates was reverse transcribed to cDNA, and 35 pmol of these two primers were used to amplify a mouse Stat1 fragment, termed M16, in a PCR reaction (30 cycles: 94°C, 1 min, 45°C, 1 min, 72°C, 2 min) (13, 14).

[^{32}P]dATP-labeled M16, using Klenow and random hexamers according to ref. 13, was used to screen a mouse thymus cDNA library (Stratagene; a gift of M. Nussenzweig) constructed in the ZAP vector. The hybridization was carried out at 42°C and washed at 42°C before the first exposure (15). Then the filters were washed in 2× SSC/0.1% SDS at 65°C for a second exposure. Stat1 clones survived the 65°C washing, whereas Stat3 and Stat4 clones were identified as plaques that lost signals at 65°C. The plaques were purified and subcloned according to Stratagene commercial protocols.

RNA Extraction and Analysis. RNA extraction from frozen tissue samples employed guanidinium thioyanate homogenization and acidic/phenol extraction, and Northern blots were carried out on poly(A)-containing RNA (13) bound to a Nytran filter (Schleicher & Schuell). Hybridization was performed as above using M16, a 5′ 1.1-kb *Eco*RI–*Sph* I fragment, or a 5′ 1-kb *Eco*RI–*Sph* I fragment as templates for making probes for Stat1, Stat3, and Stat4, respectively. A glyceraldehyde-3-phosphate dehydrogenase probe was used to ensure quality of extracted RNA.

Protein Extraction and Cell Lines. U937 and LE929 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (vol/vol) calf serum. 70Z and HT2 cells were grown in RPMI 1640 containing 10% calf serum. IL-2 (0.2 ng/ml) was added to HT2 medium. Thymocyte suspensions were prepared from thymic lobes according to standard procedures. Cell lysates were prepared according to ref. 2.

Antibodies and Immunoblots. Amino acids 688–727 of Stat3 and 678–743 of Stat4 were subcloned into pGEX1λt (Pharmacia) by PCR with oligonucleotides based on the boundary

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: STAT, signal transducers and activators of transcription; IFN, interferon; GAS, IFN-γ-activated site; IL, interleukin; EGF, epidermal growth factor; SIE, sis-inducible element.

*To whom reprint requests should be addressed.

sequence plus restriction sites (*Bam*HI at the 5′ end and *Eco*RI at the 3′ end), allowing in-frame fusion with glutathione *S*-transferase. One milligram of each antigen was used for the immunization, and three booster injections 4 weeks apart were given. Anti-Stat3C and -Stat4C sera were used at a 1:1000 dilution in Western blots using standard protocols (16).

COS Cell Transient Transfections. Plasmids used in the transfection experiments, Rc/CMV-Stat3 and Rc/CMV-Stat4, were derived by inserting the complete coding sequences of Stat3 and Stat4, respectively, into the *Not* I and *Apa* I sites of the expression vector Rc/CMV (Invitrogen). COS1 cells were grown in DMEM containing 10% (vol/vol) bovine calf serum. Rc/CMV-Stat3 or -Stat4 were introduced into the COS1 cells through DNA/calcium phosphate precipitates (13). Cell extracts were prepared 60 hr after transfection for Western blot.

DNA and Protein Sequence Analysis. All the sequence analysis was performed using the Genetics Computer Group package (17). Specifically, multiple sequence alignment was generated by PILEUP, and distance matrix was done with DISTANCE.

RESULTS

Cloning of Stat3 and Stat4. To begin searching for additional family members, we wished to switch from human DNA libraries from which the original Stat1 and Stat2 clones were obtained to mouse libraries in order to simplify the study of gene expression in differentiated tissue. Using a pair of deoxyoligonucleotide primers based on the conserved sequence of the Stat1 SH2 domain, we amplified a fragment containing the mouse Stat1 SH2 region from BALB 3T3 cell line RNA.

This probe was chosen to screen for other STAT family members because while the Stat1 and Stat2 SH2 domains are quite similar over the entire 100- to 120-amino-acid region, only the amino-terminal half of the STAT SH2 domains strongly resemble the SH2 regions found in other proteins (12, 18). Thus we hoped that the ≈300-nucleotide probe to the STAT SH2 region might preferentially select for STAT family members over other SH2-containing proteins. A cDNA library screen with the labeled 300-nucleotide probe at low stringency resulted in many strongly hybridizing signals from individual clones, seven of which were partially sequenced and shown to be mouse Stat1. When the mouse Stat1 sequence was completed it was ≈90% identical with the human Stat1 cDNA sequence, and the deduced amino acid sequence was ≈95% identical to the human sequence (Table 1). A number of mouse clones that hybridized to the SH2 probe at low stringency gave no signals after washing at high stringency. Ten additional examples of this type were selected for analysis. All but two of these hybridizing clones contained segments of the same cDNA. The sequence encoded a protein clearly similar to, but not identical with, the known Stat1 and Stat2 sequences and was designated Stat3. The other two clones that did not match Stat3 were identical in regions of overlap and were designated Stat4. The apparent start site for translation in Stat3, the methionine listed as amino acid 1 in Fig. 1, had an upstream sequence AGGATGG that agrees with the Kozak consensus (19) and as there were no other potential in frame upstream start sites, this is presumably the start site. The amino acid sequence just downstream of the ATG also agreed with the Stat1 and Stat2 amino acid sequence at the translation initiation sites (MS or AQW in all four proteins). Using this methionine as amino acid 1, the cDNA for Stat3 would encode 770 amino acids. The sequence of Stat4 as depicted in Fig. 1 encodes 749 amino acids. [As we will discuss later (Fig. 2), these clones can encode proteins of ≈92 kDa (Stat3) and ≈89 kDa (Stat4) when transfected into COS cells.]

Table 1. Distance matrix of Stat proteins

	Stat1	m-Stat1	Stat3	Stat4	Stat2
Stat1	1.0000	0.9453	0.6907	0.6671	0.5733
m-Stat1		1.0000	0.6796	0.6604	0.5821
Stat3			1.0000	0.6310	0.5403
Stat4				1.0000	0.5374
Stat2					1.0000

Pair-wise comparisons of Stat proteins were carried out by Genetics Computer Group PILEUP, and homology scores (not strictly equivalent to identity) were calculated by DISTANCE and entered in the distance matrix. m-Stat1, mouse Stat1.

The detailed comparisons of the amino acids of the several known STAT sequences are given in Fig. 1 and Table 1. As we have noted above, the Stat1α sequences of mouse and human are nearly identical, and the previously described ≈35% amino acid identity between Stat1α and Stat2 (Stat113) is obvious. The comparison of each of the two new sequences to either Stat1α or Stat2 or to each other revealed approximately 30–40% identity between any pair of the four known sequences. The regions of amino acid identity in all these comparisons are scattered throughout an ≈700-amino-acid stretch—that is, nearly the entire length of the proteins (Fig. 1; see consensus). Stat1, -3, and -4 are somewhat more highly related to each other than any of the three are to Stat2 (Table 1). Thus the genes encoding these proteins presumably were duplicated early in evolution, but the maintenance of the scattered sequence identity over such a long region (700 amino acids) suggests several different domain functions in the proteins must be maintained.

The chromosomal locations of Stat1, -2, -3, and -4 have now been determined and are described elsewhere (29; N. Copeland, N. A. Jenkins, Z.Z., Z.W., and J.E.D., unpublished results). Only two of the four genes are on the same chromosome. Stat1 and Stat4 are located in the centromeric region of mouse chromosome 1 (human 2q 32-q34) with the other two genes on other chromosomes.

There are several regions in which there is a particularly high concentration of sequence identity between all four sequences. (Reference numbers on the consensus lane in Fig. 1 are used throughout the text. For the corresponding region of a specific protein, amino acid numbers can be determined from coordinates listed on the right of Fig. 1.) For example in the amino terminus, 26 of the first 110 amino acids are invariant. This is followed by a region of lower homology, which nevertheless contains one conserved feature: from amino acid 207 to amino acid 235, there is a heptad leucine repeat (some residues are analogous hydrophobic repeat amino acids and not leucine residues). When this region was displayed in a helical wheel diagram (17), there was a hydrophobic face and a charged face in each of the four proteins, suggesting an amphiphilic helical structure that could allow coil–coil interaction between subunits (data not shown). A deletion of 22 amino acids in this region inactivates Stat1α so that no phosphorylated protein is detected (T. Improta, C. Schindler, C. Horvath, I. M. Kerr, G. R. Stark, and J.E.D., unpublished results).

Particularly striking is the high degree of amino acid conservation in all four proteins beginning at amino acid ≈250 and extending for much of the next ≈230 amino acids with several blocks of absolute sequence identity. We have no current idea of the function of this region. The putative SH3 and the SH2 domains follow between the position ≈510 and position 710. The SH3 domain is known in other proteins to mediate interactions with proline-rich sequences (20), and the SH2 forms a pocket that binds phosphotyrosine, with different SH2 groups having affinity for different phosphopeptides (21). The SH2 group in the Stat1 has recently

```
        M*QW  *Q*L*  *  Q* Q Y     P***R  LA W*E Q*W  A*      *  55
Stat1   MSQWYELQQLDSKFLEQVHQLYDDS.FPMEIRQYLAQWLEKQDWEHAA..NDVSF 52
mStat1  MSQWFELQQLDSKFLEQVHQLYDDS.FPMEIRQYLAQWLEKQDWEHAA..YDVSF 52
mStat3  MAQWNQLQQLDTRYLKQLHQLYSDT.FPMELRQFLAPWIESQDWAYAA..SKESH 52
mStat4  MSQWNQVQQLEIKFLEQVDQFYDDN.FPMEIRHLLAQWIETQDWEVAS..NNETM 52
Stat2   MAQWEMLQNLDSPFQDQLHQLYSHSLLPVDIRQYLAVWIEDQNWQEAALGSDDSK 55

        AT*      L *** * *R   *    L  HN***     *Q  *   P  ** **  110
Stat1   ATIRFHDLLSQLDDQYSRFSLE.NNFLLQHNIRKSKRNLQDNFQEDPIQMSMIIY 106
mStat1  ATIRFHDLLSQLDDQYSRFSLE.NNFLLQHNIRKSKRNLQDNFQEDPVQMSMIIY 106
Stat3   ATLVFHNLLGEIDQQYSRFLQE.SNVLYQHNLRRIKQFLQSRYLEKPMEIARIVA 106
Stat4   ATILLQNLLIQLDEQLGRVSKE.KNLLLIHNLKRIRKVLQGKFHGNPMHVAVVIS 106
Stat2   ATMLFFHFLDQLNYECGRCSQDPESLLLQHNLRKFCRDIQP.FSQDPTQLAEMIF 109

          L EE **L  A   Q            *   Q  ** **  **   *        165
Stat1   SCLKEERKILENAQRFNQ..AQSGNIQSTVMLDKQKELDSKVRNVKDKVMCIEHE 159
mStat1  NCLKEERKILENAQRFNQ..AQEGNIQNTVMLDKQKELDSKVRNVKDQVMCIEQE 159
mStat3  RCLWEESRLLQTAATAAQQGGQANHPTAAVVTEKQQMLEQHLQDVRKRVQDLEQK 161
mStat4  NCLREERRILAAANMPIQ.GPLEKSLQSSSVSERQRNVEHKVSAIKNSVQMTEQD 160
Stat2   NLLLEEKRILIQAQRA.QL.EQGEPVLETPVESQQHEIESRILDLRAMMEKLVKS 162

           * * QD * *  K             *              L     LD  R*  220
Stat1   IKSLEDLQDEYDFKCKT..LQNREHETNGVAKSDQKQEQLLLKKMYLMLDNKRKE 212
mStat1  IKTLEELQDEYDFKCKT..SQNREGEANGVAKSDQKQEQLLLHKMFLMLDNKRKE 212
mStat3  MKVVENLQDDFDFNYKTLKSQGDMQDLNGNNQSVTRQKMQQLEQMLTALDQMRRS 216
mStat4  TKYLEDLQDEFDYRYKTIQT.MDQGDKNSIL..VNQEVLTLLQEMLNSLDFKRKE 212
Stat2   ISQLKDQQDVFCFRYKI.....QAKGKTPSLDPHQTKEQKILQETLNELDKRRKE 212

         *     **       *     L *WK  QQ ACI *P    L* L*   T  A 275
Stat1   VVHKIIELLNVTELTQNALINDELVEWKRRQQSACIGGPPNACLDQLQNWFTIVA 267
mStat1  IIHKIRELLNSIELTQNTLINDELVEWKRRQQSACIGGPPNACLDQLQTWFTIVA 267
mStat3  IVSELAGLLSAMEYVQKTLTDEELADWKRRQQIACIGGPPNICLDRLENWITSLA 271
mStat4  ALSKMTQIVNETDLLMNSMLLEELQDWKKRQQIACIGGPLHNGLDQLQNCFTLLA 267
Stat2   VLDASKALLGRLT.TLIELLLPKLEEWKAQQQKACIRAPIDHGLEQLETWFTAGA 266

              RQ *  L        *Y  DP*               L   L   *FVVE Q 330
Stat1   ESLQQVRQQLKKLEELEQKYTYEHDPITKNKQVLWDRTFSLFQQLIQSSFVVERQ 322
mStat1  ETLQQIRQQLKKLEELEQKFTYEPDPITKNKQVLSDRTFLLFQQLIQSSFVVERQ 322
mStat3  ESQLQTRQQIKKLEELQQKVSYKGDPIVQHRPMLEERIVELFRNLMKSAFVVERQ 326
mStat4  ESLFQLRQQLEKLQEQSTKMTYEGDPIPAQRAHLLERATFLIYNLFKNSFVVERQ 322
Stat2   KLLFHLRQLLKELKGLSCLVSYQDDPLTKGVDLRNAQVTELLQRLLHRAFVVETQ 321

        PCMP  P RP**LKT   FT * RLL**  E N    *    D**            385
Stat1   PCMPTHPQRPLVLKTGVQFTVKLRLLVKLQELNYNLKVKVLFDKDVNERNTVKGF 377
mStat1  PCMPTHPQRPLVLKTGVQFTVKSRLLVKLQESNLLTKVKCHFDKDVNEKNTVKGF 377
mStat3  PCMPMHPDRPLVIKTGVQFTTKVRLLVKFPELNYQLKIKVCIDKDSGDVAALRGS 381
mStat4  PCMPTHPQRPMVLKTLIQFTVKLRLLIKLPELNYQVKVKASIDKNVSTLSN.... 373
Stat2   PCMPQTPHRPLILKTGSKFTVRTRLLVRLQEGNESLTVEVSIDRNPPQ...LQGF 373

        R*F * **  K *  E *   *L  *   L   E *   *G     *   VTEEL 440
Stat1   RKFNILGTHTKVMNMEESTNGSLAAEFRHLQLKEQK..NAGTRTNEGPLIVTEEL 430
mStat1  RKFNILGTHTKVMNMEESTNGSLAAELRHLQLKEQK..NAGNRTNEGPLIVTEEL 430
mStat3  RKFNILGTNTKVMNMEESNNGSLSAEFKHLTLREQRCGNGGRANCDASLIVTEEL 436
mStat4  RRFVLCGTHVKAMSSEESSNGSLSVEFRHLQPKEMKC.STGSKGNEGCHMVTEEL 427
Stat2   RKFNILTSNQKTLTPEKGQSQGLIWDFGYLTLVEQRSGGSGKGSNKGPLGVTEEL 428

        H **F      GL  *L T *LPVV*ISN* Q** *WAS**W*N*   *  N   495
Stat1   HSLSFETQLCQPGLVIDLETTSLPVVVISNVSQLPSGWASILWYNMLVAEPRNLS 485
mStat1  HSLSFETQLCQPGLVIDLETTSLPVVVISNVSQLPSGWASILWYNMLVTEPRNLS 485
mStat3  HLITFETEVYHQGLKIDLETHSLPVVVISNICQMPNAWASILWYNMLTNNPKNVN 491
mStat4  HSITFETQICLYGLTINLETSSLPVVMISNVSQLPNAWASIIWYNVSTNDSQNLV 482
Stat2   HIISFTVKYTYQGLKQELKTDTLPVVIISNMNQLSIAWASVLWFNLLSPNLQNQQ 483

        FF  PP      *    *SWQFSS   RGL* *QL  L *KL *        **W 550
Stat1   FFLTPPCARWAQLSEVLSWQFSSVTKRGLNVDQLNMLGEKLLGPNAS.PDGLIPW 539
mStat1  FFLNPPCAWWSQLSEVLSWQFSSVTKRGLNADQLSMLGEKLLGPNAG.PDGLIPW 539
mStat3  FFTKPPIGTWDQVAEVLSWQFSSTTKRGLSIEQLTTLAEKLLGPGVNYSGCQITW 546
mStat4  FFNNPPSVTLGQLLEVMSWQFSSYVGRGLNSEQLNMLAEKL.TVQSNYNDGHLTW 536
Stat2   FFSNPPKAPWSLLGPALSWQFSSYVGRGLNSDQLSMLRNKLFGQNCRTEDPLLSW 538

         * F K         *FW W** ***L**  *  LWN*G IMGF*S*  ER *L   605
Stat1   TRFCKENINDKNFPFWLWIESILELIKKHLLPLWNDGCIMGFISKERERALLKDQ 594
mStat1  TRFCKENINDKNFSFWPWIDTILELIKNDLLCLWNDGCIMGFISKERERALLKDQ 594
mStat3  AKFCKENMAGKGFSFWVWLDNIIDLVKKYILALWNEGYIMGFISKERERAILSTK 601
mStat4  AKFCKEHLPGKTFTFWTWLEAILDLIKKHILPLWIDGYIMGFVSKEKERLLLKDK 591
Stat2   ADFTKRESPPGKLPFWTWLDKILELVHDHLKDLWNDGRIMGFVSRSQERRLLKKT 593

          *GTFLLRFSE S  G**T *WV*  *          *V*PY K  L  ** **I*  660
Stat1   QPGTFLLRFSESSREGAITFTWVERSQNGGEPDFHAVEPYTKKELSAVTFPDIIR 649
mStat1  QPGTFLLRFSESSREGAITFTWVERSQNGGEPDFHAVEPYTKKELSAVTFPDIIR 649
mStat3  PPGTFLLRFSESSKEGGVTFTWVEK.DISGKTQIQSVEPYTKQQLNNMSFAEIIM 655
mStat4  MPGTFLLRFSE.SHLGGITFTWVDQSEN.GEVRFHSVEPYNKGRLSALAFADILR 644
Stat2   MSGTFLLRFSESS.EGGITCSWVEH.QDDDKVLIYSVQPYTKEVLQSLPLTEIIR 646

          Y **   NI   PL *LYP I ** AFG  Y         E           Y*  715
Stat1   NYKVMAAENIPENPLKYLYPNIDKDHAFGKYYSRPKEAPEPMELDGPKGTGYIKT 704
mStat1  NYKVMAAENIPENPLKYLYPNIDKDHAFGKYYSRPKEAPEPMELDDPKRTGYIKT 704
mStat3  GYKIMDATNILVSPLVYLYPDIPKEEAFGK.YC.RPESQEHPEADPGSAAPYLKT 708
mStat4  DYKVIMAENIPENPLKYLYPDIPKDKAFGKHYS..SQPCEVSRPTERGDKGYVPS 697
Stat2   HYQLLTEENIPENPLRFLYPRIPRDEAFGCYYQEKVNLQERRK........YLKH 693

          I **                 *    L **       *                770
Stat1   ELISVSEVHPSRL..........QTTDNLLPMSPEEFDEVSRIVGSVE....... 742
mStat1  ELISVSEVHPSRL..........QTTDNLLPMSPEEFDEMSRIVGP.E....... 741
mStat3  KFICVTPTTCS............NTID..LPMSPRTLDSLMQFGNNGEGAEPSAG 749
mStat4  VFIPISTIRSDST........EPQSPSDLLPMSPSAYAVLRENLSPTTIETAMNS 744
Stat2   RLIVVSNRQVDELQQPLELKPEPELESLELELGLVPEPELSLDLEPLLKAGLDLG 748

Stat1   ..FDSMMNTV  750
mStat1  ..FDSMMSTV  749
mStat3  GQFESLTFDMDLTSECATSPM  770
mStat4  PYSAE  749
Stat2   PELESVLESTLEPVIEPTLCMVSQTVPEPDQGPVSQPVPEPDLPCDLRHLNTEPM 803

Stat2   EIFRNCVKIEEIMPNGDPLLAGQNTVDEVYVSRPSHFYTDGPLMPSDF 851
```

FIG. 1. The primary sequences and comparisons of STAT family members. Complete amino acid sequences for Stat1 (human), mouse Stat1 (mStat1), Stat2 (human), mouse Stat3 (mStat3), and mouse Stat4 (mStat4) are aligned, introducing gaps (·) to maximize amino acid identity. Amino acid coordinates for individual proteins are labeled on the right; reference numbers for the whole group are indicated above the coordinates. Consensus sequence is indicated by an asterisk (*) or the identical amino acid where a residue is conserved for all Stat members. Amino acids were scored as homologous if they were within a single group of neutral (P, A, C, G, S, T), hydrophilic/acidic (Q, N, E, D), hydrophilic/basic (H, K, R), hydrophobic (L, I, V, M), or aromatic (F, Y, W).

been shown necessary for phosphorylation of that protein (22) and after phosphorylation for homodimer formation of Stat1 (1). The putative SH3 group begins about amino acid 510 and ends about amino acid 565. The conservation of this region within the four STAT proteins is considerably higher (>40% identity) than with other SH3 regions, which are not a highly conserved region compared, for example, to SH2 domains. The SH2 domain of all four proteins begins at about amino acid 585 with a tryptophan preceded by two leucine residues (21) and continues in a highly homologous fashion to amino acid ≈700. There is divergence at about 700–710, the region where the tyrosine residue that becomes phosphorylated is located in Stat1 and Stat2. It is interesting that the amino acids downstream from the phosphorylated tyrosine of Stat1 (residue 701) are absolutely conserved between mouse and human Stat1 but vary in the other STATs. These downstream amino acids are often important in the recognition of tyrosine-containing peptides (23). There is a tyrosine in a homologous position (≈710 in the consensus alignment of Fig. 1) in Stat3 and Stat4, both of which are known from other experiments to be phosphorylated on tyrosine. Stat3 is phosphorylated in cells treated by EGF and IL-6 (11). We have not discovered a ligand that causes phosphorylation of Stat4, but it becomes phosphorylated on tyrosine in cells treated with vanadate and peroxide (unpublished results).

We can anticipate a number of functional requirements for domains of the STAT proteins such as receptor and/or kinase recognition, phosphorylation target sites, dimerization domains, a binding site to facilitate nuclear transport, a DNA-binding domain, a transcriptional activation domain, and a phosphatase recognition domain. The identification of the more highly conserved amino acid stretches discussed above should assist in pinpointing some of the residues that are important in these functions.

Tissue and Cell Distribution of Stat3 and Stat4 mRNA and Protein. With the availability of mouse cDNA probes for several STAT family members, we examined mRNA preparations from a variety of tissues for the presence of the mRNA encoding the proteins (Fig. 2). For Stat1 mRNA, signals were obtained from a number of tissues tested. The highest concentrations were in thymus and heart, but it is also present in spleen, liver, and testis. In all cases, two major mRNA bands were observed, which presumably represented Stat1α (91 kDa) and Stat1β (84 kDa) forms. The proportion of the two mRNAs did not vary greatly in different tissues. A third higher molecular mass band was observed in heart, thymus, and spleen, but it has not yet been characterized.

A strong Stat3 mRNA signal was detected in brain, heart, liver, testis, and thymus, the source of the cDNA library from which the cloned cDNA was isolated. This mRNA, which is ≈5 kb, was also detected in spleen (Fig. 2*B*). Stat4 had the most restricted tissue distribution. Only weak mRNA signals from a 3.3-kb species were obtained in the thymus and spleen. However, in mRNA from the testis, there was a very large amount of Stat4 mRNA present as two forms, one of which was the same size as that found in the spleen and thymus and the other was slightly larger (Fig. 2*C* and data not shown). In agreement with this data, previous experiments suggested that Stat1 proteins were present in a large number of cell lines, such as HeLa, A431, and FS2 (4, 12). To determine whether the Stat3 and Stat4 proteins were present in cells, protein blots were carried out with antisera against each protein. To avoid cross-reactivity, antibodies were raised against the carboxyl terminus of Stat3 and Stat4, the

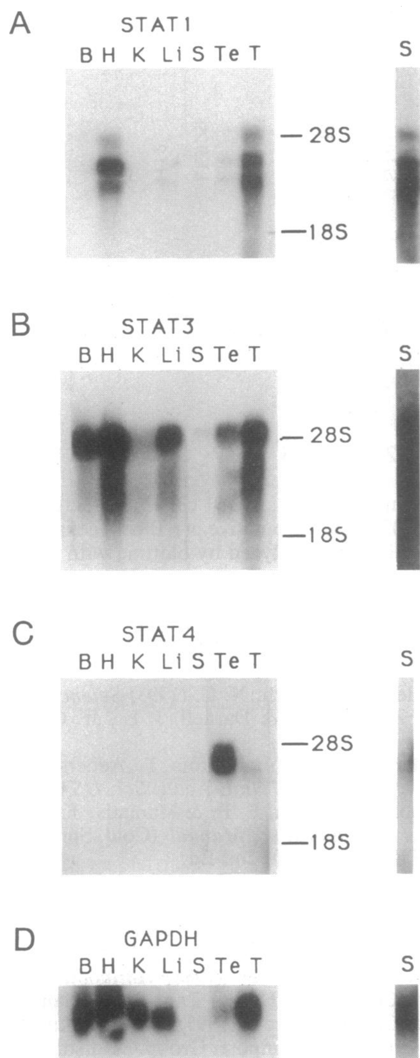

FIG. 2. Northern blot of different tissue RNA samples. Poly(A)-containing RNA was prepared from brain (B), heart (H), kidney (K), liver (L), spleen (S), thymus (T), and testis (Te). The blot was hybridized with Stat1 (*A*), Stat3 (*B*), Stat4 (*C*), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH; *D*) probes. The migration of 28S and 18S rRNA, which served as markers, are indicated. Less RNA was recovered from spleen; longer exposures of blots for the spleen sample is shown on the right.

less homologous region of the protein. The proteins were unambiguously found in several tissues where the mRNA was known to be present, but we did not test for protein in all tissues (data not shown). We also checked protein expression in several cell lines. A protein of 89 kDa, which was reactive with Stat4 antiserum, was expressed in 70Z cells, a pre-B cell line, but not in many other cell lines. Stat3 was highly expressed predominantly as a 92-kDa protein in 70Z, HT2 (a mouse helper T-cell clone), and U937 (a macrophage-derived cell) cell lines (Fig. 3). J. Ihle and colleagues have found Stat4 to be present at high level in myeloid cells during growth on erythropoietin or IL-3, but the protein is not phosphorylated in response to these ligands (29).

COS Cell Overexpression. To prove that we had obtained functional full-length clones of Stat3 and Stat4 cDNA, we subcloned the open reading frame of either cDNA into an expression vector, Rc/CMV (Invitrogen), downstream of a cytomegalovirus promoter. The resulting plasmids were transfected into COS1 cells, and proteins were extracted 60 hr posttransfection and examined by Western blot after electrophoresis. Untransfected COS1 cells expressed a low level of 92-kDa Stat3 protein but did not express a detectable level of Stat4 (Fig. 3*A*, lane 1; Fig. 3*B*, lane 1). Upon transfection of the Stat3-expressing plasmid, the 92-kDa Stat3 was increased at least 10-fold. An 89-kDa protein antigenically related to Stat3 was a minor band in most cell line extracts but was also increased posttransfection (Fig. 3*B*, lane 2). This protein therefore possibly represents another form of Stat3 protein. It could also be an antigenically related protein whose synthesis is stimulated by the increased 92-kDa protein. Transfection of RcCMV/Stat4 leads to the expression of an 89-kDa reactive band indistinguishable in size from the p89 Stat4 found in 70Z cell extracts (Fig. 3*A*, lane 3).

DISCUSSION

The work described in this paper establishes that the STAT family of proteins is larger than simply those proteins that act in the IFN pathway. What is the role of different STAT family members and how many might we expect? In the IFN-α and IFN-γ pathways, two distinct response elements were characterized (24, 25, 28), and different STAT protein complexes were shown to bind the two elements: Stat91 to the GAS element possibly in conjunction with other proteins at some other sites (26, 27) and Stat91 and Stat113 plus a 48-kDa DNA-binding protein to the IFN-α-stimulated response element (24). What remains less clear is how many clearly different response elements exist and therefore how many different binding proteins might be anticipated.

As we noted in the Introduction, Stat3 has been shown to be phosphorylated after EGF and IL-6 treatment and to bind to the sis-inducible element (SIE) site, a site from the c-*fos* promoter that is similar to but not identical with the GAS element that is present in IFN-γ-induced genes (4). Stat3 not only binds DNA, probably as a homodimer, but also can heterodimerize with Stat1 and bind DNA (Z.Z., Z.W., and J.E.D., unpublished results). Both the Stat3 homodimers and the Stat3–Stat1 heterodimers can recognize the natural SIE, whereas the Stat1 homodimer does not. Since the sequences to which the EGF (SIE)- and IFN-γ (GAS)-induced proteins bind are similar, it suggests that a small sequence variation around a consensus GAS element may be bound by different STAT complexes. Variability of STAT complexes might be further increased by alternative splicing or different translation initiation sites. The Stat1 gene gives rise to two different Stat1 proteins (Stat91 and Stat84) by alternative splicing (12). A third form of RNA species is readily detectable in multiple tissues using Stat1 probe under stringent conditions (Fig. 2). This higher molecular weight RNA band is present in cytoplasmic RNA prepared from cell lines (Z.Z., data not shown) and therefore is unlikely to be a nuclear precursor to Stat1 mRNA. Two forms of mRNA for Stat4 were also detected. Whether in these cases different proteins are generated needs to be explored. Two protein bands were detected by Stat3 antiserum, a major form at 92 kDa and a minor band at 89 kDa. The 89-kDa protein was also recognized by an antiserum against the N-terminal of Stat3 (data not shown). It is possibly a different form of Stat3. Alternative forms for proteins from a single gene could be potentially different in their interaction domains and therefore mediate transcriptional responses for different ligands.

In addition to the variations in the STAT proteins themselves, the IFN-α-induced multiprotein factor IFN-stimulated gene factor 3 illustrates another means of increasing the diversity of transcription factors containing STAT proteins. In IFN-stimulated gene factor 3 the Stat1 and Stat2 cooperate with a third protein, the 48-kDa binding protein from a different family, to yield a multiprotein factor that recognizes a DNA element outside the GAS family. Perhaps a relatively few STAT family members (say 10 or less) plus

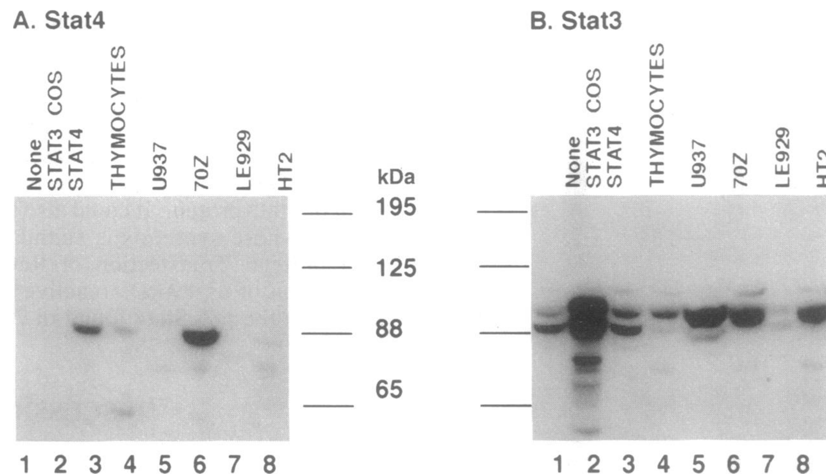

FIG. 3. Expression of Stat3 and Stat4 proteins in different cells. Protein extracts from COS1 cells, COS1 cells transfected with Stat3- and Stat4-expressing plasmids, thymocytes, U937 (a human macrophage-like cell line) cells, 70Z (a mouse putative pre-B cell) cells, LE929 (a human undifferentiated fibroblast-like cell line) cells, and HT2 (a mouse helper T-cell clone) cells were analyzed by blotting with anti-Stat4 (*A*) and anti-Stat3 (*B*) antiserum.

variations in binding within the 48-kDa family could provide the extent of required variations for activating ligand-triggered genes. However, if a very large number of specific response elements exist, there could be dozens of STAT proteins.

The overall role of the STAT proteins in intracellular signaling will, it appears, also be conditioned by differential expression in different tissues. Stat4 is restricted to the thymus, spleen, and testis. Even Stat1 and Stat3, which are widely expressed in cultured cells, are not available in many tissues. Such results might also suggest that other proteins in the family will be found in tissues where some of the presently described proteins have not been located, since all cells most probably bind ligands and respond transcriptionally.

It seems highly likely that the coming years will see an explosion in our knowledge of this protein family, but already we have reason to argue for its general importance in signaling transcriptional events in the cell nucleus in response to polypeptide ligands.

The first two authors contributed equally to this work. We thank D. Schwartz and M. Nuzzenzweig for cDNA libraries and for assistance in the culture of HT2 cells. We thank J. Ihle for sharing unpublished information. This work was supported by National Institutes of Health Grant AI32489 (J.E.D.).

1. Shuai, K., Horvath, C. M., Tsai-Huang, L. H., Qureshi, S., Cowburn, D. & Darnell, J. E., Jr. (1994) *Cell* **76**, 821–828.
2. Shuai, K., Stark, G. R., Kerr, I. M. & Darnell, J. E., Jr. (1993) *Science* **261**, 1744–1746.
3. Muller, M., Laxton, C., Briscoe, J., Schindler, C., Improta, T., Darnell, J. E., Jr., Stark, G. R. & Kerr, I. M. (1993) *EMBO J.* **12**, 4221–4228.
4. Sadowski, H. B., Shuai, K., Darnell, J. E., Jr., & Gilman, M. Z. (1993) *Science* **261**, 1739–1744.
5. Larner, A. C., David, M., Feldman, G. M., Igarashi, K.-I., Hackett, R. H., Webb, D. S. A., Sweitzer, S. M., Petricon, E. F., III, & Finbloom, D. S. (1993) *Science* **261**, 1730–1733.
6. Silvennoinen, O., Schindler, C., Schlessinger, J. & Levy, D. E. (1993) *Science* **261**, 1736–1739.
7. Ruff-Jamison, S., Chen, K. & Cohen, S. (1993) *Science* **261**, 1733–1736.
8. Fu, X.-Y. & Zhang, J.-J. (1993) *Cell* **74**, 1135–1145.
9. Bonni, A., Frank, D. A., Schindler, C. & Greenberg, M. E. (1993) *Science* **262**, 1575–1579.
10. Kotanides, H. & Reich, N. C. (1993) *Science* **262**, 1265–1267.
11. Zhong, Z., Wen, Z. & Darnell, J. E., Jr. (1994) *Science*, in press.
12. Schindler, C., Fu, X.-Y., Improta, T., Aebersold, R. & Darnell, J. E., Jr. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7836–7839.
13. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
14. Kawasaki, E. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic, San Diego).
15. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
16. Harlow, E. & Lane, D. (1988) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
17. Genetics Computer Group (1991) *Program Manual for the GCG Package* (Genetics Computer Group, Madison, WI), Version 7.
18. Koch, C. A., Anderson, D., Morgan, M. F., Ellis, C. & Pawson, T. (1991) *Science* **252**, 668–674.
19. Kozak, M. (1986) *Adv. Virus Res.* **31**, 229–292.
20. Ren, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. (1992) *Science* **259**, 1157–1161.
21. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D. & Kuriyan, J. (1993) *Cell* **72**, 779–790.
22. Shuai, K., Ziemiecki, A., Wilks, A. F., Harpur, A. G., Sadowski, H. B., Gilman, M. Z. & Darnell, J. E., Jr. (1993) *Nature (London)* **366**, 580–583.
23. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R. & Soltoff, S. (1991) *Cell* **64**, 281–302.
24. Levy, D. & Darnell, J. E., Jr. (1990) *New Biol.* **2**, 923–928.
25. Lew, D., Decker, T., Strehlow, I. & Darnell, J. E., Jr. (1991) *Mol. Cell. Biol.* **11**, 182–191.
26. Pearse, R. N., Feinman, R., Shuai, K., Darnell, J. E., Jr., & Ravetch, J. V. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4314–4318.
27. Kanno, Y., Kozak, C. A., Schindler, C., Driggers, P. H., Ennist, D. L., Gleason, S. L., Darnell, J. E., Jr., & Ozato, K. (1993) *Mol. Cell. Biol.* **13**, 3951–3963.
28. Darnell, J. E., Jr., Stark, G. R. & Kerr, I. M. (1994) *Science* **263**, in press.
29. Yamamoto, K., Quelle, F. W., Thierfelder, W. E., Kreider, B. L., Gilbert, D. J., Jenkins, N. A., Copeland, N. G., Silvennoinen, O., & Ihle, J. N. (1994) *Mol. Cell. Biol.* **14**, in press.