*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 7836–7839, August 1992
Biochemistry

# Proteins of transcription factor ISGF-3: One gene encodes the 91- and 84-kDa ISGF-3 proteins that are activated by interferon α

CHRIS SCHINDLER*, XIN-YUAN FU*, TERESA IMPROTA*, RUEDI AEBERSOLD†, AND JAMES E. DARNELL, JR.*

*The Rockefeller University, New York, NY 10021; and †Biomedical Research Centre and Department of Biochemistry, University of British Columbia, Vancouver, BC V6T 1W5, Canada

*Contributed by James E. Darnell, Jr., March 6, 1992*

ABSTRACT **ISGF-3 is an interferon-dependent positive-acting transcription factor that is cytoplasmically activated, possibly through direct interaction with the interferon receptor. The factor has been purified, its component proteins have been separated, and its peptide sequences have been obtained. From the sequences, degenerate oligonucleotide probes were constructed to screen for cDNA clones. Sequencing of the selected clones shows that the 91- and 84-kDa components represent two forms of a previously unknown (to our knowledge) protein. Several antibodies raised against these proteins prove that they indeed do encode protein components of ISGF-3. This work provides reagents to explore the modification of this cytoplasmically activated transcription factor.**

The attachment of interferon α (IFN-α) to its specific cell surface receptor activates the transcription of a limited set of genes, termed the IFN-stimulated genes (1–3). Agents that affect second messenger levels do not activate transcription of these genes, leading us to propose that protein–protein interactions in the cytoplasm beginning at the IFN receptor might act directly in transmitting the signal generated by receptor occupation to the nucleus (4). To test this hypothesis, we began our experiments in the nucleus at the activated genes. We identified a DNA element, the IFN stimulation response element (ISRE), and a cognate transcription factor, ISGF-3, whose activation paralleled the transcriptional activation pattern of the IFN-stimulation genes in cells treated with IFN-α (5). The observations that the proteins in ISGF-3 preexist in untreated cells, are promptly activated in an IFN-α-dependent fashion in the cell cytoplasm, and are subsequently translocated to the nucleus thus suggest that these proteins are a possible specific link between an occupied receptor and a limited set of genes (4, 6).

Partial purification of ISGF-3 followed by recovery of the purified proteins from a specific DNA–protein complex revealed that the complete complex was made up of four proteins (7). A 48-kDa protein termed ISGF-3γ, because pretreatment of HeLa cells with IFN-γ increased its presence, binds DNA weakly on its own (7–9). In combination with the IFN-α-activated proteins, termed collectively the ISGF-3α proteins, the ISGF-3γ forms a complex that binds the ISRE with a 50-fold higher affinity (8). The ISGF-3α proteins comprise a set of polypeptides of 113, 91, and 84 kDa. All of the ISGF-3 components initially reside in the cell cytoplasm (6, 10). However, after only about 5 min of IFN-α treatment, the active complex is found in the cell nucleus (6).

To understand the mechanism of cytoplasmic activation of the ISGF-3α proteins, their transport to the nucleus, and their interaction with ISGF-3γ, we have purified the factor in sufficient quantity to obtain peptide sequence from each protein. Degenerate oligodeoxynucleotides that encode the peptides were constructed and used in a combination of cDNA library screening and PCR amplification of cDNA products copied from mRNA to identify cDNA clones encoding each of the four proteins. In this report we present a description of the final protein preparations that allowed the cloning of cDNAs encoding all the proteins and also the primary sequence of the 91- and 84-kDa proteins,‡ which appear to arise from two differently processed RNA products derived from one gene. The observation that antisera against portions of the 84- and 91-kDa proteins bind specifically to the ISGF-3 DNA binding factor (detected by the electrophoretic-mobility-shift assay with cell extracts) indicates that these cloned proteins are indeed part of ISGF-3. The availability of the cDNA and the proteins they encode provides the necessary material to understand how the bound IFN-α receptor causes immediate cytoplasmic activation of the ISGF-3 protein complex. In separate reports we describe the cloning of the ISGF-3γ DNA binding protein (11) and the 113-kDa ISGF3-α protein (12) and show that the ISGF-3α proteins constitute a protein family (12).

## MATERIALS AND METHODS

**Protein Purification and Partial Peptide Analysis.** Nuclear extracts were made from superinduced HeLa cells (9) and chromatographed as described (7) on phosphocellulose P-11, heparin-agarose (Sigma), DNA-cellulose (Boehringer Mannheim; flow-through was collected after the material was adjusted to 0.28 M KCl and 0.5% Nonidet P-40), two successive rounds of ISRE oligonucleotide affinity column (1.8-ml column; material was eluted with a linear gradient of 0.05–1.0 M KCl), a point-mutant-ISRE-oligonucleotide affinity column (flow-through was collected after the material was adjusted to 0.28 M KCl), and a final round on the ISRE oligonucleotide column (material was eluted in a linear 0.05–1.0 M NaCl gradient adjusted to 0.05% Nonidet P-40). Column fractions containing ISGF-3 were subsequently examined for purity by SDS/PAGE and silver staining and were pooled appropriately. The pooled fractions were concentrated by a Centricon-10 (Amicon). The pools of fractions from preparations 1 and 2 were combined and electrophoresed on a 10-cm-wide 1.5-mm-thick 7.5% polyacrylamide gel. The proteins were electroblotted to nitrocellulose for 12 h at 20 V in 12.5% (vol/vol) MeOH/25 mM Tris/190 mM glycine. The membrane was stained with 0.1% ponceau red (in 1% acetic acid), and the bands of 113, 91, 84, and 48 kDa were excised and digested with trypsin, and peptides were sequenced (13, 14). The recovered peptide sequences for the 91- and 84-kDa proteins are given in Fig. 3. The following degenerate oligonucleotides were designed based on the peptide sequences t19, t13b, and t27 (where forward and reverse complements are denoted by F and R): 19F,

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: IFN, interferon; ISRE, IFN stimulation response element.

‡The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M97935 and M97936).

AAYGTIGAYCARYTNAAYATG; 13bR, RTCDATRTTNGRGTANAR; 27R, GTAYAAITYRAYCAGNGYAA (where Y is T or C, R is A or G, and D is T, G, or A).

**Northern Blot Analysis.** Cytoplasmic RNA (20 $\mu$g; 0.5% Nonidet P-40 lysate) of IFN-$\alpha$-treated (6 h) HeLa RNA was fractionated in a 1% agarose/6% formaldehyde gel (in 20 mM Mops/5 mM NaOAc/1 mM EDTA, pH 7.0) for 4.5 h at 125 V. The RNA was transferred in 20× standard saline citrate to Hybond-N (Amersham), UV-crosslinked, and hybridized with the indicated probes ($1 \times 10^6$ cpm/ml; $1.5 \times 10^8$ cpm/$\mu$g).

**Antibody Preparation.** Antibodies a42, a55, and a57 were prepared by injecting ≈500 mg of a fusion protein prepared in *Escherichia coli* carrying the GEX-3X vector (15). The following fusion proteins were used: a42, amino acids 598–705; a55, amino acids 2–66; a57, amino acids 704–739 (the terminal 38 amino acids in the 91-kDa protein that are not present in the 84-kDa protein). Rabbits were bled after the second booster injection and serum was prepared.

For Western blots, highly purified ISGF-3 was separated on an SDS/7% polyacrylamide gel and electroblotted to nitrocellulose. The filter was incubated in blocking buffer (Blotto = 2.5% nonfat dry milk/10 mM Tris, pH 8/150 mM NaCl/0.05% Tween 20) cut into strips, and probed with specific antiserum and preimmune antiserum diluted 1:500. The immune complexes were visualized with the aid of an ECL kit (Amersham).

Shift analyses were performed as described (5, 6) in a 4.5% polyacrylamide gel.

## RESULTS

To purify adequate amounts of ISGF-3 for peptide sequence analysis, HeLa cell nuclear extracts were prepared from batches of up to $2 \times 10^{10}$ cells treated overnight (16–18 h) with IFN-$\gamma$ (0.5 ng/ml) and 45 min with IFN-$\alpha$ (500 units/ml). The steps used in the large-scale purification were modified slightly from those described above in the identification of the four ISGF-3 proteins (7).

The final ISRE oligonucleotide affinity selection yielded material with the pattern after SDS/PAGE shown in Fig. 1 *Left*. This gel represented ≈1.5% of the available material purified from >200 liters of appropriately treated HeLa cells. Although bands of 113, 91, 84, and 48 kDa were clearly prominent in the final purified preparation (Fig. 1 *Right*), there were also two prominent contaminents of ≈118 and 70 kDa and a few of other contaminents in lower amounts. [We have shown by amino acid sequencing that the contaminents at 86 and 70 kDa are the *ku* antigen, a widely distributed protein that binds DNA termini. However, in the specific ISGF-3–ISRE complex, there is no *ku* antigen and we therefore assign it no role in IFN-dependent transcriptional stimulation (13)]. Since we could accurately mark the mobility of proteins at 113, 91, 84, and 48 kDa by comparison with the partially purified proteins characterized in previous experiments (7), we did not attempt further purification at this stage. The total purified sample from 200 liters of HeLa cells was loaded onto one gel, subjected to electrophoresis, transferred to nitrocellulose, and stained with ponceau red. The protein bands at 113, 84, 91, and 48 kDa were separately excised and subjected to peptide analysis as described (14). Released peptides were collected, separated by HPLC, and analyzed for sequence content by automated Edman degradation analysis.

In this report we describe the use of the peptide sequence data for three of four peptides from the 91-kDa protein and a single peptide derived from the 84-kDa protein.

The recovered peptide sequences (t19, t13b, and t27) are shown underlined in Fig. 2 and the oligonucleotides constructed from them are 19F, 13bR, and 27R, respectively.

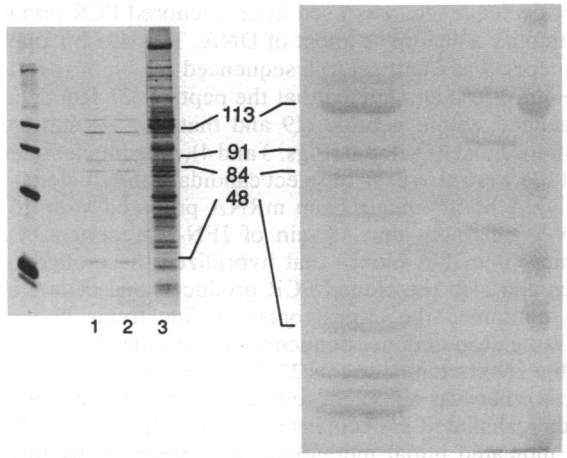


FIG. 1. Purification of ISGF-3. (*Left*) Purification of ISGF-3 showing the polypeptides present after the first oligonucleotide affinity column (lane 3) and two preparations after the final chromatography step (lanes 1 and 2). The left most lane contains protein size markers (high molecular weight markers; Sigma). ISGF-3 component proteins are indicated at 113, 91, 84, and 48 kDa (8, 9). (*Right*) Purified ISGF-3 from 2–3 × $10^{11}$ cells was electroblotted to nitrocellulose after preparations 1 and 2 (*Left*, lanes 1 and 2) had been pooled and separated on an SDS/7.5% polyacrylamide gel (lane to the left). ISGF-3 component proteins are indicated. The two lanes to the right contain protein markers (high molecular weight and prestained markers; Sigma).

When oligonucleotides 19F and 13bR were used to prime synthesis from a HeLa cell cDNA library, a PCR product of 475 base pairs (bp) was generated. When this product was cloned and sequenced, it encoded the 13a peptide internally. Oligonucleotide 27R, derived from the only available 84-kDa

```
  1 MSQWYELQQLDSKFLEQVHQLYDDSFPMEIRQYLAQWLEKQDWEHAANDV
 51 SFATIRFHDLLSQLDDQYSRFSLENNFLLQHNIRKSKRNLQDNFQEDPIQ
101 MSMIIYSCLKEERKILENAQRFNQAQSGNIQSTVMLDKQKELDSKVRNVK
151 DKVMCIEHEIKSLEDLQDEYDFKCKTLQNREHETNGVAKSDQKQEQLLLK
201 KMYLMLDNKRKEVVHKIIELLNVTELTQNALINDELVEWKRRQQSACIGG
251 PPNACLDQLQQVRQQLKKLEELEQKYTYEHDPITKNKQVLWDRTFSLFQQ
301 LIQSSFVVERQPCMPTHPQRPLVLKTGVQFTVKLRLLVKLQELNYNLKVK
351 VLFDKDVNERNTVKGFRKFNILGTHTKVMNMEESTNGSLAAEFRHLQLKE
401 QKNAGTRTNEGPLIVTEELHSLSFETQLCQPGLVIDLETTSLPVVVISNV
451 SQLPSGWASILWYNMLVAEPRNLSFFLTPPCARWAQLSEVLSWQFSSVTK
                t27
501 RGLNVDQLNMLGEKLLGPNASPDGLIPWTRFCKENINDKNFPFWLWIESI
                t19
551 LELIKKHLLPLWNDGCIMGFISKERERALLKDQQPGTFLLRFSESSREGA
601 ITFTWVERSQNGGEPDFHAVEPYTKKELSAVTFPDIIRNYKVMAAENIPE
                                                t13a
651 NPLKYLYPNIDKDHAFGKYYSRPKEAPEPMELDGPKGTGYIKTELISVSE
          t13b
701 VHPSRLQTTDNLLPMSPEEFDEVSRIVGSVEFDSMMNTV
```

FIG. 2. Protein sequence of the 91- and 84-kDa proteins of ISGF-3. One-letter amino acid code is shown for the open reading frame from clone E4 (encoding the 91-kDa protein). The 84-kDa protein, encoded by a different cDNA (E3), has the identical sequence but terminates after amino acid 701, as indicated by the arrow. Tryptic peptides t19, t13a, and t13b from the 91-kDa protein are indicated. Peptide t27, the sole recovered tryptic peptide from the 84-kDa protein, was wholly contained within peptide t19, as indicated.

peptide sequence, was used in an anchored PCR procedure to amplify a 405-bp segment of DNA. This 405-bp amplified sequence was identical to a sequenced region of the 91-kDa protein. We then realized that the peptide t27 sequence was contained within peptide t19 and that the 91- and 84-kDa proteins must be related (Figs. 3 and 4). Oligonucleotides 19F and 13a were also used to select candidate cDNA clones from a cDNA library made from mRNA prepared after 16 h of IFN-$\gamma$ treatment and 45 min of IFN-$\alpha$ treatment. Of the numerous cDNA clones that hybridized these oligonucleotides and also the cloned PCR products, one cDNA clone, E4, contained the largest open reading frame flanked by in-frame stop codons. Sequences of peptides t19, t13a, and t13b were contained in this 2217-bp open reading frame (see Fig. 2) that was sufficient to encode a protein of 739 amino acids (calculated molecular mass of 86 kDa). The codon for the indicated initial methionine was preceded by three in-frame stop codons. This coding capacity has been confirmed by translating *in vitro* an RNA copy of the E4 clone to yield a product with the nominal size of 86 kDa, somewhat shorter than the *in vitro*-purified 91-kDa protein (data not shown). Perhaps this result indicates posttranslational modification of the protein in the cell.

A second class of clones was also identified (see Fig. 3). E3, the prototype of this class, was identical to E4 from the 5′ end to bp 2286 (amino acid 701) at which point the sequences diverged completely. Both cDNAs terminated with a poly(A) tail. Primer-extension analysis suggested another ≈150 bp were missing from the 5′ end of both mRNAs. DNA probes were made from the clones representing both common and unique sequences for use in Northern blot analyses (see Fig. 3). Probes from regions common to E3 and E4 hybridized to two RNA species of ≈3.1 kilobases (kb) and 4.4 kb. Several probes derived from the 3′ noncoding end of E4 that were unique to E4 hybridized with only the larger RNA species. A labeled DNA probe from the unique 3′ noncoding end of E3 hybridized with only the smaller RNA species.

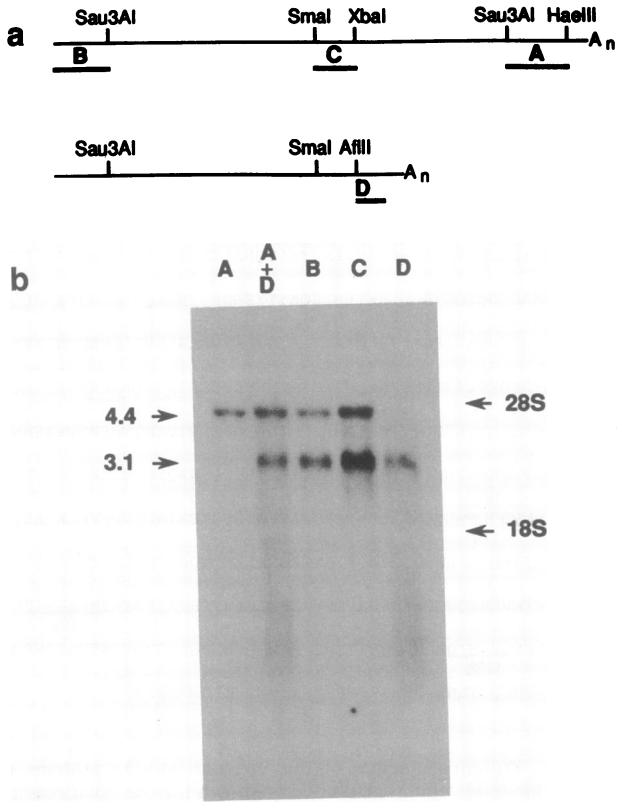


FIG. 3. Northern blot analysis. (*a*) Restriction maps for cDNA clones E4 (top map) and E3 (bottom map) show DNA fragments that were radiolabeled as probes (probes A–D). (*b*) Northern blots of cytoplasmic HeLa RNA hybridized with the indicated probes. The 4.4- and 3.1-kb species and the 28S and 18S rRNA bands are indicated.

Review of the sequence at the site of 3′ discontinuity between E3 and E4 suggested that the shorter mRNA results from choice of a different poly(A) site and 3′ exon that begins at bp 2286. The last two nucleotides before the change are GT followed by GT in E3, in line with the consensus nucleotides at an exon–intron junction. Since the open reading frame of E4 extends to bp 2401, it encodes a protein that is 38 amino acids longer than the protein encoded by E3 but is otherwise identical (the open reading frame of the shorter protein is 82 kDa).

Since there is no direct assay for the activity of the 91- or 84-kDa protein, we needed an independent method to determine whether the cDNA clones we had isolated did indeed encode proteins that are part of ISGF-3. For this purpose antibodies were initially raised against the sequence from amino acid 598 to amino acid 705 (see Fig. 2) by expressing this peptide in the pGEX-3X vector (15) as a bacterial fusion protein. This antiserum (a42) specifically recognized the 91- and 84-kDa proteins in both crude extracts and purified ISGF-3 (Fig. 4*a*). More importantly, this antiserum specifically shifted the ISGF-3 band in a mobility-shift assay using the labeled ISRE oligonucleotide (Fig. 4*b*), confirming that the isolated cDNA clones (E4 and E3) for the 91- and 84-kDa proteins represent a component of ISGF-3.

Additional antisera were raised against the amino terminus and carboxyl terminus of the protein encoded by E4. The amino-terminal 66 amino acids that are common to both proteins and the unique carboxyl-terminal 36 amino acids encoded only by the larger mRNA were expressed as fusion proteins in pGEX-3X for immunization of rabbits. Western blot analysis with highly purified ISGF-3 demonstrated that the amino-terminal antibody (a55) recognized both the 91- and 84-kDa proteins, as expected. However, the other antibody (a57) recognized only the 91-kDa protein, confirming our assumption that the larger mRNA (4.4 kb) and larger cDNA encode the 91-kDa protein whereas the shorter mRNA (3.1 kb) and cDNA encode the 84-kDa protein (Fig. 4*a*).

Thus we conclude that we secured accurate peptide sequences from ISGF-3 protein components and thus correctly identified cDNA clones encoding the 91- and 84-kDa components of ISGF-3.

## DISCUSSION

In dealing with transcription factors composed of multiple proteins, it is difficult to obtain preparations with no other contaminating proteins. In earlier work (7), it appeared clear that the 113-, 91-, and 84-kDa proteins were regularly associated with the 48-kDa DNA binding subunit to constitute ISGF-3. Nevertheless, we could not be certain that the partial peptide sequences we had obtained were in fact related to ISGF-3. Therefore, we were gratified when an antiserum raised against the protein encoded by the cloned 91/84-kDa protein sequence did indeed react with the ISGF-3 DNA binding complex and that a single antiserum was capable of recognizing both native 91- and 84-kDa polypeptides. In addition the antisera and the sequence data of the two clones encoding the 91- and 84-kDa proteins clearly show that two of the four ISGF-3 proteins are from a single gene. This simplifies somewhat the problem of further studying the activation and interactions of the four protein components of ISGF-3. Finally, as we discuss in detail elsewhere (12), the sequence of the 113-kDa protein and the sequences of the 91- and 84-kDa proteins are encoded by genes in the same family.

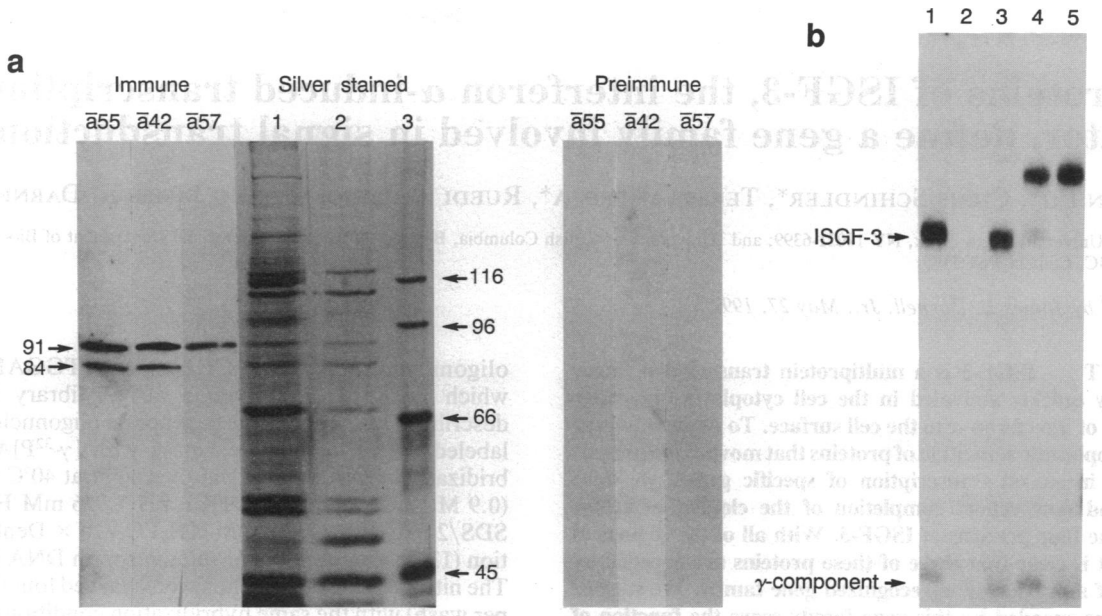


FIG. 4. Western blot and antibody-shift analysis. (*a*) Highly purified ISGF-3, fractionated on an SDS/7.0% polyacrylamide gel, was probed with antibodies a42 (amino acids 598–705), a55 (amino acids 2–66), and a57 (amino acids 704–739) on a Western blot. The silver-stained part of the gel (lanes 1–3) illustrates the location of the ISGF-3 component proteins and the purity of the material used in the Western blot. Lanes: 1, silver stain of protein sample used in all Western blot experiments (immune and preimmune); 2, material of purity equal to that shown in Fig. 1, for clearer identification of the ISGF-3 proteins; 3, protein size markers as indicated (kDa). (*b*) Antibody interference with the mobility of the ISGF-3 shift complex. Lanes: 1, complete ISGF-3 and the free ISGF-3γ component that shifts with partially purified ISGF-3; 2, competition with a 100-fold excess of unlabeled ISRE oligonucleotide; 3, shift complex after the addition of 1 μl of preimmune serum to a 12.5-μl shift reaction mixture; 4 or 5, shift complex after the addition of 1 μl of a 1:10 dilution or 1 μl of undiluted a42 antiserum to a 12.5-μl shift reaction mixture, respectively.

With the antisera available from this work and from the work on the 113-kDa gene, it should be possible to explore events that may activate the three proteins after IFN induction. For example, earlier work suggests that the 113-, 91-, and 84-kDa proteins may form a physical complex (8). In addition, staurosporine, an inhibitor of various protein kinases (16, 17), blocks the IFN-α activation of transcription and blocks ISGF-3 activation. Among the first experiments to be done with the antibodies should be the examination of the IFN-α-dependent determination of complex formation and of any protein phosphorylation.

As we have emphasized above, the cytoplasmic activation of ISGF-3 is a very prompt event specifically requiring IFN-α occupation of its receptor. Thus we have reason to believe that the ISGF proteins may interact directly with the ligand-occupied receptor to become activated.

We thank Anthony Canova and Edward Bures for excellent technical assistance and Dr. Uli Nielsch for help with Northern blot analysis. This work was supported by grants from the American Cancer Society (MV-424U) and the National Cancer Institute of Canada. X.-Y.F., now at Mt. Sinai School of Medicine, was a Fellow of the Cancer Research Institute/Fan Fox and Leslie R. Samuels Foundation (New York).

1. Larner, A. C., Jonak, G., Cheng, Y.-S., Korant, B., Knight, E. & Darnell, J. E., Jr. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6733–6737.
2. Larner, A. C., Chaudhuri, A. & Darnell, J. E. (1986) *J. Biol. Chem.* **261**, 453–459.
3. Friedman, R. L., Manly, S. P., McMahon, M., Kerr, I. M. & Stark, G. R. (1984) *Cell* **38**, 745–755.
4. Levy, D. & Darnell, J. E. (1991) *New Biologist* **2**, 923–928.
5. Levy, D. E., Kessler, D. S., Pine, R., Reich, N. & Darnell, J. E. (1988) *Genes Dev.* **2**, 383–392.
6. Levy, D. E., Kessler, D. S., Pine, R. & Darnell, J. E. (1989) *Genes Dev.* **3**, 1362–1371.
7. Fu, X., Kessler, D. S., Veals, S. A., Levy, D. E. & Darnell, J. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8555–8559.
8. Kessler, D. S., Veals, S. A., Fu, X.-Y. & Levy, D. E. (1990) *Genes Dev.* **4**, 1753–1765.
9. Levy, D. E., Lew, D. J., Decker, T., Kessler, D. S. & Darnell, J. E. (1990) *EMBO J.* **9**, 1105–1111.
10. Dale, T. C., Iman, A. M. A., Kerr, I. M. & Stark, G. R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1203–1207.
11. Veals, S. A., Schindler, C. W., Fu, X.-Y., Leonard, D., Darnell, J. E. & Levy, D. E. (1992) *Mol. Cell. Biol.* **12**, in press.
12. Fu, X.-Y., Schindler, C., Improta, T., Aebersold, R. & Darnell, J. E., *Science*, in press.
13. Wedrychowski, A., Seong, D., Paslidis, N., Johnson, E., Howard, O. M. Z., Sims, S., Talpaz, M., Kantarjian, H., Hester, J., Turpin, J., Lopez-Berestein, G., Gutterman, J., Freireich, E. J. & Deisseroth, A. (1990) *J. Biol. Chem.* **265**, 21433–21440.
14. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6970–6974.
15. Smith, D. B. & Johnson, K. S. (1988) *Gene* **67**, 31–40.
16. Reich, N. & Pfeffer, L. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8761–8765.
17. Kessler, D. S. & Levy, D. (1991) *J. Biol. Chem.* **266**, 23471–23476.