The genomic structure of the STAT genes: multiple exons in coincident sites in Stat1 and Stat2

Riqiang Yan, Sajjad Qureshi, Zhong Zhong, Zilong Wen and James E. Darnell, Jr*

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

Received September 28, 1994; Revised and Accepted December 15, 1994

GenBank accession nos U18662–U18671 (incl.)

ABSTRACT

The genomic structure of Stat2 has been determined and compared with a large portion of the Stat1 gene. There are 24 exons in the Stat2 gene and a matching number in very similar positions in the Stat1 gene. Thus a very complicated genomic structure was presumably duplicated and has been closely maintained throughout evolution.

INTRODUCTION

STAT proteins have the dual function of signal transduction and activation of transcription (1). These proteins are activated by phosphorylation on tyrosine in response to different ligands after which they form homodimers or heterodimers (2,3) that translocate to the cell nucleus where they either directly bind to DNA or act together with other DNA binding proteins in multiprotein transcription complexes to direct transcription (1). The first of these proteins to be described, Stat1, is activated by a number of different ligands including IFN- α , IFN- γ , EGF, PDGF, IL-6 and others (4–8). The same tyrosine residue is activated at least by IFN- α , IFN- γ and EGF (9). Stat2, in contrast, is activated by IFN- α but not by IFN- γ or any of the other ligands mentioned above. Stat3 is known to be activated by EGF, IL-6, LIF and perhaps other ligands but is not activated by IFN- γ (3). Stat4, another family member, is present in high concentrations in the testes but has not yet been found in a phosphorylated form in cells (3,10).

The STAT proteins differ in the DNA sites to which they bind. Stat1 homodimer binds to a site termed GAS, first defined as required for IFN- γ induction (1). Variations on this site are also used in response to IL-6 (11,12), PDGF (13) and also other ligands. The multiprotein complex ISGF-3, which is composed of Stat1, Stat2 and p48, binds to a 15 bp element designated the ISRE (interferon stimulated response element) that is different from the GAS element (14,15). The Stat1 gene produces two different spliced forms (16) and Stat3 also exists as two or more forms (10).

As a prelude to an improved understanding of this group of proteins and the domains they contain, we have begun examination of genomic clones of STAT genes. We report here the complete genomic sequence and characterization of the promoter region and exonic structure of the human Stat2 gene: it contains 24 exons and has an imperfect ISRE and no demonstrably active

GAS sites, consistent with its weak transcriptional induction by IFN- α and IFN- γ . We have compared the Stat2 gene structure to a considerable portion of the Stat1 gene which also has many exons at homologous regions. Thus, not only is there considerable amino acid conservation throughout a 700 amino acid coding region of these genes (17), the genomic structure appears to have been largely conserved.

MATERIALS AND METHODS

Genomic cloning of human Stat1 and Stat2 genes and DNA sequence analysis

A human λ EMBL4 genomic library was used for primary screening with Stat1 and Stat2 cDNA probes by the conventional cloning method (18). A human λ EMBL3 genomic library (Stratagene, La Jolla, CA) was used to screen for additional Stat1 and Stat2 genomic clones. DNA inserts were excised from positive recombinant phages and subcloned into pBluescript SK II (–) (Stratagene, La Jolla, CA). The nucleotide sequences were determined by the chain termination method (19), and the sequencing was performed by either synthesizing primers from previously determined sequences or generating nested sets of deletion mutants.

Isolation of RNA and primer extension analysis

HeLa cells were harvested with or without IFN treatment. Interferon treatment consisted of an overnight IFN- γ treatment (5 ng/ml) and a 45 min IFN- α treatment. This protocol referred to as ‘superinduction’ elevates ISGF-3 response ~10-fold. Isolation of cytoplasmic RNA was performed as described (18). RNA transcription initiation site for Stat2 gene was analyzed by primer extension (18). Primer 1 which covered the exon 1 and exon 2 was complementary to 28–61 as described (17) and had the sequence 5′-CTGCAGCATTTCCTCCACTGCGCCATTGGGCTCTG-3′. Primer 2 corresponded to 59–92 and had the sequence 5′-GCAGCTGATCCTGAAAGGGGCTGTCAAGAT TCTG-3′ (Fu *et al.*, 1992).

Construction of plasmids

A promoterless luciferase reporter construct (pZLUC) which has unique sites for *Bam*HI, *Sma*I, *Kpn*I, and *Hind*III was a gift from William Chen. A 2.2 kb fragment containing the first exon and the upstream sequences was fused to the pZLUC at the *Bam*HI and

* To whom correspondence should be addressed

SmaI sites. The restriction enzymes *BglIII* (at about -1500), *NcoI* (at -1470 and -406), *PflMI* (at -301), were used to make double digestions with *BamHI* followed by religation using Klenow to fill the gap. This full length construct was also used to produce the rest of 5' end or 3' end deletion constructs by *ExoIII* nuclease digestion (18).

DNA transient transfection and luciferase activity assay

The WI-38 human lung diploid cells were used for DNA transfection by the calcium phosphate method with 4 µg of each construct. The transfection efficiency was monitored by inclusion of 4 µg of a β-galactosidase expression plasmid (pCMVβ, Promega). For basal promoter activity assay, the cells were harvested after transfection for 72 h and luciferase activity was determined according to the standard luciferase assay procedure (Promega). The same constructs were also used to determine the transcriptional activation after IFN treatment. For IFN-α induction, the WI-38 cells were either treated with IFN-α for 12 h or 1 h before harvest (12 h treatment yielded slightly higher activity). For IFN-γ induction, the transfected cells were treated with IFN-γ for 4 h before harvest.

Gel shift assay

For the detection of a Stat2 ISRE, HeLa cells were harvested after IFN superinduction (18 h IFN-γ and 45 min IFN-α treatment). Nuclear extracts were prepared as described (14). The Stat2 ISRE sequence from -573 to -359 is shown in bold:

5'-GATCCAAGCTAGTTTCGGTTCCG....-3'
3'-.....GTTTCGATCAAAGCCAAGGCCTAG

The oligonucleotides were labeled with [α-³²P]-dNTP and Klenow. The gel shift reactions were performed at room temperature as described (14). For blocking the formation of ISGF-3 complex, 1 µl of 1:10 dilution antiserum against Stat1 C-terminal portion (16), Stat2 C-terminal portion (17) or p48 (20) was preincubated with the nuclear extract in the gel shift reaction buffer for 20 min before the addition of the labeled Stat2 ISRE probe. A competitor sequence derived from -349 to -328 of stat2 gene, containing a GAS-like site (consensus sequence for GAS is TTNCNNNA, see Darnell's *Science review*), was used for non-specific competition. The GAS like sequence is:

5'-GATCCGCGGGTTCTCGAAAATCAG....-3'
3'-.....GCGCCCAAGAGCTTTTGTCTCCTAG-5'

RESULTS

Genomic maps of the Stat2 gene

A λ EMBL3 human DNA library was screened with probes representing the carboxyl and amino terminal sections of the human Stat2 cDNA. A single λ clone was obtained that hybridized both the 5' and 3' probes of the Stat2 cDNA. The Stat2 λ clone 1 (Fig. 1) was completely sequenced mainly using oligonucleotide primers from the known cDNA sequence. The genomic sequence (see Fig. 1 and Table 1) contained 23 exons the sequence of which encoded the entire cDNA that was previously described and extended seven bases upstream of the putative ATG translation start site (17). The largest number of amino acids encoded by any exon was 104 and the shortest was seven amino acids which is unusually short compared to other known amino

Figure 1. Partial restriction map of the Stat2 gene and its intron-exon organization. The restriction enzymes indicated on the map are B (*BamHI*), E (*EcoRI*), K (*KpnI*) and X (*XbaI*). An open box represents an exon and the horizontal line indicates the intron.

Table 1. Exons/introns organization of STAT2 gene

Exon Number	Exon	Numbers of amino acid	Intron and its size (bases)
1	1-AAGTCG.....AATCAG-50	0	gtacgg.....acgcag (3509)
2	51-AGCCCA.....GAACGT-188	432/3	gtgagg.....tcccag (154)
3	189-GCAGGA.....ATTTCAG-342	391/3	gtactt.....tgtag (330)
4	343-CCCTTT.....CAATTG-438	32	gtgagg.....tttcag (175)
5	439-GAACAA.....ATGGAG-528	30	gttagt.....ctgcag (91)
6	529-AAGCTG.....CCAAG-604	251/3	gtagga.....atccag (420)
7	605-GGAAGA.....AGAAAG-690	29	gtggga.....tcccag (165)
8	691-GAGGTG.....GACATG-839	502/3	gtgaga.....accag (3075)
9	840-GTTCAC.....CCACAG-998	53	gtctag.....tcttcg (90)
10	999-AGCCTT.....AACAAG-1091	302/3	gttggt.....ccccag (209)
11	1092-GCTGCT.....TGACAG-1150	191/3	gtaaat.....tttcag (337)
12	1151-GAATCC.....ACAAGG-1171	7	gtaggt.....gaacag (188)
13	1172-CTTCCG.....TACCTG-1262	311/3	gtaaga.....tcacag (463)
14	1263-ACTCTG.....AATAAG-1309	16	gtgaga.....tcccag (79)
15	1310-AGGGGC.....CTGAAA-1396	28	gtgagt.....ccacag (166)
16	1397-ACGGAC.....CTTCAG-1495	33	gtaggg.....ctgcag (103)
17	1496-AACCAG.....TGTTTC-1531	451/3	gtacag.....ccacag (116)
18	1532-GGCAGA.....ACTAAG-1584	182/3	gtaact.....ctccag (127)
19	1585-CGAGAG.....TGATGG-1726	312/3	gtaagg.....ctccag (1595)
20	1727-ACGCAT.....ATGATG-1863	451/3	gtagct.....ctccag (192)
21	1864-ACAAGG.....AGAAAG-2046	611/3	gtggga.....tcacag (240)
22	2047-TTAATC.....AATAGA-2105	192/3	gtgaga.....atccag (2009)
23	2106-CAGGTG.....TGGAAG-2406	1031/3	gtaagt.....acccag (282)
24	2407-TCTTCA.....CAATCT-3271	462/3	(>472)

acid encoding exons that are not at the terminus of a protein (21); (there are a number of cases where AUG is the only coding sequence in the first exon). The introns in several sites were also unusually short, approximating the putative minimum length required for an intron to function [for example, 79 bases between exons 14 and 15; Table 1; (21)].

The original cDNA sequence of Stat2 contained additional bases at its 5' end not encoded in the available genomic DNA from λ clone 1; thus a second λ clone was selected with a probe from the 5' end of the cDNA in order to attempt to locate the complete first exon and the RNA start site for Stat2. This second

-702 TCAAGATCAGCCTGGGCAACATGGCGAAACCCGCTCTACAATAAATAC
 -652 AAAAAAATTATCTGGCGGAGTTATGCACGTTGTAGTCCCACTACCTGG
 -602 GAGGCTGAGCGCGGAGAAATCACCTGAGCCTGGGAGTTCGAGGCTGCAGCG
 -552 AGCCGAGATCGGCCGCTGCATTCCAGCCTGGGTGACAGAGCGAGACCATG
 -502 TCTCAAAAAATAAAAAATTAATAAAAAATTTTTCATTACCTCAGCCCTC
 -452 CTCTTCCTATCCCAAGCGCTGAAATTCGGTCCACCCCTTCCCATGGA
 -401 GCCCTTGGCCTCTCCAGGCTCCTCAAGCT**TAATTCGGTTC**CGGGCGGTCA
 -351 CCCGCGGGTTCTCGAAATCAGCTGTTTCAGTCTTGGGCTAGTCCACTAA
 -301 TTGGACTCCTCCCTCGTAGAAAGTGCTACTTGAACCTTCTCCACCAATC
 -251 GCTGAAGCTGCAGGTGTGGTTTCGGCTCAGCTTGTCCCGCTGGCGGAG
 -201 GGCGGAGTTGGCGCGCGCCAGTGAAGCTCGCAGTCTGGGAAGGGCTTGA
 -151 CTGAATGGCAGCCAGTGTGGGGTGGCGGCTGGGAATGGGGCCGCTCCG
 -101 GACTTCCGCTGCCAACTACAAGGGGCGGGTCCGAGGGGGTTAGCCGAA
 -51 GTTGTAGGCGGGGCGCGAGGTTCTAGTACCGAGCTCATACTAGGACGG
 +1
 -1 GAAGTCGCGACCAAGCCATTTGGAGGGCGCGGGGACTGCAACCCTAATCAG +50

Figure 2. Sequence of the Stat2 promoter region. About 750 bp flanking fragment including the first exon sequence of the Stat2 gene is shown, with +1 corresponding to the putative RNA transcription site. Underlines are GC box, CAAT box and some GC rich sequences. The ISRE site is in bold.

λ clone (clone 2, Fig. 1) overlapped clone 1 but extended further in the 5' direction. The sequence of the region that hybridized the 5' end of the cDNA probe was located about 3.5 kb upstream from exon 2 with a likely splice junction that lay just downstream of the nucleotide sequence from the 5' end of the cDNA.

Analysis of the 5' end sequence of the Stat2 gene

Using two different primers within the exon 1 and exon 2 and total RNA from HeLa cells induced by an overnight IFN- γ and a 45 min IFN- α treatment (the so-called superinduction protocol for producing maximal IFN- α response) (22). This treatment produces higher levels of Stat2 mRNA (16) for primer extension study and we obtained one major band with each primer that suggested an apparent RNA start site ~50 nucleotides from the 3' end of the new exon (data not shown; repeated primer extensions yielded different non-reproducible bands as well as the one constant band mentioned here). Thus we preliminarily assigned this site as the correct start site of Stat2 (shown in Fig. 2). The region upstream of this putative start site contained no TATA box but did contain several GC boxes and a CAAT box. This region close to the putative start site, resembled that of many housekeeping genes, was highly GC rich and could include potential binding sites for the Sp1 transcription activator (23).

To test the correct assignment of the initiation site, a series of DNA constructs with luciferase as a reporter gene were prepared (outlined in Fig. 3) and were transfected into HeLa cells to test the luciferase expression activity. The construct that contained about 2.2 kb of sequence upstream of the putative start site gave the highest expression signal in transfected cells. A small reduction in signalling (2-fold) was observed upon removing upstream segments down to -301. Sequences within -406 to -301 seemed to have some minor negative regulatory roles as the construct with

Figure 3. Deletion analysis of basal promoter activity of Stat2 gene. Different lengths of the Stat2 upstream sequence were fused to a promoterless luciferase reporter vector and their luciferase activity were determined. The open box represents the Stat2 gene and the hatched box represents the luciferase reporter gene. The full-length construct that contained sequence from -2200 bp to +50 bp yielded the highest luciferase activity which was set at 100%. The relative luciferase activity shown are the values determined after normalization for transfection efficiency. The number of experiments (n) is indicated in parenthesis.

sequence to -406 repeatedly had somewhat less luciferase activity than the one with sequence to -301. Removal of -301 to -121 decreased activity ~40-fold. In addition to this evidence of required sequences between -121 and -301, there was also evidence of an initiator sequence. If the entire upstream sequence from +3 to -2200 was fused to the reporter luciferase sequence there was almost no expression, suggesting the possible presence of an activator sequence in the region indicated as the start site.

In addition to the foregoing experiments on sequences required for constitutive expression, these constructs were also tested for IFN- α induction of expression since a modest increase in run-on transcription of Stat2 has been reported after IFN- α induction (24). The -301 construct did not respond to IFN- α induction while the construct with the addition of 105 bases (to -406) did respond with a small but reproducible IFN- α induction. There was a site at -380 that resembled an ISRE (Fig. 3). A double-stranded DNA oligonucleotide from this region bound to proteins in extracts of IFN- α -treated but not in untreated cells and formed a weak but clearly detectable band that migrated in the protein of ISGF-3 (arrows). Moreover, antibodies to ISGF-3 proteins removed the specific complex (Fig. 4). The same constructs did not show obvious induction of any specific DNA binding complexes after IFN- γ treatment. Thus the putative start site region contained a weak ISRE in accord with the weak

Figure 4. Gel shift assay to determine the formation of ISGF-3 complex with putative Stat2 ISRE sequence. Nuclear extracts from mock- or superinduced HeLa cells were subject to a gel shift assay with a 32 P-labeled probe containing Stat2 ISRE site (see Materials and Methods). ISGF-3 complex is indicated with an arrow, and the bands below the ISGF-3 are not inducible after interferon treatment.

transcriptional induction of endogenous Stat2 by IFN- α and it appears that we have cloned the promoter region of Stat2 to complete its exonic structure.

Partial exonic structures of Stat1 reveal the conserved exon-intron border of Stat1 and Stat2 gene

The same λ EMBL genomic library was also screened with Stat1 cDNA probes and two different clones were recovered that hybridized the two ends of the Stat1 cDNA. The large sections of the two available λ clones were sequenced. Again a large number of exons were found that corresponded to the available cDNA sequence. The sequence found in the two λ clones provided corresponding genomic sequence to about two-thirds of the Stat1 cDNA. When the Stat2 and available Stat1 exon sequences were aligned it became clear that every exon-intron border in Stat1 (exons 1–6 and 12–17) was also represented by an analogous exon in Stat2 (Fig. 5). There is a gap (exons 5–11) in the middle of the cDNA which needs to be filled by further genomic screening and sequencing. Several attempts with two different libraries have not located this missing region as part of these λ libraries. Examination of the genomic Stat1 sequence at the 3' end of the cDNA explained the differential splicing of Stat1 α and β . There is after exon 22, that encodes amino acid 700 in both Stat1 α and Stat1 β , a choice of the next exon. The exon that appears in the mRNA encoding Stat1 β (the 84 kDa protein) contains an in frame

Figure 5. Conserved exonic organization of Stat1 and Stat2 genes. Comparison of the amino acid sequence of Stat1 and Stat2 was performed by Bestfit Program (Genetics Computer Group). Lines connect identical amino acids; dots indicate conservative changes. The exon borders are denoted with an arrow and numbers at ends of lines are amino acid sequence positions.

translational stop signal after amino acid 701 (Fig. 6). There are two exons after amino acid 700 in Stat1 α (the 91 kDa protein) one that encodes 38 amino acids that are added to amino acid 699 plus one additional exon that encodes the final amino acid before the UAG stop codon. The upstream exon [and poly(A) site] that is used to produce the Stat1 β (84 kDa) mRNA is skipped in producing the Stat1 α mRNA.

DISCUSSION

The STAT proteins must interact with many other intercellular proteins such as receptors, kinases, translocation machinery, other transcriptional proteins, including other STAT proteins, tyrosine phosphatases and probably other proteins as well. Therefore understanding the possible domains in the proteins that contribute to the various functions of the STAT proteins are important targets of research. In some proteins functional domains are evident in the exonic division of genes and this is one of the reasons we undertook this study. However, the Stat1 and 2 genes contain even more exons than the many functions anticipated for these proteins.

Figure 6. Alternative splicing in formation of 3' ends of Stat1 α and 1 β mRNAs. The genomic DNA sequence around the final intron–exon junction sites at final coding exon of Stat1 β (left) and α (right) is shown with the capital letter indicating exon sequence and lower case letter indicating intron sequence. The open box represents the exon. The hatched box is considered an exon if it is not spliced out. The horizontal line represents the intron.

Aside from proteins that have only one or a few amino acids encoded in an initial exon, the very short exons of Stat1 and Stat2 are among the shortest known. In addition, the very short intronic sequences in Stat2 are quite unusual. In general, the introns are shorter in Stat2 than they are in Stat1 which has made the cloning of the entire genomic region representing Stat1 more difficult.

We knew that Stat1 and Stat2 proteins were similar in amino acid sequence (35% identity scattered over ~700 amino acid coding region) indicating a duplicated gene at some time in evolution. In comparing Stat1 with the available region of Stat2 the genomic structure is also extremely similar; each one of the identified exons in Stat1 has an identified matching exon in Stat2 dividing the proteins in almost identical sites. These results make a powerful case that the exonic structure existed before gene duplication.

The only domains in the STATs that have been recognized by sequence comparison with previously defined proteins are the SH3 and SH2 regions of the STAT proteins. The proposed SH3 region (amino acids 504–558) is divided by three introns.

The Stat SH2 domain is divided into three exons, the same as the SH2 region of other proteins such as src, however the intron and exon borders are not correlated (25,26). The tyrosine that is phosphorylated is present in a separate exon (residue 690 in Stat2 and 701 in Stat3). There are additional regions that are highly conserved among all the Stat members such as (T/S)LPVV (V/I)ISN motif which are found in a single exon.

The existence of many exons in the two STAT proteins as well as the existence of one case of differential splicing in Stat1 raises the possibility that this family of genes might exhibit differentially spliced mRNAs leading to protein variation. It has been noted, in addition during recombinant DNA experiments manipulating the size of exons, that exons shorter than 51 nucleotides lead to frequent exon skipping and one exon in Stat2 is 21 nucleotides long and the homologous exon in Stat1 is 30 nucleotides long (27). We have made a brief search of HeLa cell mRNA and of a series of human cDNA libraries with PCR technology without yet finding differentially spliced variants of Stat2. Further attempts along these lines however seem warranted because different

transcriptional responses to different polypeptide ligands might depend on such splicing events.

A previous transcriptional analysis of the endogenous gene showed that the Stat2 gene was weakly activated upon IFN- α treatment (24). The core ISRE of Stat2 gene (AGTTTCGGTTCC) is similar to ISG15 ISRE (AGTTTCGGTTTC), but the Stat2 ISRE sequence is ~5-fold weaker in formation of the ISGF-3 complex. A weak ISRE site in Stat2 gene could provide limited positive autoregulation.

ACKNOWLEDGEMENTS

R. Yan was a Winston Fellow and effective 7/1/94 is an Aaron Diamond Fellow. S. Qureshi is a Cancer Research Institute Fellow. This work was supported by NIH grants AI32489 and AI34420.

REFERENCES

- Darnell, J. E., Jr., Kerr, I. M. and Stark, G. M. (1994) *Science* **264**, 1415–1421.
- Shuai, K., Horvath, C. M., Tsai-Huang, L. H., Qureshi, S., Cowburn, D. and Darnell, J. E., Jr. (1994) *Cell* **76**, 821–828.
- Zhong, Z., Wen, Z. and Darnell, J. E., Jr. (1994) *Science* **264**, 95–98.
- Lerner, A. C., David, M., Feldman, G. M., Igarashi, K.-I., Hackett, R. H., Webb, D. S. A., Sweitzer, S. M., Petricola III, E. F. and Finbloom, D. S. (1993) *Science* **261**, 1730–1733.
- Ruff-Jamison, S., Chen, K. and Cohen, S. (1993) *Science* **261**, 1733–1736.
- Silvennoinen, O., Schindler, C., Schlessinger, J. and Levy, D. E. (1993) *Science* **261**, 1736–1739.
- Sadowski, H. B., Shuai, K., Darnell, J. E., Jr. and Gilman, M. Z. (1993) *Science* **261**, 1739–1744.
- Shuai, K., Stark, G. R., Kerr, I. M. and Darnell, J. E., Jr. (1993) *Science* **261**, 1744–1746.
- Shuai, K., Ziemiecki, A., Wilks, A. F., Harpur, A. G., Sadowski, H. B., Gilman, M. Z. and Darnell, J. E., Jr. (1993) *Nature* **366**, 580–583.
- Zhong, Z., Wen, Z. and Darnell, J. E., Jr. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4806–4810.
- Hocke, G. M., Barry, D. and Fey, G. H. (1992) *Mol. Cell. Biol.* **12**, 2282–2294.
- Wegenka, U. M., Buschmann, J., Luttkicken, C., Heinrich, P. C. and Horn, F. (1993) *Mol. Cell. Biol.* **13**, 276–288.
- Wagner, B. J., Hayes, T. E., Hoban, C. J. and Cochran, B. H. (1990) *EMBO J.* **9**, 4477–4484.
- Levy, D. E., Kessler, D. S., Pine, R. I. and Darnell, J. E., Jr. (1989) *Genes Dev.* **3**, 1362–1372.
- Reich, N., Evans, B., Levy, D. E., Fahey, D., Knight, E., Jr. and Darnell, J. E., Jr. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6394–6398.
- Schindler, C., Fu, X.-Y., Improt, T., Aebersold, R. and Darnell, J. E., Jr. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7836–7839.
- Fu, X.-Y., Schindler, C., Improt, T., Aebersold, R. and Darnell, J. E., Jr. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7840–7843.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Veals, S. A., Schindler, C., Leonard, D., Fu, X.-Y., Aebersold, R., Darnell, J. E., Jr. and Levy, D. E. (1992) *Mol. Cell. Biol.* **12**, 3315–3324.
- Hawkins, J. D. (1988) *Nucleic Acids. Res.* **16**, 9893–9906.
- Levy, D. E., Lew, D. J., Decker, T., Kessler, D. S. and Darnell, J. E., Jr. (1990) *EMBO J.* **9**, 1105–1111.
- Boisclair, Y. R., Brown, A. L., Casola, S. and Rechler, M. W. (1993) *J. Biol. Chem.* **268**, 24892–24901.
- Pine, R., Canova, A. and Schindler, C. (1994) *EMBO J.* **13**, 158–167.
- Tanaka, A., Gibbs, C. P., Arthur, R. R., Anderson, S. K., Kung, H.-J. and Fujita, D. J. (1987) *Mol. Cell. Biol.* **7**, 1978–1983.
- Anderson, S. K., Gibbs, C. P., Tanaka, A., King, H.-J. and Fujita, D. J. (1985) *Mol. Cell. Biol.* **5**, 1122–1129.
- Dominski, Z. and Kole, R. (1991) *Mol. Cell. Biol.* **11**, 6075–6083.