

---

**Mutant analysis of protein interactions with a nuclear factor I binding site in the SL3-3 virus enhancer**

---

Pia Nilsson, Bengt Hallberg, Anders Thornell and Thomas Grundström

---

Department of Applied Cell and Molecular Biology, University of Umeå, S-901 87 Umeå, Sweden

---

Received March 2, 1989; Revised and Accepted May 11, 1989

---

**ABSTRACT**

Nuclear factor I (NFI) is shown to be of importance for the activity of the enhancer element of a T-cell leukemogenic murine retrovirus, SL3-3, and for the regulation of this element by glucocorticoid. Each nucleotide of the binding site of the NFI proteins was mutated, and the effects of the mutations were quantitated with an electrophoretic mobility shift assay. Mutations in the inverted repeat of the binding site have symmetric effects which strongly support the notion that NFI proteins preferentially bind to dyad symmetry sites. Such binding sites were shown to be more than 100 fold stronger than the corresponding single binding sites. We find dyad symmetry sequences which are much stronger NFI binding sites than NFI sites identified in different genes and also stronger than previously proposed consensus binding sequences for NFI.

**INTRODUCTION**

The temporal and spatial restriction of fundamental cellular processes such as replication and transcription are believed to involve interactions of proteins with specific DNA recognition elements. Eukaryotic protein-coding genes contain a complex array of cis-regulatory elements that mediate induced, repressed, or basal transcription rates. Enhancer elements are cis-acting DNA segments which stimulate transcription from homologous and heterologous promoters of protein coding genes over long distances in an orientation independent manner when located either upstream or downstream from the RNA start site (for reviews and references see 1-4).

We have recently localized the enhancing DNA segments of the long terminal repeat (LTR) of a T-cell leukaemogenic murine retrovirus, SL3-3 (5). Two copies of a 72 bp sequence followed by a repetition of the first 34 bp of the repeat were found to be the most important DNA segments for the function of this enhancer in T-lymphocytes, where the enhancer is most active. The 34 bp repeat contains a DNA sequence, 5'-CCGGCCCAGGGCCAA-3', which shows homology with known binding sites for the transcription factor nuclear factor I (NFI). NFI was first identified in nuclear extracts of HeLa cells by virtue of its ability to enhance initiation of adenovirus DNA replication *in vitro* (6). NFI (TGGCA binding protein; CTF) has been shown to be important for the transcriptional activation of many different genes (7-10 and refs. therein).

Comparisons of different binding sites for NFI resulted in suggestions of a consensus sequence: TGGA/CN<sub>5</sub>GCCAA (11), PyTGGA/CAN<sub>3</sub>A/TGCCA (12), TGGA/CA/TN<sub>3</sub>A/TGCCAA (13), and TTGGCTN<sub>3</sub>AGCCAA (8). Several alterations of the NFI binding site of adenovirus in the region of homology to the consensus sequence have been shown to impair binding (13-16). Gronostajski found that approximately 50% of DNA segments containing the general sequence TGGN<sub>6</sub>GCCAA bind to NFI with

relatively high affinities and found a bias for a C or A residue in the first position of the central 6 bp region of strong binders (17). An analysis of contact points between NFI and the binding site in adenovirus 2 revealed that 22 out of 23 contacts distributed over a segment of 19 bp are accessible from one side of the DNA helix (18). The partial dyad symmetry of the binding site of adenovirus was found to be reflected by a corresponding symmetry in the contact points.

Different procedures for purification of 'nuclear factor I site binding proteins' from different sources have resulted in the isolation of different proteins or sets of proteins with apparent molecular weights ranging from 30 kDa to 66 kDa on SDS PAGE gels (8, 16, 19–24). Therefore, in the following the term nuclear factor 1 (NFI) does not represent a single protein but all the proteins which specifically bind to 'the NFI binding sequences'. Isolations of cDNA clones for NFI have shown that the heterogeneity of the proteins is both due to the existence of multiple genes (24) and to differential splicing of the mRNA (25). Ristiniemi and Oikarinen have recently reported that histone H1 is included in the set of proteins recognizing this type of DNA sequences (21).

In this communication we show that the enhancer of the SL3-3 virus contains binding sites for NFI which are of importance for the activity of the enhancer. To determine the importance of individual nucleotides for binding of NFI, we performed a systematic mutant analysis of the binding site with quantitation of the binding in electrophoretic mobility shift experiments. Binding to dyad symmetry sites was more than 100 fold stronger than binding to the corresponding single binding site, supporting the notion that NFI binds as a dimer. We show that the dyad symmetry sequences 5'-PyPyPyTGGCACAGTGCCAPuPu-3' are much stronger NFI binding sites than identified high affinity binding sites of different genes, and also stronger than previously proposed consensus binding sequences for NFI.

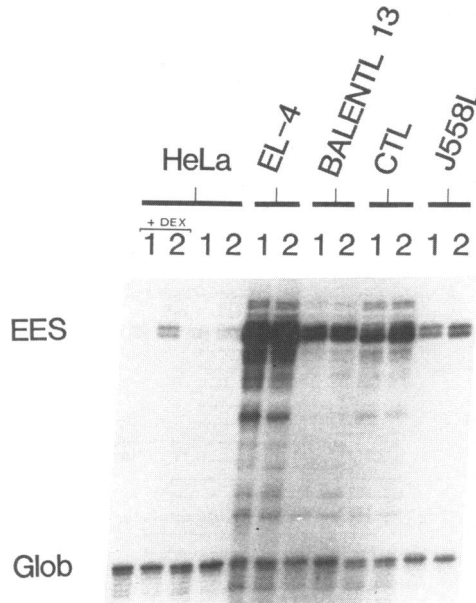
## MATERIALS AND METHODS

### *Construction of mutants and in vivo transcription assay*

Plasmid pESG134 (5) contains a version of the SL3-3 enhancer with two copies of the 34 bp repeat and one copy of the 38 bp segment in front of the SV40 early promoter region and the coding sequence of the rabbit  $\beta$ -globin gene. Plasmid pESG130 is a derivative of pESG134 containing a mutation of the CGG sequences at positions 133–135 and 277–279 to ATT. The PstI-ApaI and ApaI-BamHI DNA segments containing the mutation (positions 37 to 203 and 200 to 446) were assembled with the shotgun ligation technique (26) from the same sets of oligonucleotides as employed for construction of pESG134 (5), except that the proper nucleotides to create the mutations were replaced. The assemblies were confirmed by dideoxy sequencing, and plasmid pESG130 was constructed by standard subcloning techniques. The cotransfected reference plasmid pESG003 (5) has a structure corresponding to that of pESG134, except that it has the enhancer of SV40 and the promoter of rabbit  $\beta$ -globin. Lymphocytes and HeLa cells were transfected with the DEAE-Dextran and calcium phosphate techniques, respectively, as previously described (5). RNA were prepared with the LiCl-Urea method and the specific RNA were quantitated by S1 nuclease mapping as previously described (5).

### *Cell lines, cell culture and preparation of nuclear extracts*

The following cell lines were used: X.63 Ag8.653 (in the text shorter denoted X-63), a non-immunoglobulin secreting mouse myeloma (27, 28); K46, a mouse B cell lymphoma (29); EL-4, a mouse T helper cell lymphoma (30, 31); CTL, a monoclonal cytotoxic mouse T cell line (32); J558L, a mouse myeloma (33); 18–81 and 230–238, two pre-B cell



**Figure 1.** Quantitative S1 nuclease analysis of RNA from transient expression in different cell lines using (1) plasmid pESG 130, harbouring a substitution of nucleotide positions 133–135 and 277–279 from CGG to ATT, and (2) the corresponding wild type plasmid pESG 134. +Dex indicates expression performed with the addition of dexamethasone to 1  $\mu$ M 24 hours after addition of the DNA. EES corresponds to RNA transcribed from the early early SV 40 promoter start sites of the two plasmids and Glob to RNA transcribed from the co-transfected reference recombinant (Materials and Methods).

lines (34, 35); BALENTL 13, a mouse T helper cell line (36); Jurkat, a human T helper cell line (37); CHO, a Chinese hamster ovary fibroblast cell line (38); and HeLa, a human carcinoma. The cell lines were grown as previously described (39). Nuclear extracts from cell lines and primary cells from spleen, thymus, kidney and liver of CBA mice were prepared with eight protease inhibitors present as previously described (39).

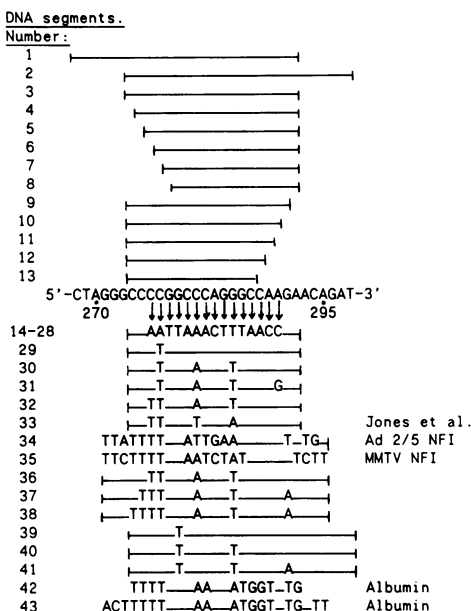
*Electrophoretic mobility shift assays*

Double-stranded DNA segments were prepared from oligonucleotides as previously described (39), and the electrophoretic mobility shift experiments were performed as

Table I. Effect in different cell lines of mutation in NFI binding site.

Recombinant	Relative transcription (per cent)				
	HeLa	EL-4	BALENTL 13	CTL	J558L
pESG134	100	100	100	100	100
pESG130 (NFI binding site mutant)	76	54	88	39	63

Specific transcription was during transient expression of plasmids with the SL3-3 enhancer in front of the early SV 40 promoters (see text). The values represent the average amounts of RNA ( $\pm 20\%$ ) of several independent transfection experiments (determined by quantitative S1 nuclease analysis) expressed relative to the value for pESG 134 which was taken as 100% in each cell line.



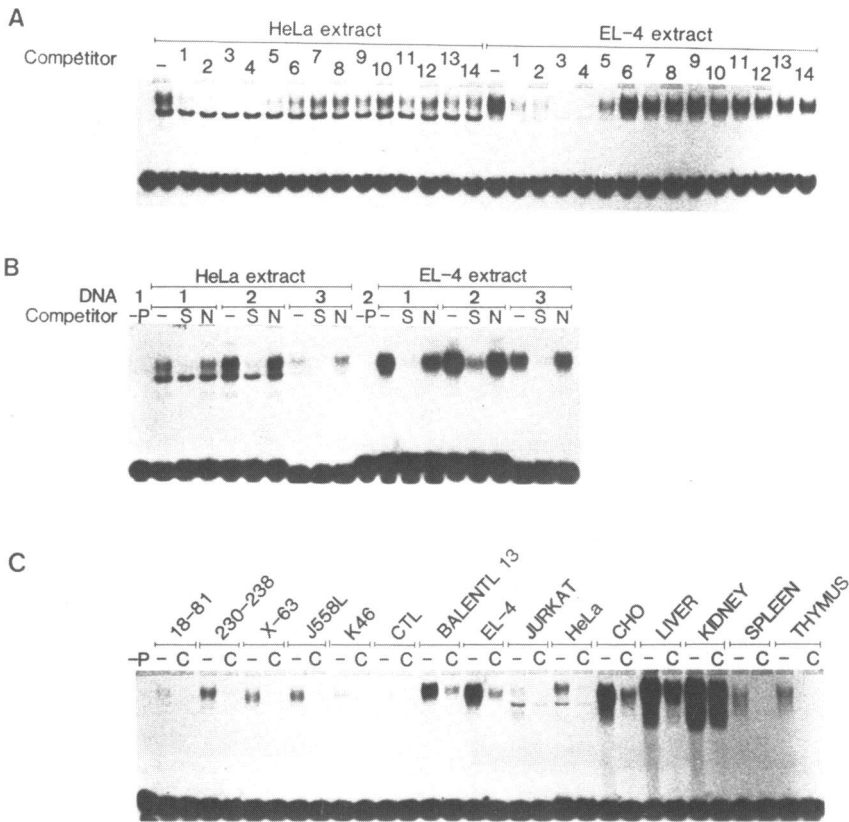
**Figure 2.** Double-stranded DNA segments used in NFI binding experiments. A repeated segment of the U3 region in the long terminal repeat of SL3-3 virus is indicated. DNA segments number 14 to 28 contain the indicated C ⇒ A and G ⇒ T one base pair substitution, respectively. Number 33 contains the consensus NFI binding sequence proposed by Jones et al. (8) and numbers 34 and 35 are NFI binding sites of adenovirus 2/5 (49) and mouse mammary tumor virus (50, 51), respectively. Numbers 42 and 43 contain the NFI binding sequence of the mouse albumin gene, positions -109--127 and -106 to -130, respectively (9, 10). Solid bars indicate positions with the same nucleotide as in SL3-3 wild type.

previously described (39). The complexes were quantitated by densitometric scanning of autoradiograms corresponding to different exposures of the gels. The purified NFI site binding protein histone H1 was from liver of rat (21).

## RESULTS

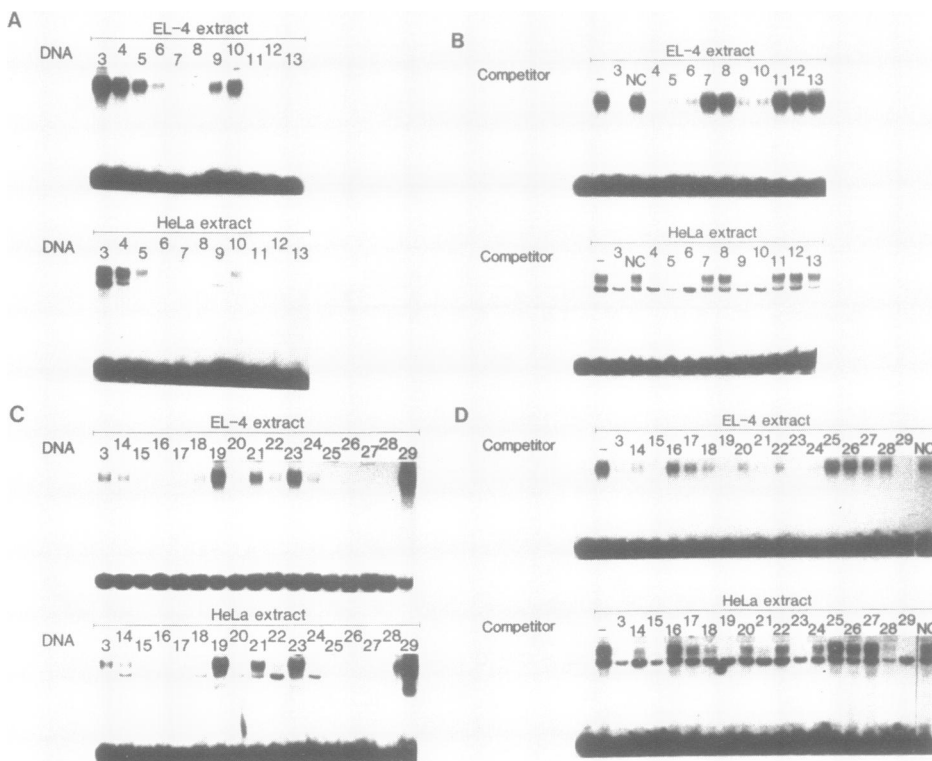
### *Mutation in potential nuclear factor I binding sites affects transient expression*

The most important region for the activity of the SL3-3 virus enhancer *in vivo* in T lymphocyte cell lines contains three copies of the DNA sequence 5'-CCGGCCCAGGGCCA-3' (positions 132-146, 204-218 and 276-290; 5) with homologies to known binding sites for nuclear factor I (NFI; 8, 11-13). To study if this DNA sequence was relevant for transcription *in vivo* in different cell lines, the enhancement of transcription during transient expression of a mutant in this sequence and the corresponding wild type were compared. The plasmids used, pESG134 (5) and the mutant derivative pESG130, carry the SL3-3 enhancer in front of the early promoters of SV40 and a rabbit  $\beta$ -globin gene. The plasmids have one 72-bp repeat less than the LTR of SL3-3, a deletion which has small effects on the level of transcription in different cell lines (5). Plasmid pESG130 has a 3 bp substitution in the proposed NFI binding site (CGG at positions 133-135 and 277-279 to ATT). Each of the three nucleotide positions was subsequently found to be important for binding of NFI *in vitro* (see below). The plasmids



**Figure 3.** Electrophoretic mobility shift experiments with the NFI binding site of the SL3-3 enhancer. Each reaction contained 2  $\mu$ g of nuclear extracts and where indicated a 50 fold excess of competitor DNA segment. A) Competition between labelled DNA segment number 2 (Fig. 2) and DNA segments containing binding sites for NFI or other transcription factors. Competitors; 1–5: NFI binding sites. 1: SL3-3 DNA segment number 2. 2: Proposed site of moloney murine leukaemia virus, positions 8053–8077 (52). 3: adenovirus 2/5, positions 19–43 (49). 4: mouse mammary tumor virus, positions –82––58 (50, 51). 5: mouse albumin gene, positions –130––106 (9, 10). 6–8: CCAAT-sequences. 6: murine *Ea1* gene, positions –66––37 (53). 7: herpes simplex virus TK gene, positions –96––67 (54). 8: mouse  $\beta$ -globin gene, positions –63––92 (55). 9: Sequence from the SL3-3 enhancer, positions 237–261 (39). 10–12: Motifs of the SV40 enhancer; positions 195–219, 256–280 and 237–261, respectively (44). 13–14:  $\mu$ 50 and  $\chi$ -3 motifs of the immunoglobulin heavy chain enhancer (56), respectively. B) Competitions of NFI binding with DNA segments from SL3-3 of different lengths. 1–3: end-labelled DNA segments numbers 1–3 (Fig. 2). –P: no nuclear extract added. Competitors; S: the same DNA segment as the labelled one. N: a non-homologous sequence, a SEF1 binding sequence of SL3-3 (39). C) Binding to the NFI site of SL3-3, using DNA segment number 3 (Fig. 2), with nuclear extracts from the indicated cell lines (described in Materials and Methods) and organs of CBA mice. Abbreviations: –P: no nuclear extract. C: the homologous DNA segment added as competitor.

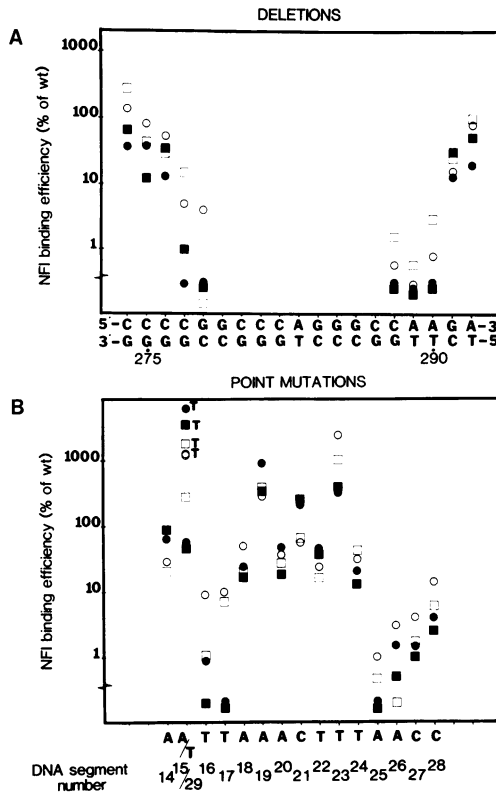
were cotransfected with a control plasmid followed by transient expression and quantification of transcription by S1 nuclease mapping (Materials and Methods). The levels of transcription obtained for different cell lines are shown in Fig. 1, and the average results from four to eight measurements are summarized in Table I. The mutation decreased transcription to different degrees in the cell lines tested. This shows that a part of the effect of the repeated



**Figure 4.** Binding of NFI to 1 fmol of labelled DNA segments of different lengths or with different single nucleotide substitutions in electrophoretic mobility shift experiments. 2  $\mu$ g of nuclear extracts of EL-4 or HeLa cells were used, and in the competition experiments a 100 fold molar excess of the indicated DNA segments (shown in Fig. 2). A) Binding of NFI to two sets of DNA segments which are shorter at either side. B) Competition of binding of NFI with the set of shorter DNA segments using labelled DNA segment number 3. C) Binding of NFI to DNA segments containing single point mutations. D) Competition of binding of NFI with DNA segments containing single point mutations using labelled DNA segment number 3. NC: non-homologous competitor, a SEF1 binding site of SL3-3 (39).

region on transcription is at least in many cell lines due to interactions with the postulated NFI binding sequences. As expected from the previously reported lack of significant effect of deleting the repeated DNA segment in HeLa cells (5), we found a very low effect of the mutation in this cell line. A decrease in transcription which may not be significant was also obtained for one of the lymphocyte cell lines, BALENTL 13.

We (5) and Celander et al. (40) have localized a glucocorticoid response element (GRE) to the repeated DNA segments of the SL3-3 enhancer. The response is especially pronounced in HeLa cells (5). A DNA sequence in the repeats with homology to a consensus GRE is located next to the proposed NFI binding sites (positions 146–160, 218–232 and 290–304). Interestingly, the 3 to 4 fold stimulation of transcription from pESG134 in HeLa cells by the glucocorticoid dexamethasone was not obtained with the mutant plasmid (Fig. 1). Thus, although the proposed binding site has a very minor effect on the basal level of enhancer activity in HeLa cells (Table 1), it is nevertheless essential to enable a glucocorticoid response in this cell line.



**Figure 5.** Summary of effects of deletions and single point mutations on binding of NFI to the site in the SL3-3 enhancer. The values represent the averages of quantitations of multiple experiment such as those in Fig. 5, expressed relative to DNA segment number 3 (Fig. 2). Abbreviations; ●:Binding to labelled DNA segments using nuclear extract of HeLa cells. ■:Binding to labelled DNA segments using nuclear extract of EL-4 cells. ○:Competition of binding of NFI from nuclear extract of HeLa cells. □:Competition of binding of NFI from nuclear extract of EL-4 cells. A) Effects of deletions. The symbols indicate the position of the last nucleotide deleted. The wild type SL3-3 sequence is indicated. B) Effects of single point mutations. The symbols indicate the position of the nucleotide substituted in the DNA segments used, numbers 14–29 (Fig. 2). Their different nucleotide substitutions are indicated. T indicates values with the C to T substitution of DNA segment number 29.

*Nuclear factor I binds to the SL3-3 enhancer*

To analyze if NFI binds to the DNA sequence containing the studied mutation, an electrophoretic mobility shift experiment was performed with DNA segment number 2 (Fig. 2) and nuclear extracts from both the murine T helper cell line EL-4, because of the T cell preference of the enhancer, and HeLa cells (used in several previous studies of NFI; 6, 8, 17, 19, 41 and refs. therein). Complexes with decreased mobilities could be detected when using either of the two nuclear extracts (Fig. 3A). Most of these complexes could be competed away with a 50 fold molar excess of the homologous non-labelled DNA segment and with segments containing known NFI binding sites of adenovirus 2 and 5 (Ad 2/5), mouse mammary tumor virus (MMTV), the mouse albumin gene and a proposed NFI binding site in moloney murine leukemia virus (Mo-MuLV; competitors 2 to 5 in fig. 3A). In contrast, competition could not be seen with DNA segments containing

Table II. Effects of mutations on competition for binding of NFI.

DNA sequence	DNA segment number	Relative competition efficiency	
		EL-4 extract	HeLa extract
CCCTGGCCCAGGGCCAAGA	29	1.0	1.0
-----A---T-----	30	12	12
-----A---T---G--	31	18	8.1
·---T---A---T-----	32	18	11
--T---T---A-----	33	9.7	6.8
---C-----	3	0.06	0.08
TTATTTGGATTGAAGCCAATATGA	34	1.0	1.0
--C-----A-CT-T---GTCTT	35	1.2	0.7
CCCC--CCCAGG-----G-AC-GAT	2	0.07	0.09
GGGCC---CACAGT-----G-AC-	36	8.6	4.9
GGGC-----CACAGT-----A-AC-	37	ND	ND
GGG-----CACAGT-----A-AC-	38	7.4	6.5

Competition between different synthetic DNA segments for binding of NFI from nuclear extracts to DNA segments number 29 (upper half) and number 34 (lower half). The values are the mean of at least six measurements, and are expressed relative to competition with the homologous DNA segment which was taken as 1 for each cell extract. Solid bars indicate that the nucleotide is identical to that of the labelled DNA segment used. ND, not determined.

binding sequences for a set of other identified nuclear factors, including recognition sequences for CCAAT-binding proteins (41–43), which appear grossly similar to those of NFI. The competable complexes were not obtained as a single band but as a series of bands with different mobilities (Fig. 3A). Interestingly, more efficient competitions occurred with the NFI binding sites of Ad 2/5 and MMTV compared to those with Mo-MuLV, the albumine gene or SL3-3 itself.

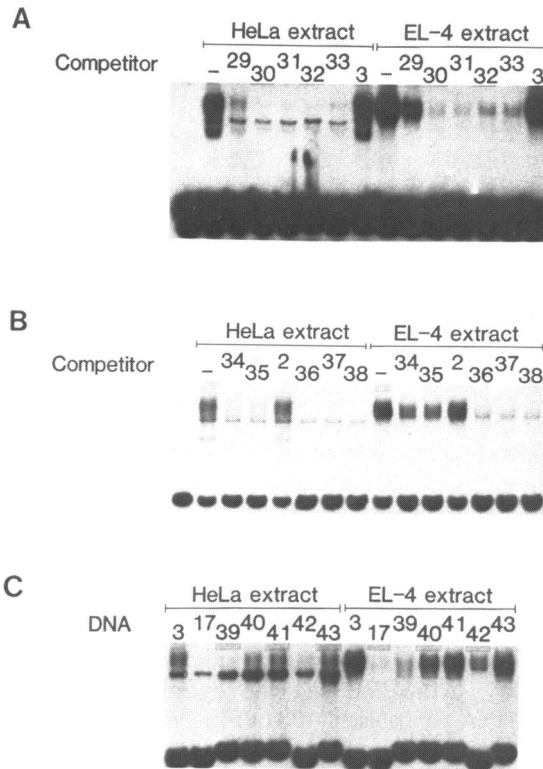
#### *NFI proteins are present in a broad range of cell types*

To investigate the occurrence of NFI binding activities in lymphoid cells compared to other cell types, binding to DNA segment 3 (Fig. 2) was analysed for nuclear extracts prepared from different cell lines and from tissues of CBA mice. The shorter DNA segment, containing the NFI binding site (Fig. 3B), was used to minimize the possibility of complexes due to non-NFI interactions. A distribution of efficiently competable complexes with different mobilities could be detected in all cell types (Fig. 3C). With the two human cell lines HeLa and Jurkat, a weak non-competable band could also be seen. The amounts of NFI complexes varied considerably between the different cell lines. Relatively high amounts were detected in the mouse T helper cell line extracts and the chinese hamster ovary fibroblast line, CHO. Even higher amounts of NFI were detected in the liver and kidney cell extracts. No correlation could be seen between the levels of NFI detectable with extracts from different cell lines and the level of the effect on transcription by mutation of the NFI binding site in that cell line (cf. Table 1 and Fig. 3C).

#### *Deletion analysis of the limits of the NFI binding site*

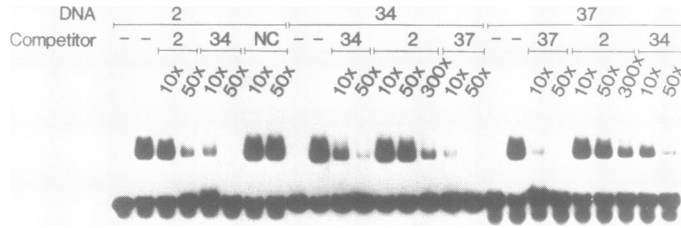
Quantitation of binding of NFI to DNA segments 1, 2 and 3 (Fig. 2) in experiments such as that displayed in Fig. 3B showed that SL3-3 positions 268 to 273 were not of significant importance for binding, and although positions 293 to 298 improved NFI binding, about 50 per cent of the binding capacity remained in the DNA segment containing only positions





**Figure 6.** A) Competition of NFI binding to an up-mutation of the SL3-3 sequence, DNA segment number 29 (Fig. 2), using NFI binding sites with different modifications. 1  $\mu\text{g}$  of nuclear extracts and a 10 fold excess of competitor DNA sequences (Fig. 2) were used. B) Competition of binding to the NFI binding site of adenovirus 2/5, DNA segment number 34, using improved NFI binding sites. 1  $\mu\text{g}$  of nuclear extracts and a 5 fold excess of competitor DNA sequences (Fig. 2) were used. C) Binding of NFI from nuclear extracts to the wild type NFI binding site of SL3-3 and to different half sites of the dyad symmetry sequence. 1 fmol of DNA segments (numbers as in Fig. 2) was incubated with 2  $\mu\text{g}$  of extracts.

274 to 292. To further determine the borders of the NFI binding site of SL3-3 we constructed double-stranded oligonucleotides where 1 to 5 base pairs were deleted from one end or the other. These DNA segments (numbers 4 to 13 of Fig. 2) were used both as labelled probes and as competitor DNA segments in the electrophoretic mobility shift assay, because mutations could potentially affect the ability to form stable complex and to compete for NFI binding differentially. The results of typical such experiments are shown in Fig. 4A and B. The average results of quantitation of multiple such experiments are shown in Fig. 5A. The results with the two types of assays confirm each other. However, with either nuclear extract most deletions decreased the ability to form a stable complex more than the ability to compete. The step-wise deletion of positions 274 to 276 from the DNA segment had a less than three fold effect on the ability to compete and a less than ten fold effect on the ability to form the complex. The subsequent deletion of position 277 led to a much larger decrease in the binding of NFI, and with deletion of position 278 very little binding remained. Deletions from the other side of the DNA segment had a larger effect despite



**Figure 7.** Competition analysis of binding of a purified NFI site binding protein, histone H1, in an electrophoretic mobility shift experiment. 1 fmol of labelled DNA segments number 2, 34 and 37 (Fig. 2) was incubated with 10 ng, 3 ng and 3 ng, respectively, of histone H1 together with the indicated excess of competitor DNA sequences (Fig. 2). NC: non-homologous competitor, positions 195–219 in the SV40 enhancer (44).

that the axis of sequence symmetry is exactly in the middle of the segment. Only positions 292 and 291 could be deleted with less than a 10 fold decrease in the binding of NFI, and with further deletions less than 3 per cent of the ability of NFI to bind or to compete remained. The larger effect of the second set of deletions strongly indicates that the 5'-CC-3' sequence at positions 276 to 277 is not as efficient, and thereby not as important, for binding of NFI as the 5'-TT-3' sequence on the other side of the 5'-GGCCC-3' dyad symmetry.

*Base pair substitution mutants in the NFI binding site*

To determine the importance of each individual nucleotide for the binding of NFI, we made DNA segments with one base pair changed in each. The purines were changed to their non-complementary pyrimidine and the pyrimidines changed to their non-complementary purine (DNA segments 14–28 of Fig. 2). As the NFI binding site of SL3-3 contains one divergence from all suggested consensus sequences (8, 11–13), a C instead of a T at position 277, we also made this substitution. The results of typical binding and competition experiments with the mutations are shown in Fig. 4C and D, respectively. The average results of quantitation of multiple such experiments are shown in Fig. 5B. The results with the two types of assays confirm each other. However, most mutations decreased the ability to form a stable complex more than the ability to compete. All substitutions in the dyad symmetry of the binding site, 5'-GGCCC-3', had very similar effects on the two sides of the axis of symmetry. Mutations of the two G residues were drastic down mutations, and mutations of the first and third C were less drastic down mutations. In contrast, the mutations of the middle C (positions 281 and 285) were up-mutations. All mutations outside of the inverted repeat were down mutations except at position 277, where the change from C to T increased binding drastically, while the change to an A had a much smaller effect.

To study if the three nucleotide substitutions which drastically improved binding of NFI act synergistically, they were combined (DNA segment number 30 of Fig. 2). This DNA segment was found to compete about 12 fold better than the best single point mutant (Fig. 5A and Table II). Dyads with the sequences 5'-CTGGCAC-3' and 5'-TTGGCAC-3' in the palindrome were found to compete about equally well, and about 100 to 300 fold better than the SL3-3 wild type sequence. These binding sequences were found to compete about 1.5 fold better than a DNA segment with the consensus binding site for NFI proposed by Jones et al. (8), despite that the important C residue in the end of the dyad, TGGCTC, (see above), was used in the DNA segment although it is not included in that proposed

consensus (Fig. 6A and Table II; DNA segment number 33 of Fig. 2). The NFI binding sites of adenovirus 2/5 and mouse mammary tumor virus were found to be 5 to 9 times less efficient competitors than one of the improved binding sites (cf. DNA segments 34–36 in Fig. 6B and Table II).

The NFI sites of SL3-3 contain C residues at several positions (274, 275 and 291) where T residues often are found in NFI binding sequences. Nucleotide substitutions changing the last two or all three of these positions to T residues did not significantly affect the competition efficiency (Fig. 6B and Table II). The relative efficiencies of the binding sites found in the competition experiments were confirmed in experiments where binding of NFI to DNA segments 29–37 were quantitated, but much smaller differences between the DNA segments were obtained (data not shown). The latter results probably indicate that with the level of binding strength of DNA segments 29, 34 and 35 (Table II) a high proportion of the available NFI in the reactions is binding.

NFI has been reported to bind to a non-palindromic sequence in the mouse albumin gene (9, 10). To study the formation of complexes with a non-palindromic sequence we analysed a series of DNA segments with many divergences from an efficient NFI binding sequence in one half of the SL3-3 sequence and with the other half stepwise improved. A DNA segment with an extremely poor 'left' half of the binding site (number 39 in Fig. 6C) showed a drastically decreased binding efficiency. Up-mutations in the 'right' half (segments number 40 and 41) compensated most of this decrease, leading to 63 and 84 % of the wild type binding in HeLa, and 55 and 56 % of the wild type binding in EL-4, respectively. A 25 bp segment (number 43) with the NFI site of the mouse albumin gene gave a similar level of binding of NFI (Fig. 6C). Thus, a non-repeated NFI binding sequence can bind NFI with almost the same efficiency as to the wild type NFI site of SL3-3 if the remaining sequence is close to the optimal NFI site. It is notable that none of the point mutations, including those with a non-repeated binding sequence, decreased the mobility shifts of the complexes.

#### *Binding with purified histone H1*

The binding analysis presented so far has been performed with nuclear extracts and not with purified NFI site binding protein to ensure that none of the proteins relevant for the formation of the protein-DNA complexes was purified away. We wanted to compare the binding of NFI from crude nuclear extracts with the binding of a purified NFI site binding protein. A protein binding to the NFI recognition sequence in the mouse  $\alpha_2(I)$  collagen promoter has recently been purified from rat liver using DNA binding sequence affinity chromatography (21). The protein was shown to be histone H1 (21). Competition analysis of the binding of purified histone H1 shows that it has about the same relative binding efficiency to the wild type NFI binding site of SL3-3, an optimized NFI binding site (number 37) and the site of adenovirus 2/5 as described above for NFI from the crude nuclear extracts (Fig. 7).

## DISCUSSION

We have demonstrated binding of nuclear factor I to the sequence 5'-CCGGCCCAGGGCCAA-3' in the enhancer of SL3-3 virus. These interactions were shown to be important for the activity of the enhancer element *in vivo* in many cell lines and for its regulation by glucocorticoid in HeLa cells. The formation of complexes *in vitro* in electrophoretic mobility shift experiments were comparable with known binding sites

sequences which appear grossly similar to the recognition sequences of NFI (41–43). The interactions were detectable in a broad range of cell lines of different species and primary cells from different organs of mice. No correlation could be seen between the level of the effect of the mutation in the NFI binding site on transcription *in vivo* and the amounts of complexes formed *in vitro* in that cell type. For example, the cytotoxic T cell line CTL, where the largest effect of the mutation was seen, gave extracts resulting in relatively low amounts of the complexes. These discrepancies could have many reasons. Maybe a favourable context of efficient interaction(s) with another transcription factor(s) binding nearby on the enhancer in a particular cell type is of high importance *in vivo*. It is notable in this context that the 72 bp repeat which contains the NFI site is a more efficient enhancer by itself in CTL than in other cell lines studied (5), and that this cell line shows higher amounts of the transcription factor proteins SEF1 which bind to multiple sites in the repeated sequence (39 and Anders Thornell, Bengt Hallberg and Thomas Grundström, manuscript in preparation). Several studies have demonstrated the importance of co-operative activation for the function of enhancers (44–46). A recent report that an NFI binding site can mediate transcriptional activation by transforming growth factors  $\beta_1$  and  $\beta_2$  (47) raises other possible reasons for the discrepancies such as cell type differences in growth factor receptor contents or in the secondary message pathways. A fourth possible reason emanates from the report by Ristiniemi *et al.* that histone H1 is one of the major proteins binding to NFI sites (21). Histone H1 is considered to be a general repressor of transcription (48 for review). Thus, the obtained NFI protein-DNA complexes may represent both proteins activating and repressing transcription in unknown proportions which may vary between cell lines.

The large distribution of complexes obtained with extracts from all different cell lines and organs (Fig. 3) indicates the presence of differences in the set of proteins in the different NFI complexes in these cells. None of the NFI complexes of different mobilities was found to be selectively affected by anyone of the mutations used in the analysis of the NFI binding site of SL3-3, neither with nuclear extracts from EL-4 or from HeLa. This finding is indicative of a conserved DNA binding domain(s) within the proteins of the different NFI complexes. None of the extensive set of mutants used, which includes changes at all possible nucleotide positions of the DNA segments, shifted the mobilities of the complexes. This provides strong arguments against the possibility that any other protein in the complex than the NFI site binding proteins interacts specifically with a sequence of the DNA segment. The differences between nuclear extracts of HeLa and EL-4 in the effects of the different mutations on the binding efficiencies were in general relatively small, and smaller than the differences when comparing direct binding and ability to compete for NFI binding to the same sequence, arguing that these proteins from different species and cell types are relatively similar in their preferred DNA binding sequences.

We found that in competition experiments was the NFI binding site of SL3-3 about 8 to 17 fold less efficient in NFI binding as compared to the previously identified NFI sites of adenovirus 2/5 (Ad 2/5) and mouse mammary tumor virus (MMTV; Table II). When the C at position 277 of the SL3-3 sequence was replaced by a T (as it is in Ad2/5) the NFI binding increased 13 to 17 fold (Table II). This is a strong argument for that the C at position 277 is the major cause of the difference in NFI binding ability between SL3-3 and Ad 2/5 or MMTV. The change of a C to an A in position 281 or a G to a T in position 285 increased NFI binding between 3 and 20 fold for the two different assays and sources of extract (Fig. 4B). When the C to T mutation at position 277 was combined with the

for NFI but not with any one of a high number of other tested DNA segments with binding sites for other transcription factors, including CCAAT-binding proteins with recognition two up-mutations, the competition efficiency increased a further 12 fold with total extracts of both EL-4 and HeLa (Table II). This is very interesting since these two nucleotides are close to the center of the NFI binding site of SL3-3, a region where several studies, mainly on the NFI site of Ad 2/5, have announced conflicting data for an NFI consensus binding sequence. Gronostajski et al. (11) proposed any of the four nucleotides, Nowok et al. (12) proposed A and T/A, Leegwater et al. (13) proposed A/T and Jones et al. (8) proposed T and A, respectively, at these two positions. We find that dyad symmetries of the sequences 5'-PyPyPyTGGCA/TC-3' bind NFI with high efficiencies, and that palindromes with 5'TGGCAC-3' are about 1.5 fold better than such with 5'TGGCTC-3'. We also find a clear importance of the C residue in the 3' end of the palindrome, a position previously not claimed to be included in the recognition sequences of NFI.

A non-dyad NFI binding sequence was found to bind NFI with about the same efficiency as the wild type NFI site of SL3-3 did if the remaining sequence was close to the optimal NFI site. Such a level of NFI binding can be clearly significant *in vivo* as shown in Fig. 1 and Table I. However, an efficient inverted repeat binding site can be an about 150 to 400 fold stronger binding site *in vitro* (Fig. 6C and Table II). Dyad symmetries with the sequences 5'-PyPyPyTGGCAC-3' were found to compete for binding of NFI from HeLa and EL-4 about 150 to 350 fold better than the wild type SL3-3 sequence and 9 to 45 fold better than the NFI binding sites of adenovirus 2/5 and mouse mammary tumor virus. About the same relative efficiency of binding to the wild type NFI binding site of SL3-3, an optimized NFI binding site and the site of adenovirus 2/5 was obtained with a purified NFI site binding protein, histone H1 (Fig. 7). Dyad symmetries with the sequences 5'-PyPyPyTGGCAC-3' are thereby much stronger NFI binding sites than identified high efficiency binding sites of different genes, and also stronger than previously proposed consensus binding sequences for NFI.

## ACKNOWLEDGEMENTS

We would like to thank Christine Grundström for skilful technical assistance and Peter Brodin for help with illustrations. We are grateful to Jouko Oikarinen for the generous gift of a highly purified NFI site binding protein. This work was supported by grants from the Swedish Natural Science Research Council (Dnr 4892-108), the Swedish Cancer Society (Dnr 2235-B89-03XA), the National Swedish Board for Technological Development (Dnr 512-87-347) and the Research foundation of the oncologic clinic of Umeå (549/88). Thomas Grundström was recipient of a special researcher position from the Swedish Natural Science Research Council (Dnr 4892-301).

## REFERENCES

1. Mc Knight, S. and Tjian, R. (1986) *Cell* 46, 795-805.
2. Sassone-Corsi, P. and Borrelli, E. (1986) *Trends Genet.* 2, 215-219.
3. Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) *Science* 236, 1237-1245.
4. Jones, N.C., Rigby, W.J. and Ziff, E.B. (1988) *Genes and Dev.* 2, 267-281.
5. Hallberg, B. and Grundström, T. (1988) *Nucleic Acids Res* 16, 5927-5944.
6. Nagata, K., Guggenheimer, R.A., Enomoto, T., Lichy, J.H. and Hurwitz, J. (1982) *Proc. Natl. Acad. Sci. USA.* 79, 6438-6442.
7. Sippel, A.E., Fritton, H., Theisen, M., Borgmeyer, U., Strech-Jurk, U., Igo-Kemenes, T. (1986) In Grodzicker, T., Sharp, P.A. and Botchan, M. R. (eds), *Cancer Cells 4: DNA Tumor Viruses: Gene Expression and Replication.* Cold Spring Harbor Laboratory Press, NY.

8. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) *Cell* **48**, 79–89.
9. Lichtsteiner, S., Wuarin, J. and Schibler, U. (1987) *Cell* **51**, 963–973.
10. Cerrahini, S., Raymondjean, M., Carranca, A.G., Herbolme, P. and Yaniv, M. (1987) *Cell* **50**, 627–638.
11. Gronostajski, R.M., Adhya, S., Nagata, K., Guggenheimer, R.A. and Hurwitz, J. (1985) *Mol Cell Biol* **5**, 964–971.
12. Nowock, J., Borgmeyer, U., Püschel, A.W., Rupp, R.A.W. and Sippel, A.E. (1985) *Nucleic Acids Res* **13**, 2045–2061.
13. Leegwater, P.A.J., van der Vliet, P.C., Rupp, R.A.W., Nowock, J. and Sippel, A.E. (1986) *EMBO J* **5**, 381–386.
14. de Vries, E., van Driel, W., Tromp, M., van Boom, J. and van der Vliet, P.C. (1985) *Nucleic Acids Res* **13**, 4935–4952.
15. Schneider, R., Gander, I., Müller, U., Mertz, R. and Winnacker, E.L. (1986) *Nucleic Acids Res* **14**, 1303–1317.
16. Meisterernst, M., Gander, I., Rogge, L. and Winnacker, E.L. (1988) *Nucleic Acids Res* **16**: 4419–4435.
17. Gronostajski, R.M. (1987) *Nucleic Acids Res* **15**, 5545–5559.
18. de Vries, E., van Driel, W., van den Heuvel, S.J.L. and van der Vliet, P.C. (1987) *EMBO J* **6**, 161–168.
19. Rosenfeld, P.J. and Kelly, T.J. (1986) *J Biol Chem* **261**, 1398–1408.
20. Meisterernst, M., Rogge, L., Donath, C., Gander, I., Lottspeich, F., Mertz, R., Dobner, T., Föckler, R., Steltzer, G. and Winnacker, E.L. (1988) *FEBS* **236**, 27–32.
21. Ristiniemi, J. and Oikarinen J. (1989) *J Biol Chem* **264**, 2164–2174.
22. Rupp, R.A.W. and Sippel, A.E. (1987) *Nucleic Acids Res* **15**, 9707–9726.
23. Paonessa, G., Gounari, F., Frank, R. and Cortese, R. (1988) *EMBO J* **7**, 3115–3123.
24. Gil, G., Smith, J.R., Goldstein, J.L., Slaughter, C.A., Orth, K., Brown, M.S. and Osborne, T.F. (1988) *Proc Natl Acad Sci USA* **85**, 8963–8967.
25. Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) *Nature* **334**, 218–224.
26. Grundström, T., Zenke, W.M., Wintzerith, M., Matthes, H.W., Staub, A. and Chambon, P. (1985) *Nucleic Acids Res* **13**, 3305–3316.
27. Kearney, J.F., Radbruch, A., Liesegang, B. and Rajewsky, K. (1979) *J Immunol* **123**, 1548–1550.
28. Köhler, G. and Milstein, C. (1975) *Nature (London)* **256**, 495–497.
29. Kim, K.J., Kanellopoulos-Langevin, C., Merwin, R.M., Sachs, D.H. and Asofsky, R. (1979) *J Immunol* **122**, 549–554.
30. Gorer, P.A. (1950) *Brit. J. Cancer* **4**, 372–379.
31. Mohit, B. and Fan, K. (1971) *Science* **171**, 75–77.
32. Baker, P.E., Gillis, S. and Smith, K.A. (1979) *J Exp Med* **149**, 273–278.
33. Oi, V.T., Morrison, S.L., Herzenberg, L.A. and Berg, P. (1983) *Proc Natl Acad Sci USA* **80**, 825–829.
34. Siden, E.J., Baltimore, D., Clark, D. and Rosenberg, N.E. (1979) *Cell* **16**, 389–396.
35. Alt, F., Rosenberg, N., Lewis, S., Thomas, E. and Baltimore, D. (1981) *Cell* **27**, 381–390.
36. Kim, K.J., Weinbaum, F.I., Mathieson, B.J., McKeever, P.E. and Asofsky, R. (1978) *J Immunol* **121**, 339–344.
37. Gillis, S. and Watson, J. (1980) *J Exp Med* **152**, 1709–1719.
38. Kao, F.T., Johnson, R.T. and Puck, T.T. (1969) *Science* **164**, 312–314.
39. Thornell, A., Hallberg, B. and Grundström, T. (1988) *Mol Cell Biol* **8**, 1625–1637.
40. Celander, D., Hsu, B.L. and Haseltine, W.A. (1988) *J Virol* **62**, 1314–1322.
41. Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) *Cell* **53**, 11–24.
42. Oikarinen, J., Hatamochi, A. and de Crombrughe, B. (1987) *J Biol Chem* **262**, 11064–11070.
43. Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) *Cell* **50**, 863–872.
44. Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. (1986) *EMBO J* **5**, 387–397.
45. Ondek, B., Shepard, A. and Herr, W. (1987) *EMBO J* **6**, 1017–1025.
46. Fromental, C., Kanno, M., Nomiya, H. and Chambon, P. (1988) *Cell* **54**, 943–953.
47. Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B. and de Crombrughe, B. (1988) *Cell* **52**, 405–414.
48. Weintraub, H. (1985) *Cell* **42**, 705–711.
49. Nagata, K., Guggenheimer, R.A. and Hurwitz, J. (1983) *Proc Natl Acad Sci USA* **80**, 6177–6181.

50. Buetti, E. and Kühnel, B. (1986) *J Mol Biol* *190*, 379–389.
51. Cordingley, M.G., Riegel, A.T. and Hager, G.L. (1987) *Cell* *48*, 261–270.
52. Speck, N.A. and Baltimore, D. (1987) *Mol Cell Biol* *7*, 1101–1110.
53. Dorn, A., Durand, B., Marfing, C., Le Meur, M., Benoist, C. and Mathis, D. (1987) *Proc Natl Acad Sci USA* *84*, 6249–6253.
54. Graves, B.J., Johnson, P.F. and McKnight, S.L. (1986) *Cell* *44*, 565–576.
55. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) *Cell* *21*, 653–668.
56. Sen, R. and Baltimore, D. (1986) *Cell* *46*, 705–716.

**This article, submitted on disc, has been automatically  
converted into this typeset format by the publisher.**