Review

Roles of the NFI/CTF gene family in transcription and development

Richard M. Gronostajski *

Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic Foundation and Department of Biochemistry, 9500 Euclid Ave., Case Western Reserve University, Cleveland, OH 44195, USA

Received 20 December 1999; received in revised form 15 February 2000; accepted 13 March 2000

Received by A.J. van Wijnen

Abstract

The Nuclear Factor I (NFI) family of site-specific DNA-binding proteins (also known as CTF or CAAT box transcription factor) functions both in viral DNA replication and in the regulation of gene expression. The classes of genes whose expression is modulated by NFI include those that are ubiquitously expressed, as well as those that are hormonally, nutritionally, and developmentally regulated. The NFI family is composed of four members in vertebrates (NFI-A, NFI-B, NFI-C and NFI-X), and the four NFI genes are expressed in unique, but overlapping, patterns during mouse embryogenesis and in the adult. Transcripts of each NFI gene are differentially spliced, yielding as many as nine distinct proteins from a single gene. Products of the four NFI genes differ in their abilities to either activate or repress transcription, likely through fundamentally different mechanisms. Here, we will review the properties of the NFI genes and proteins and their known functions in gene expression and development. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: CAAT box transcription factor, Cancer, CTD, Gene expression, Nuclear Factor I

1. Introduction

The Nuclear Factor I (NFI) family of site-specific DNA-binding proteins plays wide reaching roles in animal physiology, biochemistry and pathology. While first described as being required for the replication of Adenovirus DNA, this family of transcription/replication proteins has been implicated in the replication of several other viruses and has been shown to regulate the transcription of a large variety of cellular and viral genes. In addition, NFI proteins have been associated with changes in the growth state of cells and with a number of oncogenic processes and disease states. Since the role of NFI in adenovirus DNA replication has been recently reviewed (de Jong and van der Vliet, 1999), we will focus here on the evolution of the NFI gene family and on the role of NFI proteins in gene expression and development.

2. Discovery of viral and cellular NFI-binding sites

After the initial observation that NFI protein isolated from nuclear extracts of human HeLa cells greatly stimulated the initiation of adenovirus DNA replication (Nagata et al., 1982), it was shown that NFI was a site-specific DNA-binding protein that bound to the adenovirus origin of replication (Nagata et al., 1983). Direct isolation of NFI-binding sites from cellular DNA (Gronostajski et al., 1985) and comparison with viral and cellular sites identified by DNA-binding assays (Hennighausen et al., 1985; Leegwater et al., 1985; Nowock et al., 1985) demonstrated that NFI protein bound as a dimer to the dyad symmetric consensus sequence TTGGC(N5)GCCAA on duplex DNA. Sequences flanking the consensus and in the degenerate 5 nt spacer region appear to modulate the NFI-binding affinity (Gronostajski, 1986, 1987). Quantitative analysis of binding showed that while NFI bound very tightly to a dyad symmetric site \( (K_d \sim 10^{-11} \text{ M}) \), NFI could also bind specifically to individual half sites (TTGGC or GCCAA) with a somewhat reduced affinity \( (K_d \sim 10^{-9} \text{ M}) \) (Meisterernst et al., 1988a). The identifi-
Evolution of the NFI multigene family

NFI cDNAs isolated from rat (Paonessa et al., 1988), human (Santoro et al., 1988), hamster (Gil et al., 1988), mouse (Inoue et al., 1990) and porcine (Meisterernst et al., 1988, 1989) sources indicated that multiple NFI genes are present in vertebrate genomes. Several different nomenclatures arose for the NFI genes, leading to confusion regarding the number of NFI genes in mammals. The Sippel laboratory identified four NFI genes in the chicken [designated NFI-A, NFI-B, NFI-C (for CAAT box), and NFI-X (for hamster NFI-X)] (Rupp et al., 1990; Kruse et al., 1991) and developed a consistent nomenclature for the four vertebrate NFI genes (Fig. 1). Homologs of these four NFI genes have been described in every vertebrate species examined from Xenopus (Roulet et al., 1995; Puzianowska-Kuznicka and Shi, 1996) to mouse (Chaudhry et al., 1998) and humans (Apt et al., 1994; Kulkarni and Gronostajski, 1996), and likely represent all of the NFI genes in vertebrates. The four NFI genes are distributed across three chromosomes in both, with NFI-C and NFI-X being together on 19p13.3 in humans and Nfia and Nfib together on chr. 4 in mice (Fig. 1, right). The four vertebrate NFI genes appear to have arisen by gene duplication during chordate evolution. A single NFI gene has been identified in the cephalochordate Amphioxus, which may be the progenitor of the four vertebrate genes (Fletcher et al., 1999). A single NFI gene is also present in both the nematode C. elegans (nfi-1) and Drosophila, but no NFI genes are present in any of the sequenced prokaryotic or simple eukaryotic genomes, suggesting that the NFI gene itself arose during evolution of the metazoan lineage or was lost independently in the simple eukaryotic and prokaryotic lineages (Fletcher et al., 1999). This feature distinguishes the NFI gene family from some other families of site-specific DNA-binding proteins, including the Hox genes and zinc-finger DNA-binding proteins, for which many prokaryotic homologues have been described. The significance of this restriction of the NFI genes to the metazoan lineage is still unclear, but it has been noted that the increase in diversity of NFI genes is coincident with the increase in complexity of the vertebrate body plan.

Comparative genomics has revealed several intriguing features of the NFI genes. The porcine and human NFI-C genes were the first genes for which the genomic structure was determined, showing strong conservation of all 11 exons of the genes (Meisterernst et al., 1989) (Fig. 1, NFI-C). The rat NFI-A genomic structure was determined next, and 11 exons were identified, all of which were similar in length to the NFI-C exons, ± 5 aa (Xu et al., 1997) (Fig 1, NFI-A). The amino acid sequence homology between the rat NFI-A and human NFI-C exons ranges between 100% in a conserved five-residue C-terminus, to 91% in the 177 aa DNA-binding domain encoding exon, to a low of 39% in the exon following the DNA-binding domain (overall homology 61%). The human NFI-X genomic sequence was recently completed by the Human Genome Sequencing project, and it likewise has 11 exons with sizes comparable to those of rat NFI-A and human NFI-C (Accession Nos AC004660 and AC007787, Fig. 1). This high degree of structural homology shows that little divergence of the genes has occurred since their generation prior to the establishment of the avian lineage (which contains all four genes). However, this structural homology breaks down when the vertebrate genes are compared to the C. elegans gene. Unlike the four vertebrate genes, which show a significant homology throughout their sequences, no homology is seen outside the NFI-DNA-binding domain between the C. elegans gene and the four vertebrate NFI genes. In addition, while the DNA-binding domains of the C. elegans and vertebrate NFI proteins are relatively conserved at the sequence level, they differ dramatically at the genomic level. While all four mouse (and presumably human) NFI genes have their DNA-binding domains encoded by an unusually large 532 nt second exon, in C. elegans, this large exon is interrupted by four additional introns and has a 3' extension of 213 nt that has no sequence homology to the vertebrate genes (Fletcher et al., 1999). These four additional introns are missing in the single Amphioxus NFI gene identified, suggesting that the exons were...
Fig. 1. Domains and alternative splicing of vertebrate NFI genes. The five lines illustrate several general features (top) and alternatively spliced products of the four NFI genes from vertebrates. Pan-specific gene names are on the right with the human (Qian et al., 1995) and mouse (Fletcher et al., 1999) chromosome locations shown below the name. As described in the text, the general structure (top line) is composed of 11 coding exons (boxes) with the N-terminal DNA-binding and dimerization domain (labeled DNA binding and dimerization) encoded predominantly by exon 2 (gray box). Within the second exon are four conserved cysteine residues (labeled C) required for DNA-binding and redox regulation of binding, a basic alpha helical domain (labeled Basic helix) and the Ad Pol-binding domain (labeled Ad Pol binding). Numbers above the line are approximate residue numbers and those below the line are exon numbers. The C-terminal regions of each protein encode specialized domains noted in the text (labeled Transactivation and repression) including the proline-rich transactivation domain (labeled Proline-rich). For each gene, the largest extant cDNAs contain 11 exons, and alternatively spliced isoforms are shown by angle brackets below each gene with names below. The names of each isoform are derived from the species (c, chicken; r, rat; m, mouse; p, porcine; h, human; x, *Xenopus*), gene (a, b, c or x) and particular spliced isoform (1–7). Only a subset of known isoforms is shown, and few have been confirmed in more than one species. Alternative first exons are shown by boxes or lines connected to the second exon. The names of the first exons denote conservation of the coding regions of exons 1a (8–10 aa), 1b (32–47 aa) and 1c (1 aa, M). The E1b exons of human and mouse NFI-B are predicted from GENBANK genomic or EST sequences (*NFIB*, AL136366.3; *Nfib*, AW106080). Isoform names used by previous authors were retained when possible with the exception of some NFI-A cDNAs that were previously named NFI-B due to their cloning from brain. The heptamers above NFI-A and NFI-C are regions homologous to the C-terminal domain (CTD) repeat of RNA polymerase II. Gray lines in NFI-B show predicted exons since the genomic sequence of NFI-B is not available. GENBANK Accession Nos for each isoform are available upon request. BLAST analysis (Altschul et al., 1997) and the size of NFI mRNAs suggest that each gene may have 5' or 3' untranslated exons.

either inserted recently into the nematode gene, or lost from the cephalochordate gene prior to the duplication of the four genes in the vertebrate lineage. The recently determined sequence of the single NFI gene in *Drosophila* shows that the second of the four introns present in the *C. elegans* NFI gene is also present in *Drosophila*, indicating that this intron was likely present prior to the divergence of nematodes and insects and thus is an ‘ancient’ intron (Fig. 2, bold intron). The *Drosophila* NFI gene also contains a 219 nt extension in the exon encoding the C-terminal end of DNA-binding domain, similar to the *C. elegans* gene. In addition, a splice acceptor site at the 5' end of the DNA-binding domain of both the *C. elegans* and *Drosophila* NFI genes is conserved to the nucleotide with those of all four mouse NFI genes, suggesting strongly that this early metazoan gene is the progenitor of the four vertebrate genes (Fig. 2). The lack of conservation of exons outside the NFI DNA-binding domain between the *C. elegans* and vertebrate genes may suggest that significant changes in function have developed during evolution of the gene family. However, it is also possible that there is a significant structural and functional homology between the genes that has been disguised by sequence changes during the ~ 500 Myr since the divergence of the nematode and vertebrate lineages. It will be necessary to perform functional comparisons between the *C. elegans* and vertebrates genes to resolve this issue.
Fig. 2. Conservation of exons encoding the DNA-binding domain of NFI in vertebrates and early metazoans. The exons encoding the DNA-binding domain of NFI proteins (gray boxes) from vertebrates, *C. elegans* and *Drosophila* (GENBANK AC015236) are shown together with the general structure of the vertebrate genes (white boxes). As described in the text, the vertebrate genes encode their DNA-binding domains predominantly within identically sized single exons, while the *C. elegans* and *Drosophila* genes have an extension at the 3′ end of the region and possess internal introns (triangles and lines within exon). The lengths of the DNA-binding domain encoding exons are shown above or below each gene and the intron sizes are shown near each intron. The introns in bold are at identical positions in the *C. elegans* and *Drosophila* genes and likely represent an early intron. The arrows on the *C. elegans* and *Drosophila* genes indicate additional exons present in these genes but are not shown.

4. Unusual features of NFI transcripts

While the general features discussed below (N-terminal DNA-binding and C-terminal transcriptional modulation domains) accurately describe the vertebrate NFI proteins, additional complexity is generated by alternative processing of NFI transcripts (Fig. 1). Alternative processing takes three forms: (1) alternative polyadenylation between exons 2 and 3 of NFI-B, yielding the short NFI-B3 described below (Liu et al., 1997); (2) alternative splicing of exons in the 3′ regions of all four NFI genes, yielding multiple proteins from each gene that have different C-termini fused to the same DNA-binding domain (Santoro et al., 1988; Apt et al., 1994; Roulet et al., 1995); and (3) alternative splicing/promoter usage leading to different first exons being fused to the same DNA-binding and C-terminal domains (Inoue et al., 1990; das Neves et al., 1999). The alternative splicing in 3′ regions of the NFI transcripts is phylogenetically conserved, suggesting conserved biological functions for each isoform (Kruse and Sippel, 1994a). In some instances, this alternative splicing appears to be regulated, since changes in the splicing pattern of human NFI-C were observed during differentiation of human leukemic cells in culture (Kulkarni and Gronostajski, 1996), and different relative levels of alternatively spliced NFI transcripts are present in different cell types (Apt et al., 1994; Chaudhry et al., 1997).

Other unusual features of NFI transcripts include their large 5′ and/or 3′ untranslated regions and the frequent presence of short (2–33 aa) putative open reading frames (ORFs) upstream of the predicted initiation codons. Analysis of the 5′ regions of NFI cDNAs present in GenBank shows the presence of such ORFs from eight to 98 residues upstream of the predicted initiation codons of all four NFI genes from a number of vertebrates (unpublished data). Together with the observation that the major NFI transcripts are very large [mNFI-A, -B, -C and -X mRNAs are ~10.5, ~9.7, ~7.7 and ~6.0 kb, respectively (Nebl et al., 1994; Chaudhry et al., 1997)], the presence of these short ORFs raises the possibility of translational regulation of NFI protein expression.

Another possible function for the large untranslated regions of NFI mRNAs may be in the regulation of mRNA stability. Stable introduction of an activated Ha-ras gene into mouse cells results in downregulation of NFI-C and NFI-X mRNA levels (Nebl et al., 1994). There is no change in transcription of the NFI-C and NFI-X genes with the introduction of Ha-ras, but the half-lives of their mRNA are drastically decreased. It will be important to identify the specific mRNA sequences required for destabilization of NFI-C and NFI-X transcripts and to determine whether the other NFI mRNAs are subject to similar destabilization.

5. Structural and functional domains of NFI proteins

As mentioned above, transcripts of each of the four vertebrate NFI genes are alternatively spliced generating multiple proteins from each gene. This complexity of protein and mRNA isoforms can be simplified if we consider domains that are conserved in all of the isoforms and between the four vertebrate genes. However, since recent studies have demonstrated significant differences between the functional activities of products of the four NFI genes, it should be remembered that...
generalizations can sometimes be misleading and must be continuously tested.

6. N-terminal DNA-binding/dimerization domain

The typical NFI protein is composed of an N-terminal DNA-binding/dimerization domain and C-terminal transcriptional activation and/or repression domains (Fig. 1). The N-terminal DNA-binding/dimerization domain is preceded by alternative exons encoding 8–47 aa domains of unknown function (although there is strong conservation of this region between the four genes (Meisterernst et al., 1989; Rupp et al., 1990; Kruse et al., 1991; Kruse and Sippel, 1994a) [Fig. 1, E1a–c]). Deletion analysis has shown that the DNA-binding/dimerization domain is ~200 aa in length and is ~90% identical between the four chicken, mouse, and human NFI genes (Fig. 1, DNA binding and dimerization). This N-terminal domain is sufficient for DNA-binding activity, dimerization and the stimulation of adenovirus DNA replication (Mermod et al., 1989; Gounari et al., 1990). Point mutations made within this domain have shown that dimerization is essential for DNA-binding activity but that DNA-binding activity can be abolished independently with retention of dimerization activity (Armentero et al., 1994). Also, point mutations within this domain can abolish adenovirus DNA replication while retaining both DNA binding and dimerization (Fig. 1, Ad Pol binding). Thus, the stimulation of adenovirus DNA replication requires dimerization, DNA-binding activity and additional functions of the N-terminal domain. Mutational analysis of the N-terminal domain, together with direct binding and kinetic studies, have shown that the specific interaction of the N-terminal DNA-binding/dimerization domain with the Adenovirus DNA polymerase appears essential for the recruitment of the polymerase into a replication complex and the stimulation of replication (Bosher et al., 1990; Chen et al., 1990; Mul et al., 1990; Armentero et al., 1994; reviewed in de Jong and van der Vliet, 1999).

The NFI DNA-binding domain has no detectable sequence homology with other known DNA-binding domains and thus may be structurally distinct. Four cysteine residues are conserved between all NFI DNA-binding domains, and three of the four residues are required for DNA-binding activity (Bandyopadhyay and Gronostajski, 1994) [Fig. 1C, exon 2]. The fourth cysteine residue, while not essential for DNA-binding activity, makes NFI proteins sensitive to oxidative inactivation (redox regulation). Mutation of this residue does not affect DNA binding but confers resistance to oxidative inactivation in vitro (Bandyopadhyay and Gronostajski, 1994; Bandyopadhyay et al., 1998). This feature of oxidation sensitivity is shared by a number of transcription factors and may play a role in the cellular response to oxidative damage (Abate et al., 1990; Guichmann et al., 1992; Bandyopadhyay et al., 1998).

The four cysteine residues are located within the C-terminal 2/3rd of the ~200 aa DNA-binding/dimerization domain, in a subdomain shown to be sufficient for NFI dimerization, low-affinity site-specific DNA binding, and interaction with adenovirus DNA polymerase (Dekker et al., 1996). The N-terminal ~1/3 of the ~200 aa DNA-binding/dimerization domain is highly basic (Meisterernst et al., 1989), appears to fold into a stable alpha-helical subdomain that can bind DNA non-specifically, and, when fused to the C-terminal 2/3 subdomain, can increase NFI DNA-binding affinity ~100-fold (Dekker et al., 1996) (Fig. 1, Basic helix). This subdivision of the DNA-binding/dimerization domain into two independent subdomains may be of evolutionary interest. As mentioned above, while the vertebrate NFI DNA-binding domains are encoded predominantly by single large exons (532 nt, 177.3 aa), the C. elegans NFI DNA-binding domain is encoded by five exons that may represent the ancestral gene structure (Fig. 2). Thus, the vertebrate NFI-DNA-binding domains may be composed of subdomains that were encoded by separate exons during the early evolution of the NFI gene. It will be of interest to examine the NFI gene structure in other simple metazoa to determine the pathway of evolution of the NFI DNA-binding domain.

While the NFI DNA-binding/dimerization domain is often described as an ~200–220 aa domain encoded predominantly by a single exon, the minimum size of this domain may differ slightly between the four NFI genes, and more C-terminal regions of the proteins may influence DNA-binding affinity. For example, while initial studies demonstrated that the N-terminal 220 aa domain of the human NFI-C/CTF protein could bind specifically to DNA, a larger molecule of 399 aa bound to DNA with a somewhat higher affinity (Mermod et al., 1989). Likewise, a 199 aa fragment of porcine NFI-C bound specifically, but weakly, to NFI-site DNA, while a C-terminally extended 247 aa protein was fully active (Meisterernst et al., 1989) However, these differences in binding affinity may be due to improved protein folding of the larger molecules rather than to additional DNA contacts. For example, both the NFI-C/CTF and NFI-A DNA-binding domains can be spliced to multiple alternative third exons that have no apparent sequence homology, but each protein product has a similar strong DNA-binding affinity (Meisterernst et al., 1989; Mermod et al., 1989; Kruse and Sippel, 1994b). Thus, an ~200–220 aa NFI DNA-binding/dimerization domain is a useful functional definition.

Since NFI proteins bind to DNA as dimers, several studies have examined whether heterodimers can form...
between the products of the different NFI genes. Efficient formation of DNA-binding heterodimers has been shown between products of all four chicken NFI genes, with few or no differences being seen in DNA-binding affinity, specificity, or stability of the dimers Kruse and Sippel, 1994b). As was seen previously with homodimer formation of human and porcine NFI-C and rat NFI-A proteins (Meisterernst et al., 1989; Mermod et al., 1989), the different chicken NFI proteins needed to be cotranslated in order to form heterodimers. Mixing of preformed homodimers yielded no heterodimers. However, recent studies with a truncated human NFI-B3 isoform consisting of only the N-terminal 186 residues of NFI-B suggest that heterodimers between this protein and other NFI gene products may have different characteristics depending on the ‘partner’ NFI protein (Liu et al., 1997) (Fig. 1, NFI-B, hB3). When expressed in Drosophila cells, NFI-B3 alone has no DNA-binding activity and, when coexpressed with a larger NFI-B protein, inhibits DNA binding by the larger protein (presumably by formation of non-DNA-binding heterodimers). However, when NFI-B3 was coexpressed with full-length human NFI-X or NFI-C/CTF proteins, DNA-binding activity was retained, and DNA-protein complexes consistent with the predicted sizes of heterodimers between NFI-B3 and the other proteins were seen. Thus, it appears that NFI-B3 may be either DNA-binding proficient or deficient depending on its heterodimerization partner. Transcripts encoding the NFI-B3 protein appear to be generated by use of a polyadenylation/termination signal present in the intron between exons 2 and 3 of the human NFI-B gene, leading to the generation of a short transcript containing only exons 1 and 2 and two amino acids encoded within the intron (Fig. 1, NFI-B hB3). Although the level of this NFI-B3 transcript is very low in most human tissues, it may play a role in modulating NFI activity in MRHF fibroblasts where it appears to be the only NFI-B transcript expressed (Liu et al., 1997). Whether the other NFI genes also produce truncated transcripts of a similar nature is unknown.

7. C-terminal transactivation and repression domains

While the DNA-binding and replication activities of NFI proteins reside in the N-terminal domain, C-terminal domains have been implicated in most, though not all, regulation of gene expression by NFI. As described above, alternative splicing generates many variants of the C-terminal domains of NFI proteins, only a fraction of which have been tested for functional activity (Fig. 1). The initial cloning and characterization of NFI-C/CTF transcripts demonstrated that the C-terminal 100 residues of NFI-C/CTF1 (residues 399–499) were required for maximal transcriptional activation of an NFI-site containing promoter in Drosophila Schneider cells (Mermod et al., 1989). This C-terminal domain is unusually rich in proline residues (25%) and has been termed a proline-rich activation domain (Fig. 2, Proline-rich). When linked to heterologous DNA-binding domains this 100 residue proline-rich domain stimulates transcription five- to 10-fold in mammalian and Drosophila cells (Mermod et al., 1989; Martinez et al., 1991; Seipel et al., 1992). However, it should be noted that in Drosophila cells (which lack NFI), regions of NFI-C/CTF1 outside this proline-rich domain (residues 220–400) increased the ability of the proline-rich domain to stimulate transcription by approximately threefold (Mermod et al., 1989), suggesting that the two regions may cooperate in activating transcription. A similar proline-rich domain required for transcriptional activation in yeast has also been identified in the rat NFI-A gene (Monaci et al., 1995). A more detailed analysis of potential mechanisms of transcriptional activation by the proline-rich domain of NFI-C/CTF is given in Section 8.

NFI proteins and binding sites have also been implicated in repression of transcription from several promoters. NFI-binding sites have been identified as negative regulatory elements of a number of promoters and ‘silencers’, including those at the peripherin (Adams et al., 1995), eta-globin (Macleod and Plumb, 1991), GLUT4 genes (Cooke and Lane, 1999b). These elements have been identified in transient transfection assays using specific promoters in a variety of cell types. However, due to the overlapping expression pattern of these four alternatively spliced NFI genes, it is unclear which specific forms of NFI protein mediate repression at these elements in vivo. Protein domains that mediate transcriptional repression in transient or stable transfection assays have been identified in both rat NFI-A (Monaci et al., 1995; Osada et al., 1997a,b) and mouse NFI-X proteins. These domains encompass residues 318–509 of rat NFI-A and residues 190–280 of mouse NFI-X, and both domains can repress transcription of chimeric promoters when fused to the heterologous GAL4 DNA-binding domain (Neh and Cato, 1995; Osada et al., 1997b). There is no obvious sequence homology between the NFI-A and NFI-X repression domains, suggesting that they may function through different mechanisms. It is important to note that since some specific NFI proteins have been shown to activate transcription under one condition, and repress transcription in another, it appears likely that many instances, repression and activation by NFI proteins will be both cell-type- and promoter-specific (see Section 8).

The final activity ascribed to C-terminal domains of NFI proteins is inhibition of the DNA-binding activity
of NFI. Full-length Xenopus NFI-X proteins (414–497 aa) have only a weak DNA-binding activity in vitro, and deletion of the C-terminus yields a 321 aa protein with increased DNA-binding activity (Roulet et al., 1995) that inhibits DNA binding in cis is between residues 322–405 of the NFI-X proteins, present in all three alternatively spliced Xenopus NFI-X isoforms cloned to date, and inhibited the DNA-binding activity of the heterologous yeast GAL4 DNA-binding domain when fused in cis. This inhibitory domain also contributes to transcriptional activation by NFI-X, since deletion of the domain reduced transactivation by NFI-X proteins approximately fourfold, and fusion of the domain to GAL4 confers increased transactivation ability on a GAL4 responsive promoter. The simultaneous decrease in DNA-binding activity in vitro, but increase in transactivation activity in vivo, suggests that other proteins may interact with this region to unmask NFI consensus NFI-binding sites, and that NFI may activate transcription by direct displacement of repressive histones, either by direct competition for DNA binding or by interactions of histones with the proline-rich transactivation domain of NFI-C (Fig. 3, middle). A number of studies suggest that histone H1 can bind weakly to consensus NFI-binding sites, and that NFI may activate transcription by direct displacement of histone binding at such sites (Ristiniemi and Oikarinen, 1989; Gao et al., 1998). In addition, the C-terminal proline rich domain of NFI-C interacts directly with histone H3 in a yeast two-hybrid assay, suggesting that they may function together in vivo (Alevizopoulos et al., 1995). Together with the observation that NFI-C and partially purified coactivator fractions can overcome the repressive effects of histone H1 on promoter function in vitro (Dusserre and Mermod, 1992), these studies lend support to the model that NFI proteins may activate transcription through effects on histone (and by inference nucleosome) binding. Such a mechanism may be of particular importance at promoters known to contain phased nucleosomes, such as the mouse mammary tumor virus (MMTV) promoter discussed below.

8. Mechanisms of transcriptional modulation by NFI proteins: transactivation

As discussed above, binding sites of NFI proteins have been implicated in both activation and repression of promoters. This suggests that NFI proteins likely affect transcription through multiple mechanisms. The best studied mechanism used by NFI proteins to activate transcription is through direct interaction with basal transcription factors (Figs 3, top). The largest NFI-C isoform (NFI-C_CTF1) has an ~100 aa proline rich domain (Mermod et al., 1989) that contains a single copy of the heptapeptide repeat from the C-terminal domain of RNA polymerase II (CTD repeat, PTPSYS) (Meistererst et al., 1989) (Fig. 3, NFI-C). This proline-rich domain has been shown to function as a transactivation domain when fused to heterologous DNA-binding domains (Martinez et al., 1991), and deletion of the domain from NFI-C drastically reduces transactivation by NFI-C in Drosophila (Mermod et al., 1989), mammalian (Chaudhry et al., 1998) and yeast cells (Kim and Roeder, 1993; Wendler et al., 1994; Xiao et al., 1994). This domain has been shown to interact with both human TFIIIB (Kim and Roeder, 1994) and yeast TBP (Xiao et al., 1994) in vitro, and deletion of the CTD repeat abolishes both the interaction with TFIIIB and transactivation in yeast (Xiao et al., 1994). These studies indicate that while the CTD repeat is important for activity, the proline-rich surrounding sequences also contribute substantially to transactivation. For example, two alternatively spliced isoforms of NFI-C lacking the CTD repeat (CTF4 and CTF7) activate transcription in yeast more potently than does NFI-C_CTF1, which contains the CTD repeat (Altman et al., 1994; Wenzelides et al., 1996). Whether the mechanism of transactivation of NFI-C isoforms is through interaction with TFIIIB or TBP has not been determined. The other three NFI genes also encode proteins with proline rich C-termini, but none contains perfect matches to the CTD repeat. Thus, while products of all four NFI genes can activate reporter constructs in human cells (Apt et al., 1993, 1994; Chaudhry et al., 1998), it is unclear whether the mechanism of activation is the same for all the proteins.

A second mechanism proposed for activation by NFI proteins is through displacement of repressive histones, either by direct competition for DNA binding or by interactions of histones with the proline-rich transactivation domain of NFI-C_CTF1 (Fig. 3, middle). A number of studies suggest that histone H1 can bind weakly to consensus NFI-binding sites, and that NFI may activate transcription by direct displacement of histone binding at such sites (Ristiniemi and Oikarinen, 1989; Gao et al., 1998). In addition, the C-terminal proline rich domain of NFI-C interacts directly with histone H3 in a yeast two-hybrid assay, suggesting that they may function together in vivo (Alevizopoulos et al., 1995). Together with the observation that NFI-C and partially purified coactivator fractions can overcome the repressive effects of histone H1 on promoter function in vitro (Dusserre and Mermod, 1992), these studies lend support to the model that NFI proteins may activate transcription through effects on histone (and by inference nucleosome) binding. Such a mechanism may be of particular importance at promoters known to contain phased nucleosomes, such as the mouse mammary tumor virus (MMTV) promoter discussed below.

It is also likely that specific interactions between NFI proteins and various coactivator proteins play a role in transcriptional activation (Fig. 3, top). Early studies showed that the C-terminal proline-rich domain of NFI-C could bind to coactivator proteins needed for NFI-C-mediated activation of transcription in vitro (Tanese et al., 1991; Dusserre and Mermod, 1992). One specific coactivator shown to interact with the proline-rich domain is TAFII55, which also interacts with a number of other transcription factors including Sp1, YY1, USF and HIV TAT (Chiang and Roeder, 1995). Another potential NFI coactivator protein may be pirin, a protein identified by interaction with NFI-C in a yeast two-hybrid screen (Wendler et al., 1997). While the function of pirin is still unknown, its identification as a protein that also interacts with the ankyrin repeat of the Bcl3 oncprotein suggests that it could integrate the activity of NFI proteins with other transcription factors, such as those in the NFkB/rel family (Dechend et al., 1999). Other proteins that potentially play a role in NFI activation of transcription are the p300/CBP family of
coactivators. The C-terminal proline-rich domain of NFI-C/CTF1 cooperates in transactivation with residues 451–682 of CBP in a mammalian two-hybrid assay, suggesting potential interactions between these proteins in vivo (Leahy et al., 1999). NFI proteins also interact specifically with the Ski oncoprotein in vitro, and expression of Ski protein can potentiate activation by NFI proteins in transiently transfected Drosophila cells (Tarapore et al., 1997). Together with the observation that overexpression of p300 or of SRC-1 can cooperate in repression of the MMTV promoter by truncated forms of NFI-C (Chaudhry et al., 1999), these data suggest that NFI proteins may interact with a variety of coactivator proteins in vivo, and the relative importance of any given coactivator may be cell-type or promoter-specific.

9. Mechanisms of transcriptional modulation by NFI proteins: repression

As is seen with transactivation, it seems likely that multiple mechanisms exist by which NFI proteins can repress transcription. One mechanism postulated for repression by NFI proteins is through direct competition with more potent transactivators for binding at adjacent sites (Fig. 3, middle). Competition between NFI proteins and Sp1 for binding to adjacent sites has been proposed as a means for NFI to repress Sp1 activation of the mouse α1(I) collagen promoter (Nehls et al., 1991, 1992). Similarly, competition between binding of NFI with HNF4 and HNF1 for overlapping sites on the rat pyruvate kinase promoter and albumin enhancer, respectively, is proposed to play a role in NFI repression of these genes (Bernier et al., 1993; Yamada et al., 1997). Such competition may play a role in cell-type-specific activation/repression by NFI proteins, where the balance between activation and repression may be dependent on the specific isoforms of NFI expressed in a given cell type.

NFI-binding sites have also been seen to promote repression under conditions where competition between binding sites is unlikely (MacLeod and Plumb, 1991; Adams et al., 1995; Szabo et al., 1995; Gao et al., 1996; Osada et al., 1997a; Crawford et al., 1998; Rajas et al., 1998; Cooke and Lane, 1999b; Leahy et al., 1999). In these instances, the specific form of NFI protein mediating the repression is unknown since no studies have examined the NFI isoforms present in the relevant cells. However, several studies have shown that specific
C-terminal regions of NFI proteins can function as repressors when attached to heterologous DNA-binding domains, supporting the hypothesis that direct repression by NFI proteins occurs in vivo (see Section 7). It is unknown as to how these repression domains of NFI function, but they may be related to known active repression processes such as the recruitment of corepressor proteins by hormone receptors, or direct interaction with the basal transcription apparatus (see Hanna-Rose and Hannon, 1996; Manley et al., 1996; Pazin and Kadonaga, 1997 for reviews).

While most studies of NFI repression have focused on C-terminal repression domains, a subdomain of the NFI-C DNA-binding domain has been implicated in repression of the MMTV promoter. The MMTV promoter contains an NFI-binding site that is essential for glucocorticoid induction of the gene (Mikesiek et al., 1987; Cato et al., 1988; Bueti et al., 1989). In human JEG3 cells deficient in NFI proteins, expression of either NFI-A, -B, -C or -X isoforms greatly enhances glucocorticoid-dependent MMTV expression (Chuang et al., 1996; Chaudhry et al., 1997, 1998). In contrast, in human HeLa cells that contain high levels of endogenous NFI proteins, expression of NFI-C or NFI-X proteins represses glucocorticoid induction of MMTV expression, while NFI-A and NFI-B do not (Chaudhry et al., 1999).

Surprisingly, the region of NFI-C required for this demonstration of repression of the MMTV promoter is a 160 aa subdomain of the NFI-C that appears incapable of binding DNA, yet represses glucocorticoid-dependent transcription in vitro with other NFI proteins. Repression by NFI-C is alleviated by overexpression of the coactivators p300/CBP and SRC-1 and by high levels of glucocorticoid receptor. In addition, repression by NFI-C is receptor- and cell-type specific, occurring with glucocorticoid but not progesterone induction of the promoter, and in HeLa and COS-1 but not JEG3 or 293 cells, receptively. Since the MMTV promoter is known to contain phased nucleosomes important for its expression (Richard-Foy and Hager, 1987; Bresnick et al., 1990, 1992; Pina et al., 1990; Archer et al., 1991; Truss et al., 1993, 1995; Mows et al., 1994), it will be of interest to determine whether repression by NFI is mediated by NFI dimers with other NFI proteins. Repression by NFI-C is alleviated by overexpression of the coactivators p300/CBP and SRC-1 and by high levels of glucocorticoid receptor. In addition, repression by NFI-C is receptor- and cell-type specific, occurring with glucocorticoid but not progesterone induction of the promoter, and in HeLa and COS-1 but not JEG3 or 293 cells, receptively. Since the MMTV promoter is known to contain phased nucleosomes important for its expression (Richard-Foy and Hager, 1987; Bresnick et al., 1990, 1992; Pina et al., 1990; Archer et al., 1991; Truss et al., 1993, 1995; Mows et al., 1994), it will be of interest to determine whether repression by NFI is mediated through changes in nucleosome structure at the promoter (Blomquist et al., 1996; Truss et al., 1996; Chavez et al., 1996; and Chavez and Beato, 1997; Smith et al., 1997).

10. Hormonal and signal transduction pathways in which NFI has been implicated

NFI proteins or binding sites have been shown to affect the expression of genes regulated by a number of signal transduction pathways, including those controlled by insulin (Cooke and Lane, 1999b), TGF-β (Rossi et al., 1988; Riccio et al., 1992; Alexizopoulos et al., 1995, 1997, 1999; Sun et al., 1998), cAMP (Chu et al., 1991; Lu et al., 1992; Cooke and Lane, 1999a), steroid hor-

mones (Garlatti et al., 1996; Chaudhry et al., 1999), vitamin D (Candelliere et al., 1996), vitamin B6 (Algood et al., 1993), TNFα (Alexizopoulos and Mermod, 1996), FSH (Olsson et al., 1993), DNA-PK (Jackson et al., 1990), thyrotropin (Ortiz et al., 1999) and others. In most cases, only a single pathway-specific gene or hormone-dependent response has been studied, making generalization impossible. In addition, since the expression of NFI proteins can be affected by the growth and differentiation state of cells (Goyal et al., 1990; Kulkarni and Gronostajski, 1996), it is difficult to determine whether the effects of some hormones/growth factors on NFI proteins are direct or indirect. Even where rapid insulin-dependent changes in NFI phosphorylation have been detected and correlated with changes in expression of a gene containing an NFI-binding site (Cooke and Lane, 1999b), it is difficult to demonstrate a cause-and-effect relationship because there is no evidence that phosphorylation changes in NFI proteins can directly influence either DNA-binding or transcriptional modulation. Similarly, O-glycosylation of NFI proteins was demonstrated in 1988 (Jackson and Tjian, 1988), yet it is still unclear whether this modification affects NFI function in vivo. To resolve these issues, it will likely be necessary to: (1) determine the specific isoforms of NFI present in cell types during hormonal stimulation, (2) demonstrate the specific biochemical pathways by which hormones/growth factors affect NFI expression or modification and (3) develop well-defined in-vitro transcription assays that accurately reflect the activity of NFI proteins in vivo. An additional approach would be to analyze the hormone function in mice lacking one or more NFI gene (see Section 11).

11. NFI proteins in development and cancer

Binding sites for NFI proteins have been characterized from genes expressed specifically in almost every organ system and tissue, including brain (Elder et al., 1999; Bedfòrd et al., 1998), lung (Bachurski et al., 1992; Chaudhry et al., 1997, 1998), liver (Cereghini et al., 1987; Li and Rosen, 1995; Furlong et al., 1998), blood (Fischer et al., 1993; Knezetic and Bedon, 1997), mammary gland (Watson et al., 1991; Li and Rosen, 1995; Furlong et al., 1996), pituitary (Courtois et al., 1990), retina (Ben-Or and Okret, 1993), olfactory epithelium (Butaiokova et al., 1999; Baraaieter et al., 1999) fibroblasts (Rossi et al., 1988; Alonso et al., 1996; Iozzo et al., 1997), epithelial cells (Apt et al., 1993), adipocytes (Graves et al., 1991), chondrocytes (Szabo et al., 1995), etc.
neurons (Elder et al., 1992; Adams et al., 1995; Bedford et al., 1998) and glia (Tamura et al., 1988a,b, Amemiya et al., 1992; Miura et al., 1990; Kumar et al., 1993; Taveggia et al., 1998; Krohn et al., 1999). For most of these, the NFI-binding sites have been shown to be important for the expression of the gene. With such a diverse set of tissue-specific and developmentally regulated genes under the control of NFI proteins, it appears likely that NFI proteins play a major role in development of the corpus callosum is seen. One possibility is that a number of NFI genes has been detected in either whole embryos or specific embryonic and adult tissues (unpublished data and das Neves et al., 1999). To address this issue of functional redundancy, it will likely be necessary to disrupt all four NFI genes and examine the phenotypes of animals lacking multiple NFI gene products.

NFI proteins have been implicated in the control of cell growth in both humans and model systems. The NFI-B gene is a recurrent translocation partner of the HMGIC gene in human pleomorphic adenomas (Geurtsteins et al., 1998). The C-terminus of NFI-B is fused to HMGIC, and the abberant fusion protein is expressed in the affected tissue. Although the NFI-B fusion is found in only a small percentage of tumors, its presence suggests that abberant expression of this region of NFI-B may play a role in generation of the tumor. In contrast, overexpression of NFI proteins renders chick embry fibroblasts cells resistant to transformation by a number of nuclear oncogenes, including fos, jun and yin (Schuur et al., 1995). While the mechanism of resistance is not known, the finding that the cells are not resistant to transformation by some oncogenes that function in the cytoplasm suggests some specificity to the suppression of oncogenic susceptibility. Finally, overexpression of NFI-L is a common feature of the TGF-β signal transduction pathway and cell proliferation (Sun et al., 1998). Since NFI was first discovered as a protein required for viral DNA replication (see de Jong and van der Vliet, 1999 for a review), the finding that altered expression of NFI proteins may influence cell proliferation may indicate a direct role for NFI proteins in cellular DNA replication. The availability of systems that are deficient in, or overexpress, specific isoforms of NFI will allow further investigation into the role of NFI in cell proliferation.

12. Summary

While much information has been gathered regarding the role of NFI-binding sites and proteins in gene expression, it is difficult to put it all into a global...
perspective. This is because the control of tissue-specific gene expression during development is perhaps the most complex biological regulatory system known and is the basis for all metazoan development. Clearly, NFI-binding sites play essential roles in the expression of genes in multiple organs and tissues, and changes in the expression levels of various NFI isoforms likely result in important changes in patterns of gene expression. The widespread (though spatially and temporally unique) expression of the four NFI genes, together with the alternative splicing of NFI transcripts and ability of the proteins to homo- and heterodimerize, yields a large number of potential transcriptional activators and repressors. There are three paths that should yield the information needed to fully understand the role of NFI proteins in gene expression and development:

1. The powerful technologies of multi-gene cDNA (Schena et al., 1998) and oligonucleotide (Lockhart et al., 1996) arrays, together with both the identification and expression of specific NFI isoforms and the generation of mice and cell lines deficient in specific NFI gene products, should allow us to determine the important in-vivo targets of NFI regulation.

2. The use of simpler genetic and developmental systems that appear to contain only a single NFI gene (C. elegans, Drosophila and Amphioxus) may yield insights into the evolution of the NFI gene family and its functions.

3. The development of defined transcription assays that accurately reflect in-vivo functions, the use of in-vivo detection systems such as Chromatin Immunoprecipitation [ChiP (Orlando et al., 1997; Marzio et al., 1998; Waitelet et al., 1998; Cosma et al., 1999) and the analysis of transcription factor binding in vivo in living cells (Baumann et al., 1998; Fukushige et al., 1999; Walker et al., 1999)] should allow us to determine the basic biochemical mechanisms and pathways through which NFI proteins affect transcription (Fig. 3 bottom).

Acknowledgements

The author would like to thank Dr Christine Campbell for helpful discussions. This work was supported in part by National Institutes of Health grant HD34908 and National Science Foundation grant MCB-9612567 to R.M.G.

References


Blomquist, P., Li, Q., Wrang, O., 1996. The affinity of Nuclear Factor I for its DNA site is drastically reduced by nucleosome organiza-
tion irrespective of its rotational or translational position. J. Biol. Chem. 271, 113–119.


Bromick, E.H., John, S., Berard, D.S., LeFebvre, P., Hager, G.L., 1990. Glucocorticoid receptor-dependent disruption of a specific nucleosome on the mouse mammary tumor virus promoter is pre-

Bromick, E.H., Rories, C., Hager, G.L., 1992. Evidence that nucleo-

Brugger, N., Solge, L., Wittacker, L., Beato, M., 1990. Nuclear factor I acts as a transcription factor on the MMTV pro-
motor but competes with steroid hormone receptors for DNA bind-
ing. EMBO J. 9, 2233–2239.


Cardinais, J.R., Chapel, S., Wahli, W., 1999. Exon structure of the nuclear factor I DNA-

binding domain from C. elegans to mammals. Mamm. Genome 10, 390–396.

Caucal, A.C., Strock, P., Wehmann, J., Buterakat, P., Pesty, H., 1989. DNA sequences outside the receptor-binding site differently mod-
ulate the responsiveness of the mouse mammary tumour virus pro-
motor to various steroid hormones. EMBO J. 8, 1403–1410.


Chaudhury, A.Z., Lyons, C.E., Gronostajski, R.M., 1997. Expression patterns of the four Nuclear Factor I genes during mouse embryogen-
esis indicate a potential role in development. Dev. Dyn. 208, 313–325.

Chaudhury, A.Z., Vitale, A.D., Gronostajski, R.M., 1998. Nuclear factor I (NF-I) isoform differentially activate simple versus com-


Chavez, S., Beato, M., 1997. Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus pro-

Chen, M., Mermod, N., Horwitz, M.S., 1990. Protein–protein inter-
actions between adenosine DNA polymerase and nuclear factor I regulate formation of the DNA replication preinitiation complex. J. Biol. Chem. 265, 18 634–18 642.

Chung, C.M., Rooder, R.G., 1995. Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activa-

Chu, H.M., Fischer, W.B., Osborne, T.F., Comb, M.J., 1991. NF-I proteins from brain interact with the promethorphan cAMP induc-


de Jong, R.M., van der Vliet, P.C., 1999a. Mechanism of DNA replica-

protein acts as a bridging factor between NF-κappaB/Ral and nuclear co-regulators. Oncogene 18, 3316–3323.

Decker, J., van Oosterhout, J.A., van der Vliet, P.C., 1996. Two regions within the DNA binding domain of nuclear factor I interact with DNA and stimulate adenovirus DNA replication indepen-


Fischer, K.D., Haese, A., Nowock, J., 1993. Cooperation of GATA-1 and Sp1 can result in synergistic transcriptional activation or in-

Fletcher, C.F., Jenkins, N.A., Copeland, N.G., Chaudhury, A.Z., Gro-


specific DNA-binding proteins nuclear factor I from HeLa cells and the TGGCA protein from chicken liver. EMBO J. 5, 381–386.


Ortiz, L., Aza-Blanc, P., Zamm, M., Cato, A.C., Santisteban, P., 1999. The interaction between the forkhead thyroid transcription factor TTF-2 and the constitutive factor CTF


localization of the four genes NFIA, NFIB, and X for the human transcription factor Nuclear Factor X by FISH. Genomics 28, 66–73.


