FEBS Letters 390 (1996) 44–46 FEBS 17231

Identification of DNA binding-site preferences for nuclear factor I-A

Shigehiro Osada, Shoko Daimon, Tsutomu Nishihara, Masayoshi Imagawa*

Department of Environmental Biochemistry, Faculty of Pharmaceutical Sciences, Osaka University, 1-6 Yamada-Oka, Suita, Osaka 565, Japan

Received 25 April 1996; revised version received 27 May 1996

Abstract Nuclear factor I (NFI) proteins constitute a large family of DNA binding proteins. These proteins promote the initiation of adenovirus replication and regulate the transcription of viral and cellular genes. The binding sites for NFI have been reported in a wide variety of promoters, and they exhibit flexibility in their sequences. To clarify the DNA binding site of NFI-A, one of the NFI proteins, we performed a polymerase chain reaction-mediated random site selection, and determined the optimal sequence as 5′-TTGGCANNNN(G/T)CCA(G/A)-3′.

Key words: Nuclear factor I; Transcription factor; DNA binding site; Consensus sequence

1. Introduction

Nuclear factor I (NFI) was originally identified as a DNA binding protein that stimulates the replication of adenovirus DNA in HeLa cells [1], and it was later found that NFI also stimulates the transcription of these cells [2]. Different NFI cDNAs were cloned from human [3], hamster [4], rat [5], mouse [6], and chicken [7]. Sequence comparison and analysis of chicken genes led to the identification of four genes: *NFI-A*, *NFI-B*, *NFI-C*, and *NFI-X* [7]. In addition, each subtype possesses the different isoforms generated by alternative splicing [8]. All NFI proteins contain a highly conserved N-terminal amino acid sequence required for dimerization and DNA binding. The N-terminal 240 amino acids of the rat NFI-A protein, also called NF1-L, have been shown to be sufficient for DNA binding and dimerization in vitro [9], although the full-length cDNA of the rat NFI-A has not yet been cloned.

The NFI recognition sequence has been speculated to be the palindrome-like sequence TGG(A/C)N_5GCCAA by comparison of the sequences of high-affinity binding sites from a variety of viral and cellular genomes. Base substitutions in the symmetrical NFI binding site of the adenovirus origin showed lower affinity than did the native site [2]. Chicken NFI proteins have been purified from liver nuclear extracts as TGGCA protein species [10]. These observations indicate that NFI binds to both palindrome and half sites, although the sequences of NFI binding sites remain unclear.

To define the DNA binding site of transcription factors in an unbiased manner, the method of polymerase chain reaction (PCR)-mediated random site selection was recently used, and the consensus sequences for CCAAT/enhancer binding proteins, vitamin D receptor, and interferon regulatory factors have been determined [11–13]. Utilizing a binding-site selection procedure, we show here that NFI-A recognizes the palindrome-like sequence TTGGCANNNN(G/T)CCA(G/A) and also binds to the half-palindrome site TTGGCA.

2. Materials and methods

2.1. Expression and purification of bacterially expressed NFI-A

By using mouse NFI-A, originally termed NFI-B [6], we have cloned full-length NFI-A cDNA of the rat (unpublished). The DNA binding domain of rat NFI-A was expressed by a histidine fusion protein system (Qiagen Inc.). A *Bam*HI–*Sph*I fragment encoding 5–318 amino acids was subcloned into pQE-30 expression vector. The recombinant plasmid was transformed into M15 [pREP4]. The transformant was grown overnight at 30°C in Luria-Bertani (LB) agar plate containing 100 μg/ml ampicillin and 25 μg/ml kanamycin, inoculated into 100 ml of LB medium from the plate, and grown to $A_{600} = 0.7$, at which time isopropyl-β-D-thio-galactopyranoside was added to a final concentration of 1 mM. The cells were then allowed to grow for an additional 4 h at 25°C and harvested by centrifugation. Since we could not purify the recombinant protein by Ni^{2+}-nitrilotriacetic acid column, the NFI-A DNA binding domain was purified by a DNA affinity column with the NFI site, previously termed GPS4, in the glutathione transferase P (*GST-P*) gene [14].

2.2. Random binding-site selection

Random binding-site selection was performed as described previously [11] with a slight modification. In brief, 131 ng of double-stranded oligonucleotides containing 16 random nucleotides was used. After seven rounds of enrichment, the amplified products were labeled with incorporation [α-^{32}P]dCTP by PCR and used as a probe. The labeled probe was incubated with bacterially expressed NFI-A, and protein–DNA complexes were separated on a 6% native polyacrylamide gel. Bound DNA was eluted with a solution of 0.5 M ammonium acetate, 1 mM EDTA, 0.1% SDS, 10% methanol, and 50 μg/ml proteinase K. The eluate was phenolized and then amplified by PCR. The resultant fragments were subcloned into pBluescript KS+ (Stratagene), and the sequences were determined by the dideoxy chain termination method [15].

2.3. Gel shift analysis

Gel shift analysis was performed as described [11]. The sequences of oligonucleotides for gel shift assay were as follows (only upper strands are shown):

NFI-pal: 5′-CTAGCTATTTTGGCATCATGCCAATATG-3′

NFI-half: 5′-CTAGCTATTTTGGCATCACAGATCTATG-3′

NFI-adeno: 5′-CTAGCTATTTTGGATTGAAGCCAATATG-3′

NFI-GST: 5′-CTAGTTTCTTGGAGCAGGACCCAAAAAT-3′

*Corresponding author. Fax: (81) 6-879-8244.
E-mail: imagawa@phs.osaka-u.ac.jp

Abbreviations: NFI, nuclear factor I; PCR, polymerase chain reaction; GST, glutathione transferase P

PII S0014-5793(96)00622-9

3. Results and discussion

3.1. Selection of DNA binding sites for NFI-A

We selected the oligonucleotides which could be recognized by bacterially expressed NFI-A from the double-stranded oligonucleotides with 16 random bases. The starting amount of oligonucleotides was 131 ng (2×10^{12} molecules), which contains all possible combinations of the 16-nucleotide randomized region ($4^{16}\times1/2=2\times10^{9}$). Therefore, it seemed that the selected binding sites should be unbiased. The PCR-amplified DNA recovered after each round of selection was used as a competitor of gel shift analysis by the using the adenovirus NFI site as a probe, and the competition activity was found to increase with each selection (data not shown). DNA fragments after seven rounds of enrichment were cloned into a plasmid vector and sequenced.

Clones containing 111 kinds of DNA fragments were independently obtained, and the sequences of cloned oligonucleotides were optimally aligned by the inspection facilitated by the hexanucleotide sequence 5′-(T/C)TGGCA-3′ found in most sequences (Fig. 1). This motif, used to derive a preliminary consensus, was found in 61% of the selected oligonucleotides (N-1 to N-68). The remaining clones were aligned by using the 5′-(T/C)TGGCA-3′-like sequence as a guide. The clones N-110 and N-111, lacking a 5′-(T/C)TGG-3′ sequence, did not show any binding activity to NFI-A and also did not function as competitors (data not shown). We therefore omitted these two clones from determination of the consensus sequence.

Data on alignment of the binding-site selection from 109 independent clones were summarized according to the frequency of the nucleotide at each position, and the consensus binding site for NFI-A was also established (Fig. 2). The consensus sequence of the binding site obtained here, TTGGCANNNN(G/T)CCA(G/A), is similar to the complete palindrome sequence TTGGCANNNTGCCAA. These sequences also revealed that NFI-A could bind to palindrome or palindrome-like sequences composed of two TTGGCA half sites with three spacer nucleotides. To assess conserved sequences in the spacer region which may affect NFI binding, Gronostajski used two oligonucleotide pools which contain 5′-TGG-3′ and 5′-GCCAA-3′ motifs separated by a 6 or 7 bp spacer region. He found that the NFI binding consensus sequences are TGG(C/A)NNNNNGCCAA and TGGG(C/A)NNNNNGCCAA, derived from 6 and 7 bp spacer pools, respectively [16,17]. Invariant bases in the pools are underlined. Although the oligonucleotide pools contain invariant bases, these results resemble ours. Since he used purified NFI from HeLa cell nuclear extracts which is abundant in NFI-C, his result reflects the optimal binding site of NFI-C. This indicates that the recognition sites of NFI-A and NFI-C are indistinguishable, although it seems possible that there are some differences among the optimal binding sites of NFI proteins. We have begun to investigate the optimal binding site for other NFI members.

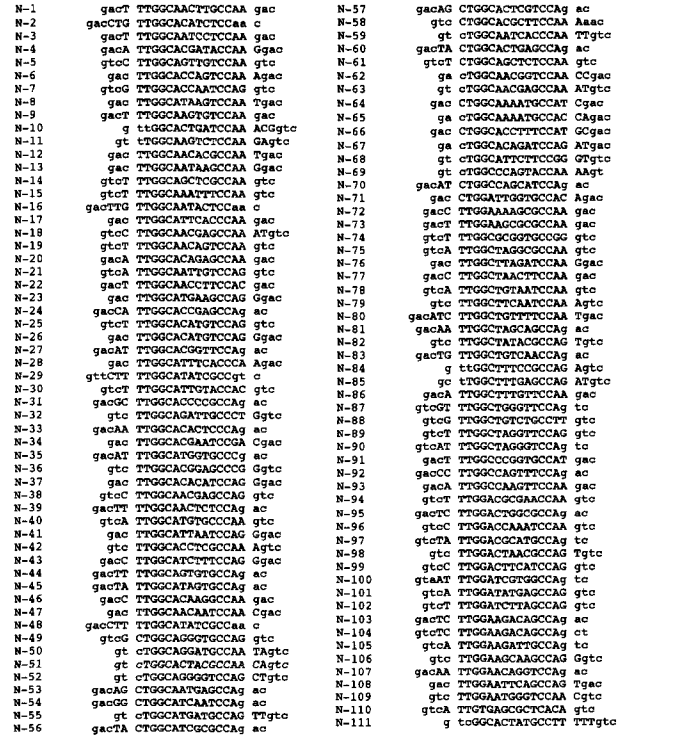

Fig. 1. The aligned sequences of oligonucleotides selected by NFI-A binding. The nucleotide sequences cloned after seven rounds of the selection are shown as aligned using the method described in the text. Sequences selected from the 16-base-long region of random nucleotides and its flanking sequences are shown by capital and small letters, respectively.

G	0	0	109	109	0	1	20	27	38	17	51	0	0	4	52
A	0	0	0	0	19	77	32	27	26	36	6	0	0	100	47
T	86	109	0	0	0	20	27	29	25	34	48	0	0	1	6
C	23	0	0	0	90	11	30	26	20	22	4	109	109	4	4
consensus	T	T	G	G	C	A	N	N	N	N	G/T	C	C	A	G/A
(palindrome	T	T	G	G	C	A	N	N	N	T	G	C	C	A	A)

Fig. 2. Summary of the selected sequences. Fifteen positions of the aligned oligonucleotides from Fig. 1 are shown with the number of occurrences in each position. A consensus sequence based on frequencies and a perfect palindrome sequence are indicated at the bottom.

Since we observed previously that the base preferences reflected the affinity of the DNA binding [11], it seems that both of the nucleotides of 5′-TGG-3′ and 5′-CCA-3′ are identified as the most critical determinants of recognition specificity and affinity. It is also likely that the side bases of both of 5′-TGG-3′ and 5′-CCA-3′ are the second most important determinant. The three bases at the center of the palindrome sequence are completely unbiased. The crystal structure of the GCN4 and AP-1 sites revealed that one of the bases near the center of the AP-1 site is not contacted by the basic and leucine zipper structure [18]. Since the unbiased three bases are located in the center, these bases may not be in contact with NFI-A. The consensus sequence speculated previously from a variety of viral and cellular genomes, TGG(A/C)N_5GCCAA, is confirmed by in our results.

3.2. Relative affinities of NFI-A for the consensus sequence and the native NFI binding sites

The oligonucleotides selected after seven rounds of enrichment contain palindrome or palindrome-like sequences. Although several NFI binding sites were identified, none of them perfectly coincided with the consensus sequence. However, some of the selected oligonucleotides after three rounds of enrichment had only one half site of the consensus sequence. To compare the affinity of several NFI sites, we per-

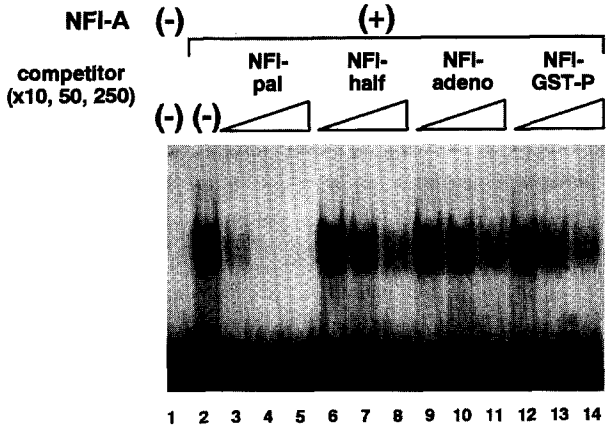

Fig. 3. Comparative in vitro binding studies among the palindrome, half, and native NFI sites. The labeled palindrome sequence was incubated without NFI-A (lane 1) or with NFI-A in the absence (lane 2) or presence of increasing amounts of the competitor (lanes 3–14: 10-, 50-, and 250-fold molar excesses).

formed a gel shift assay by using the oligonucleotide of the consensus sequence as a probe (Fig. 3). In the case of NFI-half, the sequence of one half site was TTGGCA but that of another half site was different from the TTGGCA sequence with the same composition. NFI-adeno is one of the highest-affinity binding sites among the NFI binding sites previously reported [2]. NFI-GST, originally termed GPS4, is the highest-affinity binding site in the GST-P silencer element consisting of multiple NFI and other transcription factors' binding sites [14]. As shown in Fig. 3, the perfect palindrome sequence was the most effective competitor in the NFI sites, whereas it was hard to distinguish the difference of affinity of the NFI-half site from that of the native sites.

Although the perfect palindrome sequence exists in the genomic DNA in all probability, such a sequence has not been found in the functional region in the genes. The highest binding activity might not be required for the regulation of replication and transcription reaction, or rather, the stringent binding specificity may be inconvenient. This means that the release as well as the binding of the factor to DNA is important to regulate biological functions. Using the filter binding method, the segments that bind tightly to NFI have been isolated and cloned from human genome DNA, although neither the sequences of binding sites nor the function of these segments have been characterized yet [19]. In vivo, NFI binding sites exist in or near other nuclear factor sites, and a functional cooperation has also been reported [20,21]. Therefore, experiments considering other factors will be required to clarify the functional NFI site in vivo.

Acknowledgements: This work was supported in part by grants from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] Nagata, K., Guggenheimer, R.A., Emoto, T., Lichy, J.H. and Hurwitz, J. (1982) Proc. Nat. Acad. Sci. USA 79, 6438–6442.
[2] Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) Cell 48, 79–89.
[3] Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) Nature 334, 218–224.
[4] 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.
[5] Paonessa, G., Gounari, F., Frank, R. and Cortese, R. (1988) EMBO J. 7, 3115–3123.
[6] Inoue, T., Tamura, T., Furuichi, T. and Mikoshiba, K. (1990) J. Biol. Chem. 265, 16065–19070.
[7] Rupp, R.A.W., Kruse, U., Multhaup, G., Göbel, U., Beyreuther, K. and Sippel, A.E. (1990) Nucl. Acids Res. 18, 2607–2616.
[8] Kruse, U. and Sippel, A.E. (1994) J. Mol. Biol. 238, 860–865.
[9] Gounari, F., De Francesco, R., Schmitt, J., van der Vliet, P.C., Cortese, R. and Stunnenberg, H. (1990) EMBO J. 9, 559–566.
[10] Rupp, R.A.W. and Sippel, A.E. (1987) Nucl. Acids Res. 15, 9707–9726.
[11] Osada, S., Yamamoto, H., Tsutomu, N. and Imagawa, M. (1996) J. Biol. Chem. 271, 3891–3896.
[12] Nishikawa, J., Kitaura, M., Matsumoto, M., Imagawa, M. and Nishihara, T. (1994) Nucl. Acids Res. 22, 2902–2907.
[13] Tanaka, N., Kawakami, T. and Taniguchi, T. (1993) Mol. Cell. Biol. 13, 4531–4538.
[14] Imagawa, M., Osada, S., Okuda, A. and Muramatsu, M. (1991) Nucl. Acids Res. 19, 5–10.
[15] Hattori, M. and Sakai, Y. (1986) Anal. Biochem. 152, 232–238.
[16] Gronostajski, R.M. (1986) Nucl. Acids Res. 14, 9117–9132.
[17] Gronostajski, R.M. (1987) Nucl. Acids Res. 15, 5545–5559.
[18] Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992) Cell 71, 1223–1237.
[19] Gronostajski, R.M., Nagata, K. and Hurwitz, J. (1984) Proc. Natl. Acad. Sci. USA 81, 4013–4017.
[20] Martinez, E., Dusserre, Y., Wahli, W. and Mermod, N. (1991) Mol. Cell. Biol. 11, 2937–2945.
[21] Chávez, S., Candau, R. Truss, M. and Beato, M. (1995) Mol. Cell. Biol. 15, 6987–6998.