A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif

Rukmini Kolluri, Ted Albert Torrey and Alan J.Kinniburgh*

Department of Human Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA

Received September 10, 1991; Revised and Accepted November 22, 1991

ABSTRACT

Numerous genes contain promoter elements that are nuclease hypersensitive. These elements frequently possess polypurine/polypyrimidine stretches and are usually associated with altered chromatin structure. We have previously isolated a clone that binds a class of CT-rich promoter elements. We have further characterized this clone, termed the nuclease-sensitive element protein-1, or NSEP-1. NSEP-1 binds both duplex CT elements and the CT-rich strand of these elements in a 'generic' sequence specific manner and has overlapping but distinct single- and double-strand DNA binding domains. The minimal peptide region sufficient for both duplex and single-strand DNA binding includes two regions rich in basic amino acids flanking an RNP-CS-1 like octapeptide motif. Deletion analysis shows that the single-strand DNA binding activity is mediated by the RNP-CS-1 like octapeptide motif and is the key peptide region necessary for single-strand binding. NSEP-1's affinity for CT rich promoter elements with strand asymmetry in addition to its double- and single-strand DNA binding properties suggests that it may be a member of a class of DNA binding proteins that modulate gene expression by their ability to recognize DNA with unusual secondary structure.

INTRODUCTION

Numerous genes contain *cis*-acting promoter elements that are nuclease hypersensitive. These regions are nucleosome free and are therefore hypersensitive to DNase I (1–5). Many of these elements are also sensitive to single-strand-specific nucleases. These dually sensitive elements have a strong purine/pyrimidine strand asymmetry and have been termed CT elements. The basis of these elements' nuclease sensitivities has been ascribed to an unusual DNA conformer. This conformer, termed the H-DNA form, is an intramolecular triplex/single-stranded structure. One of us (6) has previously shown that a CT promoter element of the *c-myc* gene can assume the H-DNA conformation under certain conditions *in vitro*. These conditions include low pH

and/or high levels of supercoiling which make it unlikely that H-DNA forms spontaneously *in vitro*. There is also an alternative intramolecular, triplex that may form with a G·G·C structure (1,1A). We have therefore worked under the hypothesis that cellular factors may induce these changes and produce a nuclease-sensitive conformer (H-DNA or G·G·C triplex) *in vivo*. We have previously isolated a clone that binds to a related set of CT promoter elements (7). Here we define the binding specificity of this factor, termed nuclease-sensitive element protein-1 or NSEP-1. We show that NSEP-1 binds both duplex CT elements and the CT-rich single strand of these elements in a 'generic' sequence-specific manner. That these are separable binding activities is demonstrated by the creation of mutations that disrupt single-strand binding but do not disrupt duplex CT element binding.

MATERIALS AND METHODS

Expression of NSEP-1 proteins and deletion derivatives

A full length cDNA clone was constructed in the pET expression vectors system (8,9) from sub-clones of the *Eco*RI fragments of the NSEP-1 phage clone in pUC13. Expression of this construct should produce a fusion protein containing 14 amino acids derived from the phage T7 gene 10 product of the expression vector, 6 amino acids from the pUC13 multiple cloning site *lacZ* coding sequence and 22 amino acids derived from the NSEP-1 5' untranslated region which is in frame with the methionine initiator codon (amino acid 1) of the 322 amino acid NSEP-1 protein. Deletion derivatives of this NSEP-1 expression fusion protein were engineered by restriction fragment deletion where convenient, or by polymerase chain reaction employing as primers synthetic oligonucleotides complementary to specific portions of the NSEP-1 coding sequence and containing restriction endonuclease recognition sites for sub-cloning into the expression vectors. Expression of the recombinant proteins in *E. coli* was performed as per Studier *et al.* (9). All deletion constructs refer to the amino acids of the *mature* proteins. All amino terminally deleted NSEP-1 derivatives lack the 22 amino acids of the 5' untranslated NSEP-1 region but retain the ϕ 10 gene and linker amino acids. Production of extracts of the recombinant protein-producing bacteria was by the procedure of Landschulz *et al.* (10).

* To whom correspondence should be addressed

Gel Retardation Assay

Recombinant proteins from lysates of bacterially expressed proteins were purified by fractionation on heparin-agarose columns (Figure 1) as described by Lichtsteiner *et al.* (11) except that the load and wash buffers contained 1X PBS with 5M urea. The procedure used was essentially that described previously (12) except that poly (dI-dC)·poly (dI-dC) was used at 1 µg/sample in place of the calf thymus DNA. All binding reactions employed ~1 ng of 5'-labeled oligonucleotide probe. Specific competition was tested by the addition of unlabeled probe prior to the addition of extract so that the recombinant NSEP-1 proteins were exposed to both target and competitor simultaneously. The amount of recombinant NSEP-1 extract was chosen so that the probe was in excess to the protein.

Southwestern Blotting

Cells from 1.5 ml aliquots of induced cultures were rapidly pelleted and resuspended with 100 µl aliquots of SDS-PAGE loading buffer. After heating at 100°C for five minutes, proteins in each sample (25 µl) were resolved by 7.5% SDS-PAGE. Proteins were transferred electrophoretically onto a nitrocellulose filter (0.45 µ) using 25 mM Tris, 190 mM glycine. After transfer, the filters were blocked with blotto overnight at 4°C and then washed twice at 22°C with TNE-50 (50 mM Tris, 50 mM NaCl, 1 mM EDTA, 1 mM DTT). To assay for DNA binding activity, the transfers were probed with ³²P-labeled oligonucleotides corresponding to the C-rich strand and duplex monomer of the *c-myc* NSE as per the procedure of Singh *et al.* (13).

RESULTS

Isolation of Nuclease-Sensitive Element Binding Protein cDNA Clones

We isolated several related cDNA clones by ligand screening with an oligonucleotide that is homologous to a *c-myc* promoter element. We had shown previously that this element is responsible for approximately 80% of the *c-myc* transcription in *c-myc*/CAT fusion constructs (12). We have reported the sequence for this cDNA and protein and discovered that our cDNA was 99% homologous to a relatively uncharacterized clone. This clone had been isolated using a restriction fragment that contained a nuclease-sensitive element of the EGF receptor gene promoter (14). We attribute this difference to sequencing errors and discuss our rationale for this elsewhere (7). Our first objective was to delineate the specificity of the interaction of NSEP-1 and CT elements in various genes.

NSEP-1 Binds to Related CT Promoter Elements in a 'Generic' Sequence-specific Manner

We first asked to which CT elements does NSEP-1 bind? Several end labeled oligonucleotides with different CT element sequences or unrelated but GC rich sequences were prepared and binding of NSEP-1 was assayed by the gel retardation assay (12). It can be seen that NSEP-1 bound the *c-myc* and EGFR nuclease-sensitive elements with similar avidities (Figure 2). One element with more divergent CT elements, Neu-1, binds NSEP-1 less well (Figure 2). Finally, a muscle-specific promoter element (MCAT), a GC-rich region of the Neu proto-oncogene (Neu-2) and an HLA gene promoter element (Y-box) do not bind NSEP-1 (Figure 2). The lack of binding to the Y-box element is of interest since a previous report (15) claims that NSEP-1 will bind the

Figure 1. Purification of recombinant NSEP-1 protein from *E. coli* extracts by heparin-agarose chromatography. Equivalent portions of the bacterial extract (load) and the fractions eluting from the column were resolved by electrophoresis on a 7.5% SDS-PAGE and stained with Coomassie Blue. Samples are as follows: lane 1, molecular weight markers (Mws); lanes 2 & 3, W.T. NSEP-1 (load); lane 4, flowthrough (F.T.); lane 5, wash; lane 6, 0.2 M KCl fraction; lane 7, 0.6 M KCl fraction; lane 8, 1.0 M KCl fraction.

Y-box element. We will also show that Y-box sequences cannot compete for the binding of the *c-myc* nuclease-sensitive element to NSEP-1 (see below).

To gather further information on the binding specificity of NSEP-1, we performed competition experiments. We utilized an end-labeled *c-myc* nuclease-sensitive element (NSE) oligonucleotide (labeled on the G-rich strand only) as a probe, and various unlabeled oligonucleotides as competitors for binding of NSEP-1 in an electrophoretic shift assay. As expected from the direct binding assay, the *c-myc* NSE and the EGFR CT element compete well for the binding of the labeled *c-myc* NSE probe (Figure 3). A CT element (Ki-ras NSE) with homology to these two elements competes, partially for the binding of the *c-myc* NSE probe (Figure 3). Likewise, the poorly binding Neu-1 element competes less well for *c-myc* NSE binding (Figure 3). Competition by oligonucleotides with mutated NSE sequences compete less well than the wild-type *c-myc* NSE oligonucleotide (Figure 3). Finally, two elements show very low levels of competition even at a 500-fold excess of unlabeled competitor (Neu-2 and Y-box) and two other elements (MCAT and a random oligonucleotide with the same base composition as the *c-myc* NSE but lacking any purine/ pyrimidine strand asymmetry) show no competition (Figure 3). These results are summarized in Table 1.

An inspection of the CT elements that bind NSEP-1 indicate that NSEP-1 does have a limited specificity for CT-rich, asymmetric DNA sequences (Table 1). The asymmetry is an important aspect of this. For example, the random sequence oligonucleotide has the same percentage of C and T bases but has no purine/ pyrimidine strand asymmetry and it cannot compete for DNA binding with the *c-myc* NSE oligonucleotide (Table 1, Figure 3). The purine/pyrimidine strand asymmetry

Figure 2. NSEP-1 recognizes CT rich elements. Synthetic oligonucleotides corresponding to the *c-myc* NSE, EGFR-NSE, and various other divergent CT elements (Table 1) were tested for their ability to form complexes with recombinant NSEP-1 protein. The assays were performed in the absence (odd numbered lanes) or presence of extracts (even numbered lanes).

Figure 3. Specificity of recombinant NSEP-1-dsDNA interactions. Binding of the recombinant NSEP-1 protein to the *c-myc* NSE oligonucleotide was competed for with increasing amounts (5, 20, 50, and 200-fold molar excess relative to the labeled target) of the wild-type and mutant *c-myc* NSE, EGFR-NSE and other divergent CT element oligonucleotides (Table 1).

is also important for H-DNA formation due to the nature of the triplex base-pairing (Hoogsteen base pairing occurs between pyrimidines in the third strand and purines in the Watson-Crick 'duplex'). The structure of all these CT-rich elements consists of perfect or near perfect mirror image or direct repeats. The *c-myc* NSE repeat is TCCCCACCC whereas the EGFR NSE repeat is CCTCCTCCT. We have, in fact, made randomers of T/C and have converted these to duplex T/C·A/G. We have found a significant proportion of these will bind to NSEP-1 (data not shown). Therefore, we conclude that NSEP-1 will bind C-T-rich duplexes with a high degree of purine/pyrimidine strand asymmetry. This DNA base composition and symmetry

TABLE 1

Oligomer Name	Oligomer Sequence 5'→3'	Single-stranded Oligomer binding competition	Double-stranded Oligomer binding competition
myc NSE-C	CCTTCCCGACCTCCGACCTCCG	+++	+++
myc NSE-G	GGGAGGCTGGGAGGCTGGGAGG	-	-
Mutant 11	CCTTCCCGACCCaaCGACCTCCG	N.D.	N.D.
	GGGAGGCTGGcttGGTGGGAGG	N.D.	N.D.
Mutant 13	CCTTCCCGAaaCTCCGaaCTCCG	N.D.	N.D.
	GGGAGctTGGGAGctTGGGAGG	N.D.	N.D.
Ki-ras NSE-C	CCTCCGCTCCCTCCCTCCCTCCGCG	+++	+++
Ki-ras NSE-G	CGAGGGCAAGGAGGAAAGGAGGAGG	-	-
EGFR-C	TGCTCCTCCTCCGATTCTCCTCCTCCTGCTGCT	++	++
EGFR-G	AGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGG	-	-
neu-1 Top	GCAGAAAGGAGGAGGAGGAGGAGGAGG	-	-
neu-1 Bottom	AGCCCTCCTCCTCCGACCTCCTCCTCCT	+++	+++
neu-2 Top	TTGAGGGAAGTATAAGAAATCAAGTCTCAAG	-	-
neu-2 Bottom	CTTCAGAACTTCACTTCTTATCTTCCCTCAA	++	++
MCAT Top	CGGGGCGGATTCTCGGGGGGGG	-	-
MCAT Bottom	GGGGGGCGGAAATCGGGGGGGG	-	-
Y box Top	ATTTTCTGTATGGCCAAAG	-	-
Y box Bottom	CTTTGCCCAATCAGAAAAAT	-	-
Random-1	CAGCTAGCGCTCGCTCCGAGCTGACCTCACGGA	-	-
Random-2	TGGCTAGCTCAGCTCGGAGCCAGGAGCTAGCTG	-	-
Monomer A-C	TTGAGGGAAGTATCCGACCCAGCTTGTGAAG	+	++
Monomer A-G	CTTCAGAACTGGGTGGGATCTTCTCTCAA	-	-
Monomer E-C	TTGAGGGAAGTATCAGCTCCGCTTGTGAAG	+	++
Monomer E-G	CTTCAGAGGAGGAGCTGATCTTCTCTT	-	-
(C/T) ₂₄	CGAGCTCGAGGATATCG(T/C) ₂₄ GATCCAGTACTTACA	+++	+++
poly(dC) ₁₂₋₁₈	(C) ₁₂₋₁₈	N.D.	-
poly(dT) ₁₂₋₁₈	(T) ₁₂₋₁₈	N.D.	-

Figures 4A,B. NSEP-1 recognizes single-stranded DNA probes. Gel retardation assay performed with recombinant NSEP-1 protein and various synthetic single-strand oligonucleotides (Table 1). The assay was performed in the absence (last lane in Figure 3A) or in the presence of extracts.

specificity we refer to as a generic sequence specificity in that NSEP-1 binds all of one type of DNA species.

Single Strand Binding Properties of NSEP-1

In the formation of intramolecular DNA triplexes, there is a portion of both DNA strands that is single stranded. The large kink caused by formation of the intramolecular triplex creates a steric hindrance to precise reformation of the duplex DNA after the region of intramolecular triplex. In the case of the *c-myc* NSE element, a large single strand region exists on both the C-rich strand and the G-rich strand when H-DNA forms. In the G·G·C triplex, the displaced C-rich strand is single-stranded. Because B DNA is not likely to spontaneously form non-B DNA conformers under intracellular conditions, these structures are likely to be stabilized or induced by cellular factors. Therefore, in preliminary experiments we examined NSEP-1's ability to bind

Figure 5. Specificity of recombinant NSEP-1-ssDNA interactions. Binding of the recombinant NSEP-1 protein to the C-rich strand of the *c-myc* NSE was competed for with increasing amounts (2, 5, 20 and 50-fold molar excess relative to the labeled target) of oligonucleotides corresponding to the C-rich strands of the *c-myc* and EGFR NSE regions, various CT elements and individual strands corresponding to both strands of elements that did not bind in their duplex form (Table 1).

each of the single strands of the *c-myc* nuclease-sensitive element. We found that NSEP-1 bound to the C-rich strand of the *c-myc* NSE oligonucleotide but not to the G-rich strand (data not shown). To examine the sequence specificity of this single-strand interaction we performed binding experiments with various single-stranded probes. Purified recombinant NSEP-1 binds to the C-rich strands of the EGFR, Neu-1 and Neu-2 elements but not to that of the Y-box element (Figure 4A). The higher band shift complexes are not observed at lower NSEP-1 concentrations and we therefore believe that NSEP-1 binds to multiple sites on these oligonucleotides. We will present evidence below to confirm this conclusion. NSEP-1 also binds the C-rich strand of the Ki-c-ras NSE oligonucleotide (data not shown and Table 1). We next examined G-rich strands and both strands of elements that did not bind in their duplex form (Figure 4B). We find no binding for the EGFR, Ki-c-ras, Neu-1 G-rich strand or for either strand of the MCAT or random oligonucleotides (Figure 4B). In addition NSEP-1 does not bind the G-rich strand of the Neu-2 or either strands of the Y-box element (data not shown and Table 1). We conclude that NSEP-1 does not bind single strand DNA

Figure 6. Summary of NSEP-1 mutants single and double-strand DNA binding properties. A schematic of NSEP-1 showing peptide motifs. Wild-type and deletion mutants of the NSEP-1 protein are shown. Each constructs name is followed by deleted amino acids in parentheses. Nucleic acid binding is summarized at the right. + binding observed; - no binding observed; +/- weak binding observed.

indiscriminately but rather has a marked preference for CT-rich DNA. To further define this single-strand binding specificity, we have performed competition experiments with various competitors. As in the direct binding assay, only CT-rich strands compete for the binding of purified, recombinant NSEP-1 (Figure 5). The *c-myc*, Ki-c-ras, and EGFR nuclease-sensitive elements C-rich strands compete equally well for *c-myc* NSE binding to NSEP-1 (Figure 5A). The Neu-1, CT randomer, and two monomer *c-myc* NSE C-rich strands compete approximately 1/5 as well as the above group (Figure 5). The monomer sequences are two permutations of the *c-myc* NSE repeat embedded in the 25 bp random oligonucleotide (Table 1). This demonstrates that the monomer sequences contain all the information for single strand DNA binding. In experiments with direct binding of NSEP-1 to the monomer sequences, little or no multimer is observed (data not shown). We believe this indicates that the multimer is formed by multiple NSEP-1 proteins binding to multiple sites on the nucleic acid and not by protein-protein association of two NSEP-1 molecules. The Neu-2 C-rich strand shows a 50-fold lower level of competition than the parental *c-myc* NSE C-rich strand (Figure 4B). Finally, the MCAT bottom strand, the Neu-2 top strand, either strands of the Y-box and random oligonucleotides do not compete for binding at >200-fold molar excess (Figure 5; Table 1). We have also used poly(dT) and poly(dC) as competitors and do not observe competition with either of these DNAs, even at a 50-fold molar excess (data not shown). We therefore conclude that NSEP-1 is a generic, sequence specific single-strand DNA binding protein as well as having duplex DNA binding properties. NSEP-1 binds single-strand sequence elements that are CT-rich but not pure dT or dC (Table 1).

Defining the Single-strand and Double-strand Binding Motifs of the Nuclease-Sensitive Element Binding Protein-1

NSEP-1 binds both duplex and single-strand DNA in a 'generic' sequence-specific manner. There are 7 regions of distinctive amino acid composition within the NSEP-1 polypeptide (Figure 6). There are 3 regions rich in basic amino acids, a region rich in acidic amino acids, a glutamine-rich region which overlaps the acidic region, a proline/serine/threonine-rich region and an

Figure 7. Deletion analysis with bacterially expressed proteins for identification of the double-strand and single-strand DNA binding domains. **A.** Wild-type and mutated forms of NSEP-1 were synthesized in bacteria with the use of the phage T7 expression system as per the procedure of Studier *et al.* (9). The bacterially expressed proteins were size fractionated on a 7.5% SDS-PAGE after boiling in SDS-PAGE loading buffer. The fractionated proteins were visualized by Coomassie Blue staining (top row). Samples are as follows: lane 11, vector control; lane 10, wild-type NSEP-1; lane 9, I-O (Δ a.a. 309–322); lane 8, Δ 135 (Δ a.a. 1–24); lane 7, Δ Sma1 (Δ a.a. 1–38); lane 6, Δ NH₂ (Δ a.a. 1–48); lane 5, Δ 258 (Δ a.a. 1–67); lane 4, Δ 288 (Δ a.a. 1–75); lane 3, Δ Sal1 (Δ a.a. 201–322); lane 2, Δ RNPCS (Δ a.a. 1–48; Δ 70–75; Δ 309–322); lane 1, Δ NH₂-2 (Δ a.a. 1–57). Southwestern analysis was performed on lanes 1–11 following electrotransfer to nitrocellulose. The filter was probed with ³²P-labeled double-strand oligonucleotide corresponding to the c-myc NSE region (middle row) and ³²P-labeled single-strand oligonucleotide corresponding to the C-rich strand of the c-myc NSE (bottom row). **B.** Southwestern analysis of bacterially expressed proteins corresponding to lanes 2, 6, 10 and 11 in Figure A. Top row: Coomassie Blue staining; Middle row: Southwestern analysis with ds c-myc NSE; Bottom row: Southwestern analysis with C-rich strand of c-myc NSE.

octapeptide region with homology to the ribonucleoprotein consensus sequence 1 (RNP-CS-1) motif. We wished to examine the protein motif(s) that specify both the double-strand and single-strand binding activities. To this end we prepared wild-type, truncated, and deleted NSEP-1 constructs using the T7 polymerase, T7 ϕ 10 gene promoter system of Studier (8,9). The NSEP-1 constructs were expressed in the appropriate host bacteria and cell extracts were analyzed by SDS-PAGE and Southwestern blotting. Ten constructs were analyzed using the c-myc NSE duplex probe. Blots were stripped of duplex probe by denaturation in guanidine-HCl and re-probed with the C-rich strand of the c-myc NSE. In these experiments, the films are purposely overexposed to ensure that binding and/or the lack of binding is scored correctly. However, by overexposing these Southwestern blots, we detect **minor**, proteolytic cleavage

products generated *in vivo*. The overexposure of the autoradiograph, however, exaggerates their levels.

The results of these Southwestern blots define two sets of overlapping yet distinct peptide motifs that are necessary for duplex and single-strand binding (Figure 6). We find that the Pro/Ser/Thr-rich region, the acidic region, the gln-rich region and the basic region 3 are *not* necessary for either double- or single-strand DNA binding (Figures 6 and 7A). On the other hand, basic-1, basic-2 and the intervening amino acids are apparently necessary for both duplex and single-strand binding (Figures 6 and 7). A few amino acids may be removed from the basic-1 region and still maintain duplex binding but not single-strand binding (see NH₂-2 construct; Figures 6 and 7). This small deletion of basic-1 allows double strand binding only if there is no deletion of COOH amino acids (Figure 6; compare the NH₂-2 construct with the M2 construct). Perhaps the basic-3 region near the COOH-terminus can replace a basic-1 region function such as stabilizing NSEP-1 phosphate interactions. These differences notwithstanding, the key peptide region in single-strand binding is the RNP-CS-1 region. When the RNP-CS-1 region is deleted, duplex NSE binding is maintained but single-strand DNA binding is not observed (Figures 6 and 7A,B). The minimal peptide region sufficient for both duplex and single-strand binding (construct M1) includes the basic-1, RNP-CS-1, amino acids 75–140 and the basic-2 region (Figures 6 and 7).

DISCUSSION

DNA binding properties of nuclease-sensitive element-binding protein 1

We have characterized a CT element or nuclease-sensitive element-binding protein (NSEP-1) utilizing recombinant wild-type protein as well as mutant proteins. We have found that NSEP-1 binds to a family of CT-rich elements having the property of strong purine/pyrimidine strand asymmetry. The binding of NSEP-1 is therefore 'specific' in that it binds a 'type' of DNA element to the exclusion of other sequences. We term this DNA binding specificity 'generic' sequence specificity.

NSEP-1 also binds the C-rich strand of CT or nuclease-sensitive elements. As in the case of double-strand binding, related sequence elements that are rich in CT show specific binding but not the G-rich strand of these elements or other non-C-rich single-strand oligonucleotides. One of the single-strand oligonucleotides tested, MCAT, has been shown to bind to other transcription factors (16). We therefore conclude that there are at least two classes of single-strand/double-strand binding proteins. Substituting A for C or T bases reduces NSEP-1 binding to these elements. Poly(dT) and poly(dC) do not compete for the binding of the c-myc nuclease-sensitive element, further defining the nucleotide specificity to CT rich but not pure poly(dC) or poly(dT) elements. Therefore, we conclude for the single-strand binding, as we have for the double-strand binding that the single-strand binding has generic sequence specificity.

Possible Relationships between the Binding Properties of the Nuclease-Sensitive Element-Binding Protein and Alternate DNA Topologies of CT-rich Sequence Elements

DNA sequence elements that are CT-rich and have a strong purine/pyrimidine strand asymmetry can assume alternate DNA conformations under certain conditions (6,17–20). One of these alternate, non-B DNA conformation has been termed H-DNA. H-DNA is an intramolecular triplex, single stranded structure. The

purine-rich strand of the Watson-Crick base-pairs can Hoogsteen base pair with a portion of the pyrimidine-rich strand thus forming a triplex. The triplex forms *in vitro* in supercoiled DNA at low pH. The dependency on low pH is due to the need for a proton in the formation of the C+G Hoogsteen base pair. This can be overcome by high levels of supercoiling. Alternatively, these same elements can form a G·G·C triplex at neutral pH in the presence of Mg²⁺. We have previously predicted that the proteins which bind these CT rich promoter elements that have a high degree of purine/pyrimidine strand asymmetry can either induce alternate, non-B DNA structures or stabilize them. That nuclease-sensitive elements should be considered as *in vivo* structures comes from several pieces of information. 1. Some of these elements are DNase I sensitive in isolated nuclei indicating that these elements exclude nucleosomes. 2. DNase I sensitivity can be eliminated by changes in gene activity. 3. In the case of the *c-myc* NSE element, the factors that bind this element disappear coincidentally with the change in nuclease sensitivity. Taken together, we hypothesize that the factors that bind these nuclease-sensitive, CT-rich structures help form non-B DNA structure(s) *in vivo*. As stated above, the *c-myc* nuclease-sensitive element's, DNase I sensitive structure is correlated with *c-myc* activity and the nuclease-resistant structure is found to correlate with *c-myc* down-regulation (12).

We believe that the single-strand DNA binding properties of NSEP-1 are consistent with the *c-myc* nuclease-sensitive element's ability to form a non-B DNA conformer. We envision that an NSEP-1 molecule can bind duplex NSE and 'capture' or bind the C-rich strand of the NSE as it 'breathes' or is denatured under torsional stress. The non-B DNA conformer (whichever it may be) may then form spontaneously or other factors may bind and further stabilize/induce triplex formation.

If the scenario we discuss above occurs, then one may ask why elements with poor H-DNA forming potential bind more vigorously to NSEP-1 than elements that are more likely to form H-DNA (for example the EGFR NSE has a more stable H-DNA structure than the *c-myc* NSE). We believe there are several reasons to explain these data. First, the ability to form an H-DNA structure under conditions of low pH *in vitro* may have no bearing on *in vivo* H-DNA formation, especially if sequence-specific DNA binding proteins are necessary for H-DNA formation or stability. Second, one may speculate that DNA elements with poor H-DNA forming potential would need a greater affinity for DNA binding proteins that would induce or stabilize the H-DNA structure because of the poor energetics. On the otherhand, elements that could form H-DNA more readily might not need the stabilizing effects of a DNA binding protein and therefore would need to bind these proteins only weakly or not at all. In any case, we may find that there is a family of CT element binding proteins with overlapping affinities for these elements, and that H-DNA formation at any one element is more a matter of cell physiology than chemical stability.

ACKNOWLEDGEMENTS

We thank Drs. J. Ross and W. Held for many helpful comments during the course of this work. We also thank Dr. Simon Williams for materials and advice on their use. We thank Donna Rypinski for technical assistance and Nancy Frame for secretarial assistance. This work was supported by a grant from the NIH, CA 43661, and a Cardiovascular Discovery Award from Glaxo, Inc.

REFERENCES

1. Cooney, M., Czernuszewicz, G., Postel, E.H., Flint, S.J. and Hogan, M.E. (1988) *Science*, **241**, 456-459.
- 1A. Kohwi, Y. and Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 3781-3788.
2. Johnson, A.C., Junno, Y. and Merlino, G.T. (1988) *Mol. Cell. Biol.* **8**, 4174-4184.
3. Larsen, A. and Weintraub, H. (1982) *CELL*, **29**, 609-622.
4. McGhee, J.D., Wood, W.T., Dolan, M., Engel, J. and Felsenfeld, G. (1981) *CELL*, **27**, 45-55.
5. Wu, C., Bingham, P.M., Livak, K.J., Holmgren, R. and Elgin, S.C.R. (1979) *CELL*, **16**, 797-806.
6. Kinniburgh, A.J. (1989) *Nucl. Acids. Res.*, **17**, 7771-7778.
7. Kolluri, R.V. and Kinniburgh, A.J. (1991) *Nucl. Acids. Res.* (In Press).
8. Rosenberg, A.H., Lade, B.N., Chui, D.-S., Lin, S.-W., Bunn, J.J. and Studier, F.W. (1987) *Gene*, **56**, 125-135.
9. Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods in Enzymology*, **185**, 60-89.
10. Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1989) *Science*, **243**, 1681-1688.
11. Lichtsteiner, S., Wuarin, J. and Schibler, U. (1987) *CELL*, **51**, 963-973.
12. Davis, T.L., Firulli, A.B. and Kinniburgh, A.J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9682-9686.
13. Singh, H., LeBowitz, J.H., Baldwin, A.S. and Sharp, P. (1988) *CELL*, **52**, 415-423.
14. Sakura, H., Maekawa, T., Imamoto, F., Yasuda, K. and Ishii, S. (1988) *Gene*, **73**, 499-507.
15. Didier, D.K., Schieffenbauer, J., Woulfe, S.L., Zacheis, M. and Schwartz, B.D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7322-7326.
16. Santoro, I.M., Yi, T.M. and Walsh, K. (1991) *Mol. Cell. Biol.*, **11**, 1944-1953.
17. Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D. (1985) *J. Biomol. Struc. Dyn.*, **3**, 327-338.
18. Hanvey, J., Shimizu, M. and Wells, R.D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6292-6296.
19. Johnston, B.H. (1988) *Science* **241**, 1791-1795.
20. Htun, H. and Dahlberg, J.E. (1989) *Science* **243**, 1571-1575.