Vol. 185, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 29, 1992 Pages 264-270

THE IDENTIFICATION OF A TANDEM H-DNA STRUCTURE IN THE c-MYC NUCLEASE SENSITIVE PROMOTER ELEMENT

Anthony B. Firulli, Douglas C. Maibenco and Alan J. Kinniburgh*

Departments of Molecular Diagnostics and Human Genetics
Roswell Park Cancer Institute
Elm and Carlton Streets
Buffalo, NY 14263

Received April 13, 1992

Summary: Previous studies have shown that the c-myc nuclease sensitive element (NSE) is capable of forming H-DNA in vitro. The NSE sequence exhibits strong purine/pyrimidine strand asymmetry. To study the NSE further, we have isolated the element from other c-myc sequences and have shown that the NSE alone is sufficient for the formation of H-DNA in supercoiled plasmids. We also show that the NSE forms a complex structure containing both H-y3 and H-y5 H-DNA. We term this structure tandem H-DNA.

DNA sequences that have strong purine/pyrimidine strand asymmetry can form H-DNA in vitro at acid pH when negatively supercoiled (1-12). These elements have been found in the 5' promoter regions of several eukaryotic genes as well as in the enhancer regions of a number of eukaryotic DNA viruses such as SV 40 (12). One such element is the c-myc nuclease sensitive element (NSE). The NSE has been shown to: 1) bind a number of factors from HL-60 cell nuclear extracts, 2) be involved in the transcriptional regulation of the c-myc gene, and 3) form H-DNA in vitro (9-10).

One of us (A.J.K.), had previously analyzed the c-myc NSE for its ability to form H-DNA (9). Data from in vitro chemical modification experiments was used to construct a model of the H-DNA structure formed by the NSE. In the original model,

*To whom correspondence should be addressed.

Abbreviations: NSE, nuclease sensitive element; PCR, polymerase chain reaction; BRL, Bethesda Research Laboratory; NSEP-1, nuclease sensitive element protein-1; DMS, dimethyl sulfate; DEPC, diethyl pyrocarbonate; H-y5 H-DNA, an H-DNA structure in which the 5' end of the purine rich strand participates in forming the Hoogsteen basepairs; H-y3 H-DNA, an H-DNA structure in which the 3' end of the purine rich strand participates in forming the Hoogsteen basepairs; nt, nucleotide.

0006-291X/92 $4.00

there is a 38 nucleotide (nt) region of non-B DNA in this H-DNA structure. 22 bases of this region (on the G-rich strand) are single-stranded. 3' to this single-strand region duplex DNA reforms. One peculiarity of this model is the long single-strand region found 3' of the triplex. To investigate further the large single-strand region of the NSE, we isolated the element from the larger c-myc promoter context to test the element's ability to form H-DNA independent of adjacent c-myc sequences. Our findings demonstrate that 59nt of the NSE sequence is sufficient to form H-DNA in supercoiled plasmids. More interestingly, we observe a new, H-y5 H-DNA structure in the 3' end of the element that incorporates 8nt that were originally thought to be single stranded. We term this novel H-y3 triplex/H-y5 triplex structure, tandem H-DNA.

Materials and Methods

Polymerase Chain Reaction (PCR), Cloning, Plasmid Preparation, Chemical Modification and Gel Electrophoresis: PCR of the NSE was carried out as described (13). The NSE PCR product was digested with BamH I and Bgl II (BRL) and cloned into the Bgl II cloning site of p-35AC-1 (a generous gift of Dr. Peter Johnson). Plasmid DNA was purified from E. coli by the method of Holmes and Quigley (14). Ten micrograms of pNSEAC-1 was treated with 0.5 microliters of DMS for 30 seconds or 10 microliters of DEPC for 20 minutes in the appropriate pH buffer (6, 9, 15). The plasmid was then linearized with Hind III, 5' or 3' end labeled with ^{32}P-γ-ATP or ^{32}P-α-ATP, and cut with Xba I. The labeled, smaller DNA fragment was purified by electrophoresis through a 2% agarose gel, electroeluted, cleaved with piperidine and electrophoresed through a 10% polyacrylamide gel as described (9).

Results and Discussion

Chemical Probing of the NSE: In order to examine the structural properties of the NSE, the element was amplified from c-myc sequences using PCR, and then cloned as a monomer (data not shown). By removing the NSE element from its endogenous location, secondary effects of flanking c-myc sequences on the ability of the NSE to form H-DNA could be determined.

We first reacted the NSE containing plasmid, pNSEAC-1, with DMS and DEPC at acid and neutral pH. Under conditions of acidic pH and negative supercoiling, sequences with purine/pyrimidine strand asymmetry can form H-DNA (1-9). DMS and DEPC are reactive with the N-7 of purines, DMS G>>A and DEPC A>G. DEPC is much more reactive with single stranded DNA while DMS is reactive with single stranded and duplex DNA. When DNA assumes an H-DNA conformation, the N-7 and C-6 positions of Watson and Crick base paired purines are hydrogen bonded to the pyrimidine-rich strand through the N-3 and C-4 enol. Thus triplex formation will block DMS access to the N-7 of purines and result in hypomethylation of these bases. The

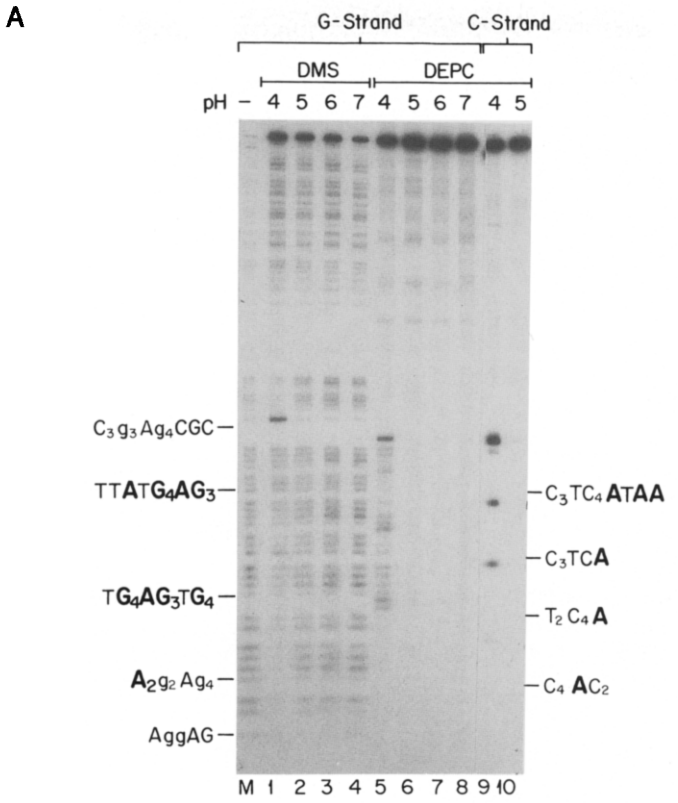
strand complementary to the third strand is single stranded in the H-DNA conformer and thus purines in this strand should be hyper-reactive with DEPC.

When the G-rich strand of the NSE was assayed after treatment with DMS at pH 4, there were two regions of hypo-reactivity observed (Figure 1A). The first, previously described (9), is located in the 5' end of the NSE at positions 8, 9, 11-14, 16, and 17 (Figure 1A, lane 1). Comparison of the same bases reacted with DMS at pH 5-7 showed uniform DMS reactivity, indicating that the NSE exists as a duplex (Figure 1A lanes 2, 3, 4). A second region of DMS hypo-reactivity was observed in the 3' end of the NSE. Gs 49-52 and 54-56 were hypo-reactive when compared to DMS reactions carried out at pH 5-7 (Figure 1A compare lane 1 with lanes 2, 3, 4). Interestingly, there was a DMS hyper-reactive G at position 47 at pH 4. The reason for this hyper-reactivity is unclear, however it was not observed at pH 5-7. These data support the idea that a second, previously unidentified triplex exists in the 3' end of the NSE. This second triplex was not observed in the previous study because the site of end label was within 5 base pairs of this triplex and the region was not well resolved when analyzed by gel electrophoresis (9).

To test this second triplex hypothesis further, DEPC was used to probe the G-rich strand of the NSE at pH 4. A long stretch of hyper-reactive purines is observed. These purines occur at positions 18-23, 25-32, and 34-41, and 43. These hyper-reactive bases establish the existence of a single stranded region (Figure 1A,lane 5). When DEPC reactions were carried out at pH 5,6 and 7, no hyper-reactivity was observed in the NSE region. This large single stranded region is consistent with a second triplex at positions 49 to 56, as suggested by the DMS reactivity.

To confirm this interpretation, we examined the DEPC reactive pattern of the C-rich strand, since the reactivity of the A bases that punctuate the C-rich strand should be consistent with this second triplex. We observed a pattern of hyper-reactive As at positions 24, 33, 42, 44, and 45 and a slightly hyper-reactive A at position 15 (Figure 1A, lane 9). At pH 5, 6 and 7 the hyper-reactivity was no longer observed (Figure 1A, lane 10; and data not shown). This demonstrates that the single strand regions exist only under conditions that favor the H-DNA conformer. This is consistent with the DMS and DEPC results of the G-rich strand. We feel that the slightly reactive A at position 15 is similar to the hyper-reactive G at base 47.

We have summarized the DMS and DEPC chemical probing of the NSE (Figure 1B). The data are most consistent with a tandem H-DNA structure (Figure 2). In this model, Hoogsteen base pairing occurs between nucleotides 22, 23, and 25-30 of the C-rich strand and nucleotides 8-10, 12-14, 16, and 17 of the G-rich strand (Figure 2). The



B **C-STRAND NSE CHEMICAL MODIFICATION**

```
   1   5   10   15   20   25   30   35   40   45   50   55
                   ^         ^         ^        ^ ^^
5' TGAGTCTCCTCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAAGCGCCCCTCCCGGG 3'
3' ACTCAGAGGAGGGGTGGAAGGGGTGGGAGGGGTGGGAGGGGTATTCGCGGGGAGGGCCC 5'
```

G-STRAND NSE CHEMICAL MODIFICATION

```
   1   5   10   15   20   25   30   35   40   45   50   55
5' TGAGTCTCCTCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAAGCGCCCCTCCCGGG 3'
3' ACTCAGAGGAGGGGTGGAAGGGGTGGGAGGGGTGGGAGGGGTATTCGCGGGGAGGGCCC 5'
      vv vvvv vvvvvvvv ^^^^^^^^ ^^^^^^^^ ^    * vvvv vvv
```

Figure 1. A) Dimethyl sulfate (DMS) and Diethylpyrocarbonate (DEPC) modifications of pNSEAC-1. Ten micrograms of super-coiled pNSEAC-1 was reacted with DMS or DEPC at the various pH values as described. The plasmid was then digested with Hind III, endonuclease, phosphatased, and either 3' end labeled (G-rich strand) or 5' end labeled (C-rich strand) as described. M, DMS marker, lane 1,DMS,G-rich Strand, pH 4; lane 2,DMS, G-rich strand, pH 5; lane 3,DMS, G-rich strand, pH 6; lane 4,DMS, G-rich strand pH 7; lane 5,DEPC, G-rich strand, pH 4; lane 6,DEPC, G-rich strand pH 5; lane 7,DEPC, G-rich strand, pH 6; lane 8,DEPC, G-rich strand pH 7; lane 9,DEPC C-rich strand pH 4; lane 10,DEPC, C-rich strand, pH 5.

B) Summary of the chemical modification data obtained on pNSEAC-1 from A. v, Hypo-reactive to DMS, *, hyper-reactive to DMS, ^ hyper-reactive to DEPC.

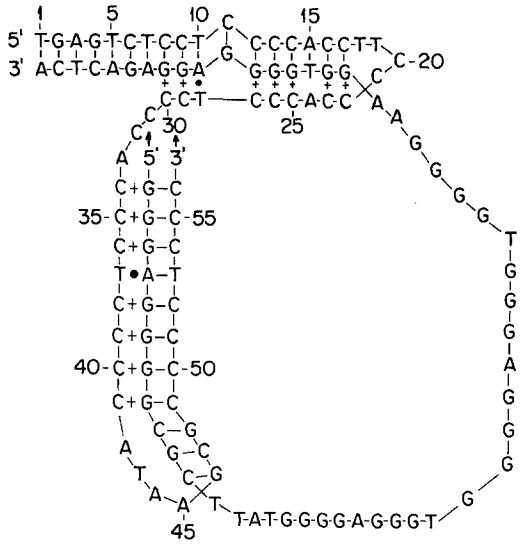


Figure 2. Tandem H-DNA structure of the c-myc NSE. An H-DNA structure was drawn using the model of Hoogsteen base pairing proposed by Frank-Kamenetskii and co-workers. Watson and Crick base pairs are represented by dashes, TAT Hoogsteen base pairs by dots and C+GC Hoogsteen base pairs by "plus" signs.

3' end of the G-rich strand accepts the third strand of the triplex. This H-DNA form is present as a H-y3 H-DNA isomer. The second H-DNA region is formed with nucleotides 34-41 of the C-rich strand which are Hoogsteen base paired with nucleotides 49-56 of the G-rich strand (Figure 2). Although there is no direct chemical evidence that the A at position 53 is involved in a Hoogsteen base pair, it should form a Hoogsteen base pair with the T residue at position 37 on the G-rich strand. If the model is correct, the A at base 53 should not be hyper-reactive with DEPC, because, to be involved in a triplex, the base must be Watson and Crick base paired. The DEPC data show that the A at position 53 is not hyper-reactive and thus not single stranded. In the second H-DNA structure, the 5' end of the G-rich strand accepts the third strand of the triplex making this structure a H-y5, H-DNA isomer. Due to the unique nature of this H-y3/H-y5 H-DNA structure, we term this structure tandem H-DNA.

We have shown that the c-myc NSE forms an H-y3 triplex H-y5 triplex tandem H-DNA structure (Figure 2). This model extends our previous H-y3 triplex/single strand model of the NSE (9). Certain data obtained previously were inconsistent or difficult to explain in the original model. We can now reconcile these discrepancies. For example, the long single strand portion of the G-rich strand found in both studies, is now seen to be due to the participation of the C-rich strand in two DNA triplexes.

Modifications to the C-rich strand of the NSE that are inconsistent with the single H-y3 DNA model, but fit the tandem model are also found. From previous work, the T residue at position 37 of the C-rich strand was found to be non-reactive with the single stranded T reactant, osmium tetroxide, at acid pH (9). In the single H-DNA model, this nucleotide should be hyper-reactive due to its location in a single stranded region. In the tandem H-DNA model, the T at position 37 is directly involved in a Hoogsteen base pair with the A at position 53 of the G-rich strand, thus its lack of reactivity is predicted only by the tandem H-DNA model. Further reconciliation of the previous data with the tandem H-DNA model is observed with the As at position 42, 44, and 45 (9). These are found to be hyper-reactive with DEPC under acidic conditions. The single H-DNA model does not account for the single-strand character of the two strands 3' to the first H-DNA structure. The tandem H-DNA model predicts that these bases should be hyper-reactive because they are found in the single strand loop of the C-rich strand required to form the H-y5 H-DNA triplex in the 3' end of the NSE.

Observations by Htun and Dahlberg as well as others have shown that in H-DNA forming DNA, the H-y3 H-DNA isomer is the dominant structure. This is due to its ability to release more torsional stress in supercoiled plasmids than the H-y5 isomer (8). In the tandem H-DNA structure of the NSE, an obvious H-y5 H-DNA structure is seen (Figure 1A, 2A). However, from the chemical modification data, it appears that these two H-DNA structures are not competing with each other as simple isomers would. It is therefore likely that the two H-DNA structures of tandem H-DNA exist within the same plasmid. That these and not the "anti-tandem" H-DNA structure forms (H-y5/H-y3), indicates that the tandem structure relieves more torsional stress. The initial nucleation event probably occurring in the H-y3 H-DNA structure since this releases more torsional stress. The second triplex may then form, further releasing the torsional stress of the plasmid. Since equal intensities of chemical modification are observed in each H-DNA structure, it seems likely that the majority of plasmid DNA molecules contain sufficient energy to form both H-DNA structures. From an energetic standpoint, the tandem structure should be favored.

The role of tandem H-DNA in the regulation of the c-myc gene is at present conjectural. The c-myc NSE is a positive transcription element and has been shown to bind a number of factors from HL-60 cell nuclear extracts. Some of these factors may be involved in the formation of the tandem H-DNA in vivo. A cDNA clone, NSEP-1, has been isolated from an expression library using an NSE oligonucleotide as a probe (16). The recombinant NSEP-1 protein has been shown to have double-strand and C-rich strand (single-strand) DNA binding activities. The single strand binding of the

NSE by a recombinant factor supports the notion that the NSE can form a non-B DNA structure in vivo (17).

Acknowledgments

We wish to thank Drs. Simon Williams and Peter Johnson for materials and helpful discussions. We also wish to thank Donna Rypinski for excellent technical assistance and Nancy Frame for excellent secretarial assistance. This work was supported by grants to A.J.K.; CA 41661 and a Cardiovascular Discovery Award from Glaxo, Inc. D.C.M. was supported by a Surgical Oncology Fellowship Award from the N.I.H. (T32 CA 09581).

References

1. Lyamichev, V.I., Mirkin, S.M., and Kamenetskii, F. (1985) J. Biomol. Struct. Dyn. 3, 227-338.
2. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A., and Kamenetskii, F. (1987) Nature 330, 495-497.
3. Lymanichev, V.I., Mirkin, S.M., Kamenetskii, F., and Cantor, C.R. (1988) Nucl. Acids Res. 16, 2165-2178.
4. Voloshin, O.N., Mirkin, S.N., Lyamichev, V.I., Belotserkovskii, B.P., and Kamentskii, F. (1988) Nature 333, 475-476.
5. Hanvey, J., Shimizu, M., and Wells, R.D. (1988) Proc. Natl. Acad. Sci. USA 85, 6292-6296.
6. Htun, H., and Dahlaberg, J.E. (1988) Science 241, 1791-1795.
7. Johnston, B.H. (1988) Science 241, 1800-1804.
8. Htun, H., and Dahlberg, J.E. (1989) Science 243, 1571-1575.
9. Kinniburgh, A.J. (1989) Nucl. Acids Res. 17, 7771-7778.
10. Davis, T.L., Firulli, A.B., and Kinniburgh, A.J. (1989) Proc. Natl. Acad. Sci. USA 86, 9682-9686.
11. Hanvey, J.C., Shimizu, M., and Wells, R.D. (1989) J. Biol. Chem. 264, 5950-5956.
12. Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, M., and Wohlrab, F. (1988) FASEB. 2, 2939-2949.
13. Innis, M., Gelfand, D.H., Sninsk, J.J., and White T.J. (1990) PCR Protocols Academic press Inc. NY.
14. Holmes, D.S., and Quigley, M. (1981) Anal. Biochem. 114, 193-201.
15. Maxam, A.M., and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
16. Kolluri, R., and Kinniburgh, A.J. (1991) Nucl. Acids Res. 19, 4771.
17. Kolluri, R., Torrey, T.A., and Kinniburgh, A.J. (1992) Nucl. Acids Res. 20, 111-116.