A conserved heptamer upstream of the IgH promoter region octamer can be the site of ^a coordinate protein-DNA interaction

Nicholas F.Landolfi¹, Xiao-Ming Yin, J.Donald Capra and Philip W.Tucker*

The Department of Microbiology, The University of Texas Southwestern Medical Center, Dallas, TX 75235, USA

Received February 3, 1988; Revised and Accepted May 20, 1988

ABSTRACT

Immunoglobulin genes contain a conserved eight base sequence element 5' to the site of transcription initiation. can serve as a site for the binding of nuclear proteins which are presumably involved in the cell type specific expression of this family of genes. In studying the binding of nuclear proteins to this conserved sequence element, we have detected a protein interaction that involves, in addition to the octamer, nucleotides which are immediately upstream. We have characterized this additional contact as a sequence specific interaction with a heptameric sequence element (CTCATGA) that is conserved among Ig heavy chain promoters. Protein binding to the heptamer is unique in that it is dependent upon the proximity and orientation of, as well as protein interaction with, the conserved octamer.

ç.

INTRODUCTION

Multiple promoter region sequence elements play a role in the B lymphocyte-restricted expression of immunoglobulin (Ig) genes (1-9). It is presumed that these sequences are binding sites for trans-acting factors with transcription-modulating activity. The conserved ATGCAAAT octameric sequence that is found upstream of the TATA box in all Ig heavy (V_H) and light (V_L) chain variable region genes is one such element $(1,10)$. This sequence, which alone can impart lymphoid specificity to a heterologous promoter (11,12), is a binding site for nuclear proteins which are found in all cells (13-18) as well as for a protein which is restricted to lymphoid cells (13,15). Another conserved Ig promoter region element is a heptamer with the consensus CTCATGA (9,19). This sequence is located upstream of the octamer in all Ig heavy chain genes with between 2 and 22 nucleotides separating these sequence elements (9). This heptamer is among sequences required for the full, lymphoidspecific activity of the Ig promoter (8,9).

In previous studies of nuclear proteins that bind specifically to Ig promoter region sequences, we detected four distinct DNA-protein complexes in a gel mobility shift assay using a nuclear extract prepared from the B lymphocyte cell line BCL₁ and a fragment containing the promoter region of the V_H gene expressed by this cell (13). We arbitrarily designated these DNA-protein complexes species 1-4 on the basis of the extent of migration retardation. The formation of all of these species could be inhibited by pre-incubating the extract with a synthetic DNA fragment that contained the octameric sequence ATGCAAAT (13). Thus, each of these complexes represents a sequence specific interaction with the conserved octamer. Analysis of nuclear extracts prepared from cell lines representing distinct lineages revealed that two of the complexes (species ¹ and 2) could be detected with extracts from a variety of cell types, while another complex (species 3) was only detected with lymphoid cell extracts and may represent an interaction involved in the lymphoid cell restricted expression of Ig genes (13).

Further characterization of the sequences involved in the formation of these complexes revealed that detection of the slowest migrating species (species 1) required both the conserved octameric element and sequences immediately upstream (13). The other complexes (species 2-4) required only the eight nucleotides of the octamer. We confirmed the involvement of upstream nucleotides in the formation of species ¹ by DNAse ^I protection analysis. While all four migration retarded species exhibited protection of the nucleotides of the octamer, species ¹ exhibited an additional area of partial protection upstream of the octamer (20). Here we report that the additional protein contact detected in species ¹ is a sequence-specific interaction with a conserved heptameric element that has been demonstrated to be involved in the lymphoid cell-type specific activity of Ig promoters. Protein binding to the heptamer is unique in that it is dependent upon the proximity and orientation of, as well as protein interaction with, the conserved octamer.

MATERIALS AND METHODS

Gel mobility shift assay.

Nuclear extract preparation, fragment labeling, and gel mobility shift analysis were carried out as decribed (13,20). The Ig promoter region fragments analyzed, and their descriptive references, are provided in the legend to Figure 1. Oligonucletides were synthesized on an Applied Biosystems 380-A automated DNA synthesizer.

Methidiumpropyl-EDTA protection analysis.

Methidiumpropyl-EDTA (MPE) protection ("footprinting") analysis was carried out exactly as described (20) except that MPE (23) was employed as the endonucleolytic agent. A detailed comparison of MPE and DNAse protection analysis will appear elsewhere (25).

RESULTS

The additional protein contact detected in species 1 is a sequence specific interaction.

We initially hypothesized that the lack of complete protection from DNAse I digestion in the area immediately upstream of the

Figure 1. *a*) The interaction of five V_H promoter regions with a nuclear extract prepared from BCL₁ cells as assessed by a gel mobility shift assay. Odd lanes received no extract. The positions of the four arbitrarily numbered migration retarded species are indicated on the left. The specific fragments utilized were: BCL_1 , the 108 base $\mathit{Rsa}\;$ I fragment (13); 91A3, the 151 base *Xba I/Nco 1* fragment (21); 186, the ~150 base *Xba I/Nco I* fragment (21); 1.3, the ~200 base Xba *I/Pst I* fragment (21); Vl, the 252 base Mbo II/BamH ^I fragment (22). b) Sequence comparison of the relevant portion of the genes in a. A dash indicates homology with the BCL₁ sequence. The conserved octameric (0) and heptameric (H) elements are shaded. Sequences are from the references in a.

octamer in species 1 indicated that the additional contact was not a sequence specific interaction (e.g., electrostatic interaction with phosphate groups) (20). To test this hypothesis, we prepared ³²P-labeled octamer-containing promoter region fragments from four other V_H genes that differed in sequence upstream of the octamer, and analyzed these in the gel mobility shift assay with the $BCL₁$ nuclear extract. The $BCL₁$ promoter region fragment resulted in the detection of the four previously characterized migration retarded species (Fig. la, lane 2) (13,20). The four other promoter region fragments resulted in DNA protein complexes that migrated equivalently to species 2, 3 and 4 detected with the $BCL₁$ promoter fragment (Fig. 1). Species 1 however was only detected with the BCL₁ and 91A3 V_H promoter regions (Fig. 1a, lanes 2 and 4). The formation of all of the migration retarded species could be inhibited by the addition of a duplex oligonucleotide containing the conserved octameric sequence (data not shown), establishing the sequence specificity of the interactions.

Comparison of the sequences of these fragments (Fig. lb) reveals that the octamer is completely conserved in each of these promoter regions. In addition, the BCL₁ and 91A3 V_H fragments exhibit an additional area of high homology immediately upstream of the octamer that is characteristic of the members of the J558 V_H family (8,19). The 1.3 and 186 V_H promoters, while also members of the J558 family (21), each differ from the area of homology shared between $BCL₁$ and 91A3 by two nucleotides (Fig. 1b). The V1 promoter, a member of the S107 V_H family (22), exhibits no homology in this area. These results implied that the formation of species 1 was a sequence specific interaction, and that the target sequence was within the area of identity between the $BCL₁$ and 91A3 promoters, immediately upstream of the octamer.

High resolution localization of the boundry of the protein contact upstream of the octamer in species 1 reveals the involvement of the conserved heptamer.

Our previous DNAse I protection analysis of species 1 had implicated a maximum of 13 nucleotides upstream of the octamer as having some involvement in the formation of species 1 (20). To delineate more precisely the nucleotides involved in the formation of species 1, we employed protection or "footprinting" analysis using methidiumpropyl-EDTA-Iron(II) (MPE) as the DNA cleaving agent. MPE is a reagent superior to DNAse 1 for protection analysis because this synthetic DNA intercalator, which cleaves DNA in the presence of ferrous ion and oxygen (23), exhibits no sequence specificity and thus allows the assessment

of every nucleotide (25). MPE protection analysis of species $\mathbf{1}$ revealed significant protection of eight bases upstream of the octamer on the coding strand, as well as protection of the eight bases of the octamer and 2-3 bases 3' (Fig. 2b, lane 3). This result allows a more precise assignment of the nucleotides involved in protein interaction in species 1, and localizes the additional protein DNA contact detected in species 1 to the eight nucleotides upstream of the octamer, which corresponds to the area of homology in the 91A3 and BCL₁ genes (Fig. 1b).

The nucleotides implicated in the formation of species 1

Figure 2. MPE protection analysis of species 1. The 165 base Mbo II/Nco *I* fragment from the BCL₁ V_H gene was labeled at the
Nco *I* site, incubated with BCL₁ v_H gene was labeled at the
 Nco *I* site, incubated with BCL₁ nuclear extract, and treated with
10 uM MPE for fiv processed and analyzed as described (20, 25). Lane 1 represents
the A+G specific cleavage pattern of the fragment; lane 2 represents the MPE cleavage pattern of the unbound fragment. Lane 3 represents the pattern of cleavage of the DNA (coding strand) isolated from species 1. The sequence of the fragment, with the octamer (0) and heptamer (H) bracketed, is indicated on the left.

comprise a heptameric sequence recently identified by Siu et al. (19) and Eaton and Calame (9) that is conserved (maximum one base mismatch with consensus) among all V_H promoters. Analogous to the octamer, it is also located in some non-Ig promoters (9). An interesting characteristic of this element is that it is separated from the octamer by a variable distance (2-22 nucleotides) (9). The $BCL₁$ and 91A3 promoter fragments possess the consensus heptamer at an identical position; two nucleotides separate the heptamer and the octamer in these promoters. Two bases also separate these elements in the 1.3 and 186 genes, however these heptamers each deviate from the consensus by two nucleotides. It is important to note that while 1.3 is a pseudogene by a number of criteria, it is not known if the 186 gene is functional (21). The Vl promoter of the S107 family possesses a six of seven match with the consensus heptamer, however this element is separated from the octamer by 14 nucleotides (Fig. lb).

Analysis of the sequence specificity and spacing/orientation requirements of the protein-heptamer interaction.

The ability to detect species 1 with the two V_H promoters that possess heptamers which are a 100% match with the consensus and the inability to detect this complex with fragments that deviate from the heptamer consensus sequence by two nucleotides (1.3 and 186) implied a strict sequence specificity for this interaction. The heptamer of the Vl promoter, however, in addition to possessing a single mismatch with the consensus, is separated by 14, rather than two, nucleotides from the octamer. To determine if the mismatch was responsible for the inability of the heptamer in the Vl gene to act as a site of protein interaction, we synthesized complimentary oligonucleotides which were identical to the BCL₁ promoter sequence except that they contained the transversion present in the heptamer of Vl (third base of the heptamer - Fig. 3e). The oligonucleotides were $32P-$ labeled, annealed and employed in the gel mobility shift assay. Species ¹ was detected when the transversion of Vl was introduced into the heptamer of BCL_1 (Fig. 3a).

The above result, which clearly demonstrated that the heptamer of Vl could serve as a site of protein interaction when separated from the octamer by two nucleotides, suggested that detection of species 1 may be a function of the spacing between these two sequence elements. To address this possibility, complimentary oligonucleotides containing two, four or six bases between the heptamer and the octamer were synthesized, $32P$ labeled, annealed and employed in the gel mobility shift assay. Species ¹ was not detected when the distance between the heptamer and the octamer was four nucleotides or greater, although this

Figure 3. Analysis of sequence requirements for the detection of the heptamer-specific interaction. a) The heptamer of the V1 gene can serve as a site of protein interaction. Complimentary oligonucleotides either identical to the BCL₁ sequence (lanes 1-
2) or containing the transversion present in the V1 heptamer (V1hep, lanes 3-4, see Fig. 3e) were synthesized, end-labeled, annealed and tested in the gel mobility shift assay with BCL₁ nuclear extract. Lanes 1 and 3 received no extract. b) Spacing between the heptamer and octamer is essential for the detection of species 1. Complimentary oligonucleotides with 2 (lanes 1-2), 4 (lanes 3-4), or 6 (lanes 5-6) nucleotides between the heptamer and the octamer (see Fig. 3f for sequences) were synthesized, end-labeled, annealed and used in the gel mobility shift assay with BCL₁ nuclear extract. The odd numbered lanes received no Analysis of mutations in conserved bases of the extract. $c)$ Complimentary oligonucleotides of the BCL₁ sequence heptamer. containing single base changes in the first (Hep $\Delta 1$, lanes $1-2$), fourth (Hep $\Delta 4$, lanes 3-4), sixth (Hep $\Delta 6$, lanes 5-6) and seventh (Hep $\Delta 7$, lanes 7-8) nucleotides of the heptamer (see figure 3e for sequences) were synthesized, labeled, annealed and used in the gel mobility shift assay with BCL_1 nuclear extract. The odd numbered lanes received no extract. d) The relative orientation of the octamer and the heptamer are essential for the detection Complimentary oligonucleotides of the BCL1 of species 1. sequence (lanes 1-2) and the identical sequence with the octamer in reverse orientation (oct-inv, lanes 3-4, see Fig. 3e for sequence) were synthesized, end-labeled, annealed and used in the beginned in column terms and the conservation of the disponential model in the conservation of the con heptamer in V1-hep and the HepA series are underlined.
octamer in V1-hep and the HepA series are underlined. The f Sequences of the oligonucleotides employed in c . The heptamer and octamer are shaded.

protein-DNA complex was clearly detected with a duplex oligonucleotide which contained only two bases between these sequence elements (Fig. 3b). Therefore, the spacing between the heptamer and octamer is essential for the detection of species ¹ in this assay.

Sequence comparison of several heptamers have indicated that the first, fourth, sixth and seventh nucleotides of the consensus CTCATGA sequence are highly conserved (invariant in 19 cases) (9). We examined the importance of each of these residues in the ability to detect species ¹ by synthesizing complementary oligonucleotides which systematicaly change each of these nucleotides. Of the four single base changes introduced, only the $A\rightarrow C$ transversion in the fourth base of the heptamer completely abolished the ability to detect species 1 (Fig. 3c, lane 4). Changing the sixth base $(G \rightarrow T)$, however, consistently decreased the amount of species 1 detected (Fig. 3c, lane 6). Thus, although we have not examined all possible permutations, our data indicate that this interaction can tolerate single base deviations from the consensus, as long as the spacing between the heptamer and the octamer is two nucleotides. We have, however, detected instances where a single mismatch either diminishes or abrogates the interaction.

We have also examined the effect of altering the AT dinucleotide which separates the heptamer and the octamer in members of the J558 family. Changing the adenine to a cytosine slightly decreases the amount of species ¹ detected, while changing the thymidine to a guanosine abolishes the ability to detect this species, however the amount of species 2, 3, and ⁴ is also slightly decreased (data not shown). An examination of V_H promoter region sequences reveals that a guanosine immediately preceeding the octamer has not been observed (1,8,9,10,19) implying a selection against this nucleotide at this position. The significance of this is unclear, however these data suggest that the nature of the dinucleotide spacer between the heptamer and the octamer also contributes to the interaction between nuclear proteins and these two sequence elements.

The octamer exists in a different orientation relative to the TATA box in different genes (1,10). To determine if the orientation of the octamer relative to the heptamer was important in the formation of species 1, we constructed a fragment identical to the $BCL₁$ promoter except that orientation of the octamer was reversed. Figure 3c establishes that inverting the octamer in relation to the heptamer abolishes the ability to detect species 1.

DISCUSSION

The experiments reported here indicate that the conserved heptameric sequence CTCATGA that is located a variable distance upstream of the ATGCAAAT octamer in IgH promoter regions is a site of protein interaction. The detection of this interaction, which manifests itself as species 1 in the gel mobility shift assay, has stringent requirements. No migration retarded species (protein-DNA complexes) are detected in the gel mobility shift assay using a fragment which contains only the heptamer sequence, and unlabeled fragments which contain the heptamer alone do not inhibit the formation of species 1 (data not shown). These data indicate that the affinity of the heptamer-specific protein(s) for this sequence alone is insufficient to maintain the integrity of the interaction upon analysis by gel electrophoresis. Species 1 is only detected when: 1) both the heptamer and the octamer are on the same fragment, 2) these elements are separated by two nucleotides, 3) they are in the same orientation, and 4) both sites are contacted by a protein(s).

These data lend themselves to a number of interpretations of the nature of species 1. It is possible that a single protein, with sequence specificity for both the octamer and the heptamer is responsible for species 1. Insufficient affinity for either of the sites alone can be invoked to explain the lack of a complex which comigrates with species 1 when only one of these elements is present on the indicator fragment. Alternatively, species 1 may result from the interaction of a protein (that has sequence specificity for the heptamer) with a DNA fragment that already has a protein bound to the octamer (i.e., species 2). The inability to detect the interaction of this protein alone with the isolated heptameric sequence could indicate a protein-protein interaction between the octamer- and heptamer-specific proteins that facilitates binding of the putative protein to the heptamer. This protein-protein interaction could stabilize the complex and thus allow the detection of species 1 in a gel mobility shift assay. Species 1 could also represent the binding of two copies of an octamer-specific protein (i.e., that responsible for species 2) to the fragment (the last four nucleotides of the heptamer along with the first two bases of the octamer and the two spacing bases comprise a six of eight match with the consensus octamer). The latter alternative is unlikely because high concentrations of the purified protein responsible for species 2 does not result in the formation of species 1.

We favor the existence of two distinct proteins for several reasons. It is easier to reconcile the differential spacing

Nucleic Acids Research

between the heptamer and octamer in various V_H promoters; it is unlikely that a single protein could possess two distinct sequence specific binding sites with sufficient flexibility to accomodate a spacing of ² to 22 nucleotides between the recognition sequences. In addition, while the proteins responsible for species 2, ³ and 4 can be biochemically separated from one another (24), species 1-forming activity is only detected in samples that also contain the protein which results in species ² (J. Hanke, personal communication). Thus, while the existence of a single protein with specificity for both sequences is a formal possiblity, we consider it unlikely. If species 1 was composed of two distinct proteins, this interaction would be analogous to the coordinate binding observed between SLl and UBF1 to the human rRNA promoter (26), and similar to the cooperativity exhibited by various procaryotic (27) and eucaryotic (28) transcription-regulating proteins.

It is interesting that, while the number of nucleotides that separates the heptamer and the octamer in the nineteen V_H promoters examined is a minimum of two and a maximum of twentytwo nucleotides, only two other spacer lengths (namely ten or fourteen nucleotides) have been observed (9). This implies that there is a constraint upon the distance between the octamer and the heptamer that is satisfied by a spacing of only two, ten, fourteen or twenty-two nucleotides. Possibly separate octamerand heptamer-specific proteins can interact successfully only when their respective binding sites are separated by one of the above distances.

It should be emphasized that we do not believe that the inability to detect species 1 with a particular V_H fragment is an indication that protein-heptamer interactions have no *in vivo* relevance to the expression of that particular gene. In fact, the striking conservation of this element, regardless of distance from the octamer, argues to the contrary. Rather, we feel that the fortuitous spacing of these two elements in the members of the J558 family has merely facilitated the detection of a heptamer-specific interaction by the gel mobility shift assay. Our data also offer an explanation for the descrepancy between our results (13) and those of others (14-18) who employed Ig genes other than those of the J558 family in their analysis, and thus have detected fewer complexes.

The role of a heptamer-specific protein is at present purely speculative. The clear contribution of the heptamer to the lymphoid-specific activity of Ig promoters, as assessed by transient transfection assays employing deletion mutants (8,9), would suggest a factor that is both lymphoid-specific and of a transcriptional-enhancing nature. However, species 1, of which the heptamer-specific interaction we detect is an integral part, is exhibited by both lymphoid and non-lymphoid extracts (13,20). Furthermore, nuclear extracts of mitogen activated B lymphocytes, which exhibit extremely high levels of Ig transcription, result in little or no species 1 (13). Similarly, mitogenic stimulation of BCL₁, which results in elevated levels of the lymphoidspecific species 3, greatly diminishes the level of species 1 (20). Therefore, it is unlikely that the heptamer-specific protein we detect is involved, at least in the positive sense, in Ig transcription. The above characteristics are more consistent with a negative regulatory function for the protein we detect, although assignment of actual functional activity will have to await direct analysis in an in vitro transcription assay. It is possible that, as we have suggested for octamer-specific proteins (20), a family of heptamer-specific proteins exists, with each possessing a distinct effector activity. Control of Ig gene expression would then be based on the cumulative activity of the proteins that occupy the sequence-specific promoter region binding sites at a given point in the differentiative pathway of the cell.

ACKNOWLEDGEMENTS

We thank Maureen Willard for technical assistance, Chhaya Das and Maya Sarkar for preparation of plasmid DNA and Helen Aronovich for oligonucleotide synthesis. We are also grateful to Dr. P. Dervan for the MPE, to Dr. G. Rathbun for providing the sequence of 1.3 and 186, to Drs. S. Eaton and K. Calame for providing a copy of their manuscript pre-publication, and to Dr. J. Hanke for comments on the manuscript. This work was supported by National Institutes of Health Research Grants AI-18016, GM-1751, and GM-31689. N.F.L. is a fellow of the Leukemia Society of America.

To whom correspondence should be addressed.

Present address: Protein Design Labs, Inc., 3181 Porter Drive, Palo Alto, CA 94304.

REFERENCES

- 1. Falkner, F.G. & Zachau, H.G. (1984) Nature (London) 310, 71-74.
- 2. Bergman, Y., Rice, D., Grosscheld, R. & Baltimore, D. (1984) Proc. Natl. Acad. Sci. USA 81, 7041-7045.
- 3. Foster, J., Stafford, J. & Queen, C. (1985) Nature (London) 315, 423-425.
- 4. Mason, J.O., Williams, G.T. & Neuberger, M.S. (1985) Cell

- 41, 479-487. 5. Grosscheld, R. & Baltimore, D. (1985) Cell 41, 885-897.
- 6. Gopal, T.V., Shimada T., Baur, A.W. Niehaus, A.W. (1985) Science 229, 1102-1104.
- 7. Picard, D. & Schaffner, W. (1985) EMBO J. 4, 2831-2838.
- 8. Ballard, D.W. & Bothwell, A. (1986) Proc. Natl. Acad. Sci. USA 83, 9626-9630.
- 9. Eaton, S. & Calame, K. (1987) Proc. Natl. Acad. Sci. USA 84, 7634-7638.
- 10. Parslow, T., Blair, D.L., Murphy, W.J. & Granner, D.K. (1984) Proc. Natl. Acad. Sci. USA 81, 2650-2654.
- 11. Dreyfus, M., Doyen, N. & Rougeon, F. (1987) EMBO J. 6, 1685-1690.
- 12. Wirth, T., Staudt, L. & Baltimore, D. (1987) Nature 329,
- 174-178.
Landolfi, 13. Landolfi, N.F., Capra, J.D. & Tucker, P.W. (1986) Nature (London) 323, 548-551.
- 14. Singh, H., Sen, R., Baltimore, D. & Sharp, P.A. (1986) Nature (London) 319, 154-158.
- 15. Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A. & Baltimore, D. (1986) Nature (London) 323, 640-643.
- 16. Hromas, R. & Van Ness, B. (1986) Nucleic Acid Res. 14, 4837- 4848.
- 17. Sive, H.L. & Roeder, R.G. (1986) Proc. Natl. Acad. Sci. USA 83, 6382-6386.
- 18. Mocikat, R. Falkner, F.G., Mertz, R. & Zachau, H.G. (1986) Nucleic Acid Res. 14, 8829-8843.
- 19. Siu, G., Springer, E.A., Huang, H.V., Hood, L. & Crews, S.T. (1987) J. Immunol. 138, 4466-4471.
- 20. Landolfi, N.F., Capra, J.D. & Tucker, P.W. (1987) Proc. Natl. Acad. Sci. USA 84, 3851-3855.
- 21. Rathbun, G.R., Otani, F., Milner, E.C.B., Capra, J.D. & Tucker, P.W. (1988) J. Mol. Biol. In press.
- 22. Clarke, C., Berenson, J., Goverman, J., Boyer, P.B., Crews, S., Sui, G. & Calame, K. (1982) Nucleic Acid Res. 10, 7731-7749.
- 23. Hertzberg, R.P. & Dervan, P.B. (1984) Biochemistry 23, 3934- 3945.
- 24. Hanke, J., Landolfi, N.F., Tucker, P.W. & Capra, J.D. (1988) Proc. Natl. Acad. Sci. USA, In press.
- 25. Landolfi, N.F., Yin, X.M., Capra, J.D., and Tucker, P.W. (1988) Analyt. Biochem., submitted.
- 26. Learned, R.M., Learned, T.K., Haltiner, M.M. & Tjian, R.T. (1986) Cell 45, 847-857.
- 27. Ptashne, M. (1986) A Genetic Switch, Blackwell Scientific Publications.
- 28. Topol, J., Ruden, D.M. & Parker, C.S. (1985) Cell 42, 527-537