A functional component of the sea urchin H2A gene modulator contains an extended sequence homology to a viral enhancer

Rudolf Grosschedl*, Marco Machler, Urs Rohrer and Max L.Birnstiel

Institut fur Molekularbiologie II der Universitat Zurich, HOnggerberg, CH-8093 Zurich, Switzerland, and *Massachusetts Institute of Technology, Center for Cancer Research, Cambridge, MA 02139, USA

Received 19 October 1983; Accepted 15 November 1983

ABSTRACT

The DNA sequences imparting a maximal rate of sea urchin H2A gene transcription in the frog oocyte nucleus were narrowed down by deletion mapping to a DNA segment -165 to -111, farupstream of the H2A mRNA cap site. C to T base changes in this area create strong down mutations, hence the primary structure of this DNA sequence is of paramount importance to the H2A gene expression. Sequence comparisons suggest that the -165 to -111 region may contain two essential sequence blocks. Most strikingly, the -135 area contains a 14 out of 17 basepair homology to the Moloney murine sarcoma virus enhancer and to topologically related ⁵' LTR-sequences of the simian sarcoma virus and the murine Friend spleen focus forming virus.

INTRODUCTION

The highly repeated cluster of histone genes of the sea urchin Psammechinus miliaris expressed during early embryogenesis has been isolated as h22 DNA and thoroughly studied in this laboratory. The nucleotide sequence of this DNA (1, and unpublished data) reveals conserved sequence motifs in the 5' flanking regions of this group of histone genes $(2,3)$. Because of their evolutionary conservation amongst histone and many other protein coding genes it has been suggested that these sequence elements constitute promoter elements for RNA polymerase II (4-6). To test the functional relevance of these sequences, the phenotypic effects of various deletions in the 5' flanking region of the H2A gene were studied by injection of the mutated DNAs into the nuclei of Xenopus laevis oocytes (7,8). These studies revealed that at least two promoter elements are required for efficient and accurate transcription of the H2A histone gene. The TATA box was shown to determine

primarily the specificity of transcription initiation (7,9); in addition, this nearly universal promoter element has been found to be important for selecting the RNA cap site and for an optimal rate of transcription for several other eukaryotic genes $(10-15)$.

A second promoter element, the modulator of the H2A gene, located at a considerable distance upstream of the TATA box dramatically influences the efficiency of transcription initiation. This control element can act in both orientations. Deletion mapping at low resolution located this regulatory sequence between nucleotides -446 and -111 upstream of the putative initiation site (8). Similar "upstream sequences" have since been identified near other genes, although the location and relative functional importance of such sequence elements seem to vary from gene to gene (see Discussion).

In this paper we refine our analysis of the distal control element of the H2A histone gene by studying the phenotypic effects of deletions and multiple point mutations in the far upstream sequences. We demonstrate that the region between nucleotides -165 and -111 contains important control elements required for maximal transcription of the H2A gene injected into frog oocyte nuclei. Contained in this DNA segment is a sequence which can also be found in a similar position near the H2A gene of the related sea urchin Paracentrotus lividus (M. Ciaccio, unpublished results). A most striking feature of the conserved H2A upstream sequence is its high degree of sequence homology to the promoter proximal part of the Moloney murine sarcoma virus (Mo-MuSV) enhancer (14 out of 17 bases) and to similar upstream LTR-sequences of the Simian sarcoma virus (SSV) and the murine Friend spleen focus forming virus (SFFV).

MATERIALS AND METHODS

DNA constructions: Restriction enzymes, T4 polynucleotide kinase, T4 DNA ligase and DNA polymerase ^I were purchased from New England Biolabs and Boehringer Mannheim and were used according to the prescribed reaction conditions. Synthetic linkers were obtained from Collaborative Research and Bal 31 nuclease from Bethesda Research Laboratories.

Construction of H2A del (-165/-111) and del (-140/-111): The \sim 1 kb Bst EII/Xho I fragment was isolated from mutant del (-164/-111) RI and (-139/-111) RI recombinants (16). After cleavage with Hpa II, the Bst EII/Hpa II DNA fragments containing the E-segment (8) were ligated to the 375 bp Taq I/Xho ^I fragment of wild-type DNA and cloned into a Bst EII/Xho ^I cleaved pBR h22 vector.

Na-bisulfite mutagenesis: The Bst EII/Xho ^I fragments of h22 wild-type (Fig.2b) and mutant del (-164/-111) RI DNA were blunt-ended by filling the staggered DNA ends with Klenov DNA polymerase and dNTPs and inserted into the Sma ^I site of M13-mp8. lug of single-stranded recombinant wild-type phage DNA was hybridized to 10ug EcoRI cut double-stranded mutant phage DNA (-164/-111) RI in the presence of $10\mu q$ E.coli RNA. The bisulfite treatment was performed at 37°C for 1.5 hr as described (17) and the reaction was terminated by passing the reaction mixture over a P10 (Bio Rad) column in O.O1M Tris/HCl pH 7.5, 0.1M NaCl and 0.OO1M EDTA. Tris-base was added to the DNA fraction and after incubation at 37° C for 8 hr, the reaction mixture was neutralized and ethanol precipitated. The filling-in of the gapped, mutated heteroduplexes was carried out as previously described (18).

Screening of the mutated recombinant phages: DNA sequence analysis of the mutated region and flanking sequences in the recombinant phages was performed by the "dideoxy-method" (19). To enable us to use the 12 nucleotide universal primer (PL-Biochemicals) near the boundary of the M13 insert, we deleted the DNA sequences between the Hinc II site at position -1 of the H2A gene and the Hinc II site in the M13 polylinker. By this manipulation the region of interest residing between nucleotides -165 and -111 of the H2A gene as well as other 5' flanking sequences were moved 276 nucleotides closer to the vector boundary and the universal primer could conveniently be used for sequencing. The sequencing reactions were performed as described (19).

DNA injection and RNA analysis: 4ng of circularized histone insert were injected into oocyte nuclei together with 0.2pCi of α ⁽³²P)GTP as described (20). After 20 hrs of incubation total

RNA was extracted and analysed by gel electrophoresis under partially denaturing conditions (21).

Nuclease Sl mapping of H2A mRNA: For the Sl mapping of H2A transcripts obtained from the point-mutated templates, unlabelled RNA was extracted from injected oocytes. lOµg of total RNA were hybridized to the BstEII/XhoI DNA fragment, ⁵' end-labelled at the Xho ^I site, and treated with S1 nuclease as described (3,16).

RESULTS

The 5' flanking region of the H2A histone gene of Psammechinus miliaris, delimited by two Taq ^I restriction sites at nucleotides -446 and -111 has been previously demonstrated to modulate the rate of H2A transcription in frog oocyte injection experiments. Deletion of this DNA sequence, termed E-segment or H2A modulator, reduces the level of H2A mRNA synthesis by a factor of 15 to 20 (8), while inversion of the DNA segment actually stimulates H2A gene transcription. The regulatory DNA signals residing in the E-segment could not be reliably identified from sequence homologies with other histone genes, due to the large size of the E-fragment. Therefore we decided to first delimit the operative regulatory DNA sequences by generating a series of deletions spanning the far upstream region.

Localization of the distal promoter element by deletion mapping

A series of 5' deletion mutants was obtained by Bal3l exonuclease digestion of the E-segment from the ClaI site (-415) near its 5' end. EcoRI DNA linkers were ligated to the truncated DNA fragments and the DNA was recloned. A similar series of 3' deletions were generated by removing the sequences from the gene proximal 3' end of the E-segment starting at an EcoRI site, which had been previously inserted between nucleotide positions -112 and -111 (16). Again, EcoRI DNA linkers were added to the 3' shortened DNA fragments and the DNA was recloned (16). Appropriate recombinants from the 5' and 3' resection series were combined with one another. In this way, a series of overlapping, internal deletions in the modulator segment was obtained. Once injected into the frog oocyte nucleus, it became

immediately clear that it was primarily the constructions lacking gene proximal sequences of the E-segment which were strong down mutations. We also found evidence that the newly inserted EcoRI DNA linkers, which replaced the deleted DNA sequences, were probably not neutral in our transcriptional analyses (R. Grosschedl, unpublished results). We therefore decided to re-investigate the gene proximal region by functional tests of deletion mutants lacking the EcoRI DNA linker sequences.

We were fortunate that in the ³' to 5' resection series of the E-segment addition of the EcoRI DNA linker GGAATTCC to nucleotides -164 and -139, respectively, created a new Hpa II site in both instances. Consequently, Hpa II restriction of the appropriate subclones of the resection series removed all but one nucleotide of the EcoRI linker, creating at the same time a ⁵' sequence overhang which could be conveniently ligated to the Taq ^I site at position -111 of wild-type DNA. In this way "clean" deletions (-165/-111) and (-140/-111) could be made (Fig.lb). The mutated DNA sequences were reintroduced into the h22 DNA repeat, where they replaced the wild-type sequences. The now mutated h22 gene repeat unit was reclaimed from the recombinant DNA, circularized in vitro, purified on ethidium bromide density gradients and injected into frog oocyte nuclei. The wild-type H2B gene residing in the same DNA repeat unit served as an internal control.

The h22 deletion mutation del(-165/-111) decreases the level of transcription of the H2A gene by a factor of ⁸ to 10 (Fig.la) that is, almost as much as the ΔE (-446/-111) deletion mutation (8). When mutation del(-140/-111) was tested by oocyte injection experiments, a reduction of H2A transcription by a factor of about ⁴ was revealed. We conclude from these results that essential upstream elements of the H2A gene reside in the 53 gene proximal nucleotides of the E-segment with both the DNA sequences -165 to -140 and -140 to -111 apparently contributing to the transcriptional activation of the H2A gene.

In the above mutants, both sequences as well as the general topology of the H2A 5' flanking sequences were changed. To reduce the perturbation analysis to a single parameter we

Nucleic Acids Research

Figure 1: Identification of functionally important H2A modulator sequences by deletion mapping.

a) Gel electrophoretic analysis of 32 P-labelled oocyte transcripts of wild-type and deletion mutants of the H2A sea urchin histone gene (for explanation see text). The autoradiograms were traced in a densitometer to determine the reduction of H2A gene transcription in relation to the internal H2B control.

b) Maps of h22 DNA and deletion mutants used to delimit the functionally important modulator sequences. For construction of the mutants see text.

attempted to identify the nucleotides essential to H2A gene transcription by introducing point mutations in the -165 to -111 region using the bisulfite treatment of single-stranded DNA (18,22) and by measuring the phenotypic effects of these sequence alterations in oocyte injection experiments. Construction and expression of multiple point mutants

To direct cytosine deamination by bisulfite treatment to the region -164 to -111 we constructed gapped DNA molecules according to Everett and Chambon (18) and Folk and Hofstetter (22). For this we cloned the Bst EII/XhoI fragment of the h22 histone DNA containing all 5' flanking sequences as well as half of the coding portion of the H2A gene into the Sma ^I site of the single-stranded vector M13-mp8 (Fig.2b). An equivalent

construction was made using the Bst EII/XhoI fragment of the deletion mutant del(-164/-111) RI where the deleted sequences were replaced by an EcoRI linker (16). The double-stranded, replicative form of the mutant phage carrying the deletion was cut with EcoRI within the DNA linker and hybridized in 10-fold molar excess to the single-stranded DNA of the wild-type recombinant phage. Gapped, circular DNA molecules were treated with sodium bisulfite (18,22). After transfection and cloning, the mutated recombinant phage were screened by dideoxysequencing of the DNAs (see Materials and Methods). None of the mutated DNAs showed sequence changes from nucleotide positions -1 to -111, but base changes in the AT-rich region upstream of position -164 were frequent. These mutants were eliminated from the present studies.

Sequence analysis of the novel recombinants revealed that our collection consisted entirely of mutants with multiple C to T changes and that, all mutants taken together, almost every C residue in the chosen DNA segment had been mutated (Fig.2c). The effects of point mutations falling between -164 and -111 on H2A transcription initiation were studied by injection of doublestranded recombinant phage DNA into Xenopus laevis oocyte nuclei. Since only the ⁵' half of the H2A gene was present in these recombinant DNAs, no H2A mRNA transcripts of defined size were expected to arise. Instead, faithful initiation of the H2A gene transcription had to be assayed by nuclease Sl mapping using a DNA probe 5' labelled at the Xho ^I site (Fig.2b). To quantify the efficiency of transcription, a H2A AC deletion mutant DNA (7) was placed in M13. The double-stranded recombinant was co-injected with the upstream H2A point mutants and served as an internal control. In the AC deletion mutant the site of transcription initiation is transposed 60 nucleotides downstream of the normal cap site and thus a shorter RNA AC transcript becomes detectable when transcripts of co-injected AC and point mutant M13 recombinants are Sl mapped simultaneously.

The S1 analysis of the transcripts obtained from the various mutants reveals that multiple point mutations in the -164 to -111 region of the H2A gene can have drastic effects on

Figure 2: Point mutagenesis of the -164 to -111 upstream sequences of the H2A gene. a) Gel analysis of transcripts. Mutant ΔC (7) and individual point mutants were co-injected into frog oocyte nuclei and their transcripts determined by S1 mapping using a Bst EII/XhoI labelled DNA probe. H2A* transcripts arise from erroneous initiation near position -70 upstream of the genuine cap site of the H2A mRNA. b) Details of the H2A gene map showing the positions of restriction sites used for cloning, S1 mapping and sequencing strategies.

c) Upstream sequences of the Psammechinus and the Paracentrotus H2A histone genes together with the C to T base changes created by deamination of gapped DNA molecules (see text). Consensus sequences are underlined. Numbers to the right indicate the factor by which H2A transcript accumulation is reduced. The values are averages of two measurements. The numbering of the sequences refers to the Psammechinus gene.

transcription (Fig.2a). For instance, mutant 27 is a strong down-mutation by a factor of 12. Other mutants are either neutral or support H2A gene transcription at a reduced rate. The results clearly demonstrate the importance of the primary structure of the far upstream sequences for the wild-type expression of the H2A gene and exclude the possibility that the down mutations obtained by deletion mapping are simply the result of a disturbance of the topology of the ⁵' flanking sequences. However, because in each mutant there are multiple base changes, the decrease in the transcriptional efficiency of the H2A gene cannot be correlated in a conclusive way with individual bases of the upstream area. Nevertheless, the strong down mutant 27 contains several base changes which in other mutants have no or few deleterious effects. As a first approximation, therefore, these changes might be considered to be more or less neutral. Mutant 27 contains uniquely two additional base changes at -140 and -139 and one at -126. Clearly, these three residues provide a focal point for a more thorough investigation by site directed mutagenesis. In all slots of this set of experiments a low level of H2A* transcript can be detected which initiates erroneously near sequence position -70. This RNA is not seen in wild-type h22 injection experiments. It is not clear why it is generated by those novel constructions.

When the sequences in DNA segment -164 to -111 are compared to topologically related areas of other genes, the nucleotide arrangement near -150 and near -133 is found again in closely similar form in the upstream sequences of the H2A gene of the related sea urchin Paracentrotus lividus (Fig.2b and 3). The more proximal sequence block near position -135 is also present, although in altered form, near position -160 of the H3 gene of the sea urchin Lytechinus pictus (41; Fig.3). However, the most

Figure 3: Compilation of sequence homologies in far upstream regions of histone genes and viral LTRs. Numbers refer to the nucleotide position (counting from the mRNA cap site) of the left-most nucleotide on each line. SSV, simian sarcoma virus; SFFV, murine Friend spleen focus forming virus; Mo-MuSV, Moloney murine sarcoma virus. In addition to the cases listed above we have found that the Rous sarcome virus LTR (42) contains at position -96 a ⁹ out of 10 bp homology to the H2A modulator. The RSV sequence aGGAAGqcAACAGAcGGN, CATG appears more closely related to the H3 -171 sequence $tGGAAGcgAACAGAtGGN₄CATG$.

striking sequence homologies that we have found to date are with the enhancer sequences in the LTR of the Moloney murine sarcoma virus (23), where in a stretch of 17 nucleotides 14 bases coincide exactly. Closely similar sequences are encountered again in the upstream regions of two other animal viruses, the SSV (24) and the SFFV (25; Fig.3). Homologies between the H3 gene of Lytechinus pictus and the LTR of Rous sarcome virus (RSV) are noted in the legend of Fig.3.

DISCUSSION

In this paper we further characterize one functional component of the H2A gene modulator (8) of the sea urchin Psammechinus miliaris. Deletion of the sequences between -165 and -111 reduces the level of H2A transcription to about 10-12% of wild-type level, while a small deletion from -140 to -111 decreases transcription to about 25% of the wild-type control (Fig.l). Strong down mutations are also obtained by C to T base changes in the -164 to -111 DNA segment (Fig.2). These point mutants confirm the results obtained with the deletion mutants and show that the primary structure of the upstream sequences is important for maximal expression of the H2A gene in the frog oocyte nucleus.

The partial reduction of transcription seen with the smaller deletion mutant del(-140/-lll)could be most easily explained if there were two important sequence elements in the upstream region of the H2A gene, the smaller deletion removing one, the larger deletion removing both of them at once. Alternatively, it might be argued that the smaller deletion removes only a section of a coherent larger regulatory sequence straddling nucleotide -140, thus eliciting only a partial reduction of transcription. A comparison with the upstream sequences of the early embryonic H2A gene of the related sea urchin Paracentrotus lividus (M. Ciaccio, unpublished results) reveals two conserved sequence blocks near position -145 and position -128 (Fig.2c), that is in a position comparable to the distal and proximal part of the large deletion in the Psammechinus H2A gene. They occur at a distance of about 20 nucleotides from one another (or two turns of helical DNA). Thus it seems possible that two regulatory sequences are present in the -165 to -111 area. It is noteworthy that even the large deletion does not quite reach the low level of transcription seen in the AE deletion mutant (8). Quite possibly, additional important sequences are located further upstream in keeping with the preliminary observation that a deletion of the sequences -421 to -328 reduces H2A gene transcription in the oocyte to about half of wild-type level (R. Grosschedl, unpublished observation).

We observe some, possibly important, differences between the in vivo oocyte injection test and the in vitro transcription analysis. We have previously reported that in the Manley-Sharp extract the sequences from -139 to -111 in combination with the TATA box, suffice for maximal H2A transcription (16). Our present oocyte injection experiments clearly show that sequences between -165 to -140 are required for a maximal rate of H2A gene transcription in addition to the upstream sequences defined previously (16).

Extensive sequence homologies between the H2A modulator and the LTRs of the Mo-MuSV, the SSV and the SFFV have been noted (Fig.3). The 17 bp homologous segment maps to the gene proximal section of the 72/73 bp reduplicated viral enhancer element of

the Mo-MuSV, which can replace the 72 bp SV40 enhancer functionally to activate the early SV40 gene (40). Hence, in both the H2A and the Mo-MuSV genetic unit the conserved sequence is associated with a functionally important upstream region.

How universal is the sequence arrangement found in the H2A gene control region? There are now many examples of sequences at a considerable distance from the site of transcription initiation which enhance transcription of the gene in vivo. A number of such control elements are more tightly linked to the TATA boxes and -50 control elements (for literature citation see ref.26) such as the H2A modulator (8) or the "distal" elements of the h22 H2B sea urchin gene (J.Mous, unpublished results), of the TK gene (27), of the rabbit β -globin gene (13-15) and the yeast his3 gene (28). Other sequences are further removed from the transcription initiation site. For instance, in the yeast, a control sequence at a distance from the ³' end of the H2B histone gene (but 5' to the co-regulated H2A gene) modulates transcription of both H2A and H2B genes during the cell cycle (29). A sequence near -275 imparts heme-dependent regulation of the CYCI gene (30). Functionally important, far upstream sequences of a Drosophila gene have been identified by the analysis of naturally occurring mutants (31). The MMTV (32,33), the lysozyme (34) and the ovalbumin gene (35) contain hormon receptor binding sites which enhance gene function. An interesting example of a tissue-specific enhancer is the geneinternal control element of the Ig gene (36,37). Finally there is the important class of viral enhancers, which act at a great distance, in an orientation independent manner, from both the ⁵' and the 3' end of the gene (38). Enhancers also augment transcription of most, but not all, genes linked to them in SV40 recombinant molecules (38,39).

A property that the H2A upstream sequences (8) share with viral enhancers (38), the hormone receptor binding sequences (32) and the yeast histone gene regulatory element (29) is that they can act in both orientations. The topology of the yeast histone gene control element suggests that it can exert its influence when present both near the 5' or the 3' end of the gene. Unlike viral enhancers, the hormone acceptor site (32) and the TK -100 control element (27) become rapidly inoperative when the distance to the TATA box is increased by insertion of DNA between these two control elements.

Thus it has become clear that distal control elements differ in their topologies and, as a rule, in their sequence. Whether, nevertheless, these control elements act in a similar way to enhance transcription, but differ in the effector molecules with which they interact, is not known at the moment, since the molecular mechanisms of transcriptional enhancement have not yet been resolved in any of these cases. Enhancers may possibly be complex structures controlling different aspects of gene regulation. The observed strong homologies found between the sea urchin modulator and some viral enhancer sequences now call for direct biochemical experiments to establish whether or not these sequences signal similar genetic functions and to determine what these might be.

ACKNOWLEDGEMENTS

We would like to thank Dr. D. Schümperli for suggesting revisions and Dr. M. Chipchase for critical reading of the manuscript. We thank S. Oberholzer and F. Ochsenbein for having prepared this manuscript and V. Dostal for plasmid preparations. This work was supported by the Kanton of Zürich and the Swiss National Research Council, grant No.3.743.80.

REFERENCES

- 1. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655-671
- 2. Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980) Nucl.Acids Res. 8, 957-977
- 3. Hentschel, C., Irminger, J.C., Bucher, P. and Birnstiel, M.L. (1980) Nature 285, 147-151
- 4. Hentschel, C. and Birnstiel, M.L. (1981) Cell 25, 301-313
5. Breathnach, R. and Chambon, P. (1981) Ann.Rev.Biochem. 50
- 5. Breathnach, R. and Chambon, P. (1981) Ann.Rev.Biochem. 50, ³ 49-383
- 6. Shenk, T. (1981) Curr.Top.Microbiol.Immunol. 93, 25-46
- 7. Grosschedl, R. and Birnstiel, M.L. (1980) Proc.Natl.Acad.Sci. USA 77, 1432-1436
- 8. Grosschedl, R. and Birnstiel, M.L. (1980) Proc.Natl.Acad.Sci. USA, 7102-7106
- 9. Grosschedl, R., Wasylyk, B., Chambon, P. and Birnstiel, M.L. (1981) Nature 294, 178-180
- 10. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310
- 11. Fromm, M. and Berg, P. (1982) J.Mol.Appl.Genet. 1, 457-481

