Poly(pyrimidine)•poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH

Jeremy S. Lee, Mary L. Woodsworth, Laura J. P. Latimer and A. Richard Morgan

Department of Biochemistry, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 0W0, and 1Department of Biochemistry, University of Alberta, Edmonton, Alberta, T6G 2H7, Canada

Received 19 April 1984; Revised and Accepted 20 July 1984

ABSTRACT

Poly(pyrimidine)•poly(purine) tracts have been discovered in the 5'-flanking regions of many eucaryotic genes. They may be involved in the regulation of expression since they can be mapped to the nuclease-sensitive sites of active chromatin. We have found that poly(pyrimidine)•poly(purine) DNAs which contain 5-methylcytosine (e.g. poly[d(TmC)]•poly[d(GA)]) will form a triplex at a pH below 8. In contrast, the unmethylated analogue, poly[d(TC)]•poly[d(GA)] only forms a triplex at pHs below 6. Synthetic DNAs containing repeating trinucleotides and poly[d(UmC)]•poly[d(GA)] behave in a similar manner. Thus the stability of a triplex can be controlled by methylation of cytosine. This suggests a model for the regulation of expression based upon specific triplex formation on the 5'-side of eucaryotic genes.

INTRODUCTION

At least two sequences have been identified which appear to be important for eucaryotic gene expression. These putative promoters include an ATA box and a CAT box which are located -30 and -80 residues respectively, on the 5'-side of the capping site (1). However, other regulatory regions are almost certainly involved since nuclease sensitive sites of active chromatin can be mapped to the 5'-flanking sequence of expressing genes (2,3,4). This suggests that the regulatory/promoter DNA has an altered secondary structure which confers nuclease sensitivity. Recently several Sl-nuclease hypersensitive sites (SHS) from the human α-globin, the chicken β-globin, the Drosophila heat-shock genes, and the sea urchin histone genes have been sequenced (5-8). They occur in regions of poly(pyrimidine)•poly(purine) DNA sequences up to 90 base-pairs in length. Other pyrimidine-purine tracts are found frequently in many eucaryotes (9).
Previously we showed that poly(pyrimidine)•poly(purine)
synthetic DNAs spontaneously dismutate at pH 5 to give triplexes
containing C.G.C base-triads (10).

\[2 \text{poly[d(TC)].poly[d(GA)]} \stackrel{\text{PH 5}}{\rightarrow} \]
\[\text{poly[d(TC)].poly[d(GA)].poly[d(c'T)]} + \text{poly[d(GA)]}\]

Although hysterisis (i.e. the presence of a metastable state—
see figure 4) is well documented in nucleic acid systems, a
physiological role seemed rather unlikely because of the low pH
required for formation (11). Here, we demonstrate that
poly(pyrimidine)•poly(purine) DNAs containing 5-methylcytosine
\(\text{m}^5\text{C}\) will form a triplex spontaneously at neutral pH. Although
most \(\text{m}^5\text{C}\) bases occur at CpG sequences, CpC and TpC dimers are
also methylated (12,13). Thus, it is possible that the
expression of eucaryotic genes may be controlled by triplex
formation in 5'-flanking regions.

MATERIALS AND METHODS

(a) Chemicals and Enzymes: Ethidium bromide and DNase I
(pancreatic) were obtained from Sigma. Nucleoside triphosphates
(including \(\text{m}^5\text{dCTP}\)), and \text{E. coli} DNA polymerase I were purchased
from P-L Biochemicals.

(b) Ethidium Fluorescence Assays: These assays rely upon
the enhancement of fluorescence which occurs when ethidium
intercalates into duplex DNA. Ethidium gives no fluorescence
increase in the presence of single-stranded DNA or triplexes. A
pH 8 ethidium assay solution (5 mM Tris-HCl, pH 8, 0.5 mM EDTA,
0.5 \(\mu \text{g/ml}\) ethidium bromide) and a pH 5 ethidium assay solution
(20 mM Na acetate pH 5, 0.2 M NaCl, 0.5 mM EDTA, 1 \(\mu \text{g/ml}\) ethidium
bromide) were used routinely. DNA synthesis and purification
were monitored with the pH 8 solution since this gives the
greatest sensitivity (1/2 \(\mu \text{g}\) of duplex gives a reading in the
middle of the linear range). The pH 5 solution was used for
triplex studies since under these conditions a triplex is stable.
Moreover, because of the presence of ethidium, the high ionic
strength and the low DNA concentration, the conversion of duplex
to triplex is insignificant during the time course of a
measurement. Thus the pH 5 solution can be used to measure the amount of duplex DNA in a mixture of duplex and triplex. Further details of these assays have been presented extensively elsewhere (10,14).

(c) Synthetic repeating-sequence DNAs: Duplex DNAs were prepared as described previously (15,16) with \( \text{dCTP} \) in place of \( \text{dCTP} \) as appropriate. The poly(pyrimidine).poly(purine) DNAs containing \( \text{mC} \) consistently replicate more slowly than both their unmethylated analogues and DNAs containing \( \text{mC} \) with both purines and pyrimidines in one strand (e.g. poly[d(TG)]•poly[d(\text{mCA})]). The presence of TANDEM (a gift from Dr. R.K Olsen, Utah State University) which inhibits poly[d(AT)] synthesis was essential (16), and addition of small amounts of DNase I (0.2 mg/ml) increased the rate of synthesis presumably by increasing the rate of slippage. Although the unmethylated polymer was used as a template, at least 20-fold net synthesis was achieved in all cases so that the extent of methylation was greater than 95%. The single-stranded DNA poly[d(TmC)] was prepared by depurination of the duplex (17). All DNAs were stored at \(-20^\circ\) in T/E buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA).

(d) Thermal denaturation profiles: A Gilford model 260 spectrophotometer equipped with a thermo-programmer was used to record temperature-absorbance transitions. The latter were reproducible to \( \pm 1^\circ \) when the temperature was raised at 0.5°/minute. Unless otherwise stated experiments were performed in a buffer of 10 mM ionic strength (10 mM buffer at a pH close to the pKa, 0.1 mM EDTA and 5 mM NaCl).

(e) pH Stability Curves: Duplexes were dismutated by incubation in 10 mM Na Acetate pH 5, until the pH 5 ethidium assay solution showed no further drop in fluorescence. Solutions were then diluted two-fold by addition to an equal volume of a buffered solution such that the final concentration of reagents was as follows: 0.2 M NaCl, 5 mM MgCl\(_2\) and 25 mM of an appropriate buffer in the pH range 5-10. The buffers used were Na acetate, K phosphate, Tris-HCl, MES, HEPES, TRICINE, CHES, and CAPS (18). The solutions also contained 5 mM Na Acetate, 1 mM Tris and 10 \( \mu \)M EDTA as residuals from the initial incubation at pH 5. Incubation was performed at these various pHs for either
**RESULTS**

While investigating the specificity of a monoclonal antibody which bound to poly(dG)•poly(d\(m^5C\)), we prepared other poly(pyrimidine)•poly(purine) DNAs containing \(m^5C\) (19). Not only did the polymers replicate sluggishly, but also they consistently gave two transitions upon melting at pH 8 and 10 mM ionic strength (Figure 1). Some of the transitions have an unusual and characteristic shape, but the total hypochromicity is typical of a duplex DNA. The first transition has a Tm which would be
Table 1. Tms of synthetic nucleic acids containing \( m^5 C \).

<table>
<thead>
<tr>
<th>Nucleic Acid</th>
<th>Tm(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>poly[d(TTG)].poly[d(( m^5 )CAA)]</td>
<td>69</td>
</tr>
<tr>
<td>poly[d(TTG)].poly[d(CAA)]</td>
<td>65</td>
</tr>
<tr>
<td>poly[d(AT( m^5 )C)].poly[d(GAT)]</td>
<td>75.5</td>
</tr>
<tr>
<td>poly[d(ATC)].poly[d(GAT)]</td>
<td>70</td>
</tr>
<tr>
<td>poly[d(TG)].poly[d(( m^5 )CA)]</td>
<td>77</td>
</tr>
<tr>
<td>poly[d(TG)].poly[d(CA)]</td>
<td>71.5</td>
</tr>
<tr>
<td>poly[d(G( m^5 )C)]</td>
<td>89( ^a )</td>
</tr>
<tr>
<td>poly[d(GC)]</td>
<td>83( ^a )</td>
</tr>
<tr>
<td>poly[d(G)].poly[d(( m^5 )C)]</td>
<td>86( ^a )</td>
</tr>
<tr>
<td>poly[d(G)].poly[d(C)]</td>
<td>71( ^a )</td>
</tr>
<tr>
<td>poly[d(TT( m^5 )C)].poly[d(GAA)]</td>
<td>64( ^b )</td>
</tr>
<tr>
<td>poly[d(TTC)].poly[d(GAA)]</td>
<td>58.5</td>
</tr>
<tr>
<td>poly[d(T( m^5 )C)].poly[d(GA)]</td>
<td>74( ^b )</td>
</tr>
<tr>
<td>poly[d(TC)].poly[d(GA)]</td>
<td>64</td>
</tr>
<tr>
<td>poly[d(T( m^5 )C( m^5 )C)].poly[d(GGA)]</td>
<td>81( ^b )</td>
</tr>
<tr>
<td>poly[d(TCC)].poly[d(GGA)]</td>
<td>73</td>
</tr>
<tr>
<td>poly[d(U( m^5 )C)].poly[d(GA)]</td>
<td>65( ^b )</td>
</tr>
<tr>
<td>poly[d(UC)].poly[d(GA)]</td>
<td>62.5</td>
</tr>
</tbody>
</table>

Tms were measured at pH 8 (ionic strength of 10 mM) except where noted.

\( ^a \) Determined at pH 8 (ionic strength 1.25 mM) (19).

\( ^b \) Tm of the first transition only. See Figure 1.

Tms were measured at pH 8 (ionic strength of 10 mM) except where noted.

\( ^a \) Determined at pH 8 (ionic strength 1.25 mM) (19).

\( ^b \) Tm of the first transition only. See Figure 1.

expected for the methylated polymer. As shown in Table I all nine methylated DNAs have higher Tms than the unmethylated parent although the increase does not correlate well with the \( m^5 C \) content of the DNAs. For example, the difference in Tms between poly[d(G\( m^5 \)C)] and poly[d(GC)] is much less than the difference between poly[d(G)].poly[d(C)] and poly[d(G)].poly[d(\( m^5 \)C)].

Under some circumstances, DNA polymerase I will inadvertantly synthesize a duplex de novo, rather than utilize the given template DNA (16). Thus, the possibility was considered that the second transition was due to some other contaminating DNA. However, several lines of evidence suggest that this is not the case. Firstly, the Tms of the second transitions did not correspond, in general, with the Tms of other duplexes. For
Figure 2. (a) Effect of ionic strength on the Tms of both transitions of poly[d(Tm C)]•poly[d(GA)]. Tms were determined in Tris-HCl buffer pH 8 supplemented with NaCl. Above an ionic strength of 100 mM only one transition was observed; (b) Effect of Ethidium bromide on the Tms of both transitions of poly[d(Tm C)]•poly[d(GA)] determined in 2.5 mM Tris-HCl pH 8. Above a concentration of 2 µg/ml of ethidium only one transition was observed.

For example, the second Tm of poly[d(Tm C)]•poly[d(GA)] is 89°C; this cannot be due to contaminating poly(dG)•poly(dm C) or poly[d(Cm C)] since at 10 mM ionic strength their Tms are over 95°C. Secondly, when the methylated polymers were replicated many fold in the presence of dCTP rather than m dCTP, the result was a duplex DNA which gave rise to only one transition. The Tm was identical to that of the original unmethylated polymer. Thirdly, as shown in Figure 1(c) for poly[d(Tm C)m C]•poly[d(GGA)] and in Figure 2(a) for poly[d(Tm C)]•poly[d(GA)], increasing the ionic strength causes the transitions to converge until only one Tm can be discerned. Similarly, (Figure 2(b)) ethidium increases the Tm of the first transition while the second transition is essentially unaffected. Then the two transitions coalesce at high ethidium concentrations. Poly[d(TM C)]•poly[d(GAA)] and poly[d(TM C)m C]•poly[d(GGA)] gave similar results (data not shown). These observations are very difficult to reconcile with a contaminated DNA preparation. Rather, they suggest that, after melting of the methylated poly(pyrimidine)•poly(purine) duplex DNA, some other structure is formed which does not bind ethidium.
and whose Tm is rather insensitive to ionic strength.

Triplexes require a protonated C and thus a pH of 8 seems rather high for triplex formation. Nevertheless, this possibility was investigated because of the propensity of poly(pyrimidine) · poly(purine) DNAs to form this structure (10). (See Figure 3). Poly[d(Tm$^5$C)] · poly[d(GA)] was allowed to equilibrate with poly[d(Tm$^5$C)] in the stoichiometry of 2:1 (phosphate). Upon melting (Figure 3(a)) it can be seen that the first transition disappears and the second transition has almost twice the hyperchromicity compared to the duplex alone. The effect of pH was also studied (Figure 3(b)). At pH 9, only one transition is observed with a Tm which corresponds to the first transition at pH 8. In other words, at pH 9 only the duplex transition is observed. At a pH of 7 (data not shown) only one transition was observed, beginning at 99°C. Thus, as expected, the Tm of the triplex increases as the pH is lowered. Finally, Figure 3(b) demonstrates that the unmethylated poly[d(TC)] · poly[d(GA)] exhibits the same behavior as its methylated counterpart but at a lower pH. All these results are consistent with triplex formation.

Several further points can be made with regard to Figures 1-3. Under the same conditions, the triplex formed from poly[d(Tm$^5$C)] · poly[d(GA)] has a higher Tm than that from poly[d(Tm$^5$Cm$^5$C)] · poly[d(GGA)]; this may reflect the destabilizing

Figure 3. Thermal denaturation profiles for (a) poly[d(Tm$^5$C)] · poly[d(GA)] · poly[d(m$^5$CT)] at pH 8; (b) poly[d(Tm$^5$C)] · poly[d(GA)] at pH 9 and poly[d(TC)] · poly[d(GA)] at pH 7. In each case the ionic strength was approximately 10 mM.
effect, in the latter case, of having two positively charged $m^5C\cdot G\cdot m^5C$ base-triads stacked together. It is noteworthy that at pH 8 no second transition is observed for poly(dC)•poly[d($m^5C$)] (19). Thus the relative stability of the duplex and triplex is determined by the (G-C) content of the DNA. The presence of a positively charged base-triad may also explain the lack of intercalation of ethidium (Figure 2). On the other hand, the insensitivity of the Tm of the triplex to ionic strength is surprising since, if anything, the charge density of a triplex is higher than that of a duplex. Similar effects of ethidium and ionic strength were also observed for poly[d(Tm$^5C$)]•poly[d(CAA)] and poly[d(Tm$^5Cm^5C$)]•poly[d(GGA)]. (Data not shown). Finally, the behavior of poly[d(Um$^5C$)]•poly[d(GA)] at pH 8 is instructive. (Figure 1(d)). This polymer has the same number of methylated pyrimidines (Tm$^5U$) as poly[d(TC)]•poly[d(GA)] which does not give two transitions at pH 8. Consequently, it would appear that the facility of triplex formation is a property of $m^5C$ rather than methylation itself. This is surprising since the pKas of C and $m^5C$ are indistinguishable (20).

If a biological role is to be assigned to triplexes, then they must be able to exist at neutral pHs. It is difficult to assess the effect of pH on the duplex to triplex dismutation from thermal denaturation studies because the temperature is constantly changing. Thus the pH stability was investigated by a more direct method to assess: (a) the pH at which the triplex will form from the duplex and (b) once formed, what pH is required to return to the duplex. The following studies on triplex stability were performed with equal pyrimidine/purine strand ratios in accord with the hypothetical situation in vivo and also under physiological salt conditions (0.2 M NaCl and 5 mM MgCl$_2$) (Figure 4).

Poly[d(TC)]•poly[d(GA)] and the methylated polymer were incubated at low pH until there was no further drop in fluorescence. Samples of each were then buffered to a higher pH and either heated at 60° for 2 hours (which will bring the system to equilibrium). Or, alternatively, they were left at 20° for 10 minutes. After addition to the pH 5 ethidium assay solution the fluorescence was read immediately. Any increase in fluorescence
Figure. 4. pH stability curves for (a) poly[d(TC)].poly[d(GA)] and (b) poly[d(Tm\textsuperscript{5}C)].poly[d(GA)]. The fluorescence in the pH 5 ethidium assay solution is shown as a function of the pH after incubation for either 2 hours at 60\textdegree\ (\bullet) or 10 minutes at 10\textdegree\ (O). See text for details.

from the original value implies that the triplex (together with the excess poly(purine) strand) is reverting back to the duplex. As shown in Figure 4(a), poly[d(TC)].poly[d(GA)] is half converted to the duplex at pH 5.8 (the pHm) after heating at 60\textdegree\ for 2 hours. By contrast the pHm for the sample left at 20\textdegree\ for 10 minutes is 7.2. Thus the dismutation exhibits hysteresis. Both poly[d(TTC)].poly[d(GAA)] and poly[d(TCC)].poly[d(GGA)] behave in a similar manner (data not shown). The result for poly[d(Tm\textsuperscript{5}C)].poly[d(GA)] is again similar, except the curves are shifted to higher pHs. The pHm under equilibrium conditions is 7.5 and for short periods of time some triplex remains even at pH 9. In other words, at neutral pH, a poly(pyrimidine).

poly(purine) DNA containing \textsuperscript{5}C would be expected to dismutate to a triplex + poly(purine). Once formed, this structure may remain metastable with respect to the duplex even when the pH is raised or the methyl groups are removed.

DISCUSSION

The pKa's of both C and \textsuperscript{5}C are 4.5 (20) yet the triplex formed from poly[d(Tm\textsuperscript{5}C)].poly[d(GA)] is stable at pH's nearly 2

6611
Figure 5. Models for the control of gene expression mediated by triplex formation in the 5'-flanking region. Two variations are shown depending on the location of the DNA which donates the additional pyrimidine strand. = Pyrimidine strand, = Purine strand. A hypothetical single strand binding protein is shown as a hatched circle.

units higher than the triplex formed from poly[d(TC)]. poly[d(GA)]. There is no compelling physico-chemical explanation for this unexpected observation. However, the discovery of triplex formation at neutral pH is intriguing. Since pyrimidine-purine tracts have been found in the 5'-flanking regions of many eucaryotic genes (see Introduction), it seems possible that triplexes may be involved in gene regulation. Two models are shown schematically in Figure 5.

In the top diagram the additional pyrimidine strand which is required for triplex formation has been donated by a segment of DNA located away from the immediate vicinity of the gene. (The folding of the DNA must satisfy the condition that the two pyrimidine strands are antiparallel). A single-strand binding protein has been included in the model; this would stabilize the triplex by sequestering the free purine strand. In this configuration the gene would be switched "off" since RNA polymerase would not be able to gain access to the promoter. As demonstrated in this work, if the pyrimidine strands contain significant levels of m5C, then the triplex configuration will be favoured under physiological conditions. An alternative
configuration is shown at the bottom of Figure 5, where the donating DNA is part of the 5'-flanking region of the gene itself. In this case, the looped out purine strand might form an entry point for RNA polymerase, so that the gene would be switched "on".

These models have several attractive features and can rationalize some of the observations concerning SI-nuclease hypersensitive sites (SHS) in pyrimidine-purine tracts. For example the displaced purine strand would be digested and all four strands may be nuclease sensitive at the ends of the structure. The model also explains the requirement for negative supercoils before SI sensitivity is observed (6). Since the donating duplex must be unwound, then triplex formation will be favoured by a negative superhelix density in the loop of DNA.

On the other hand there are several problems. For example, methylation of the pyrimidine strand has not yet been demonstrated. Indeed the relationship between methylation and gene regulation is still unclear (recently reviewed in (21)). Also, in the absence of a topoisomerase triplex formation will create topological problems.

Other authors have suggested that SHS are due to slippage (5-8) but it is not clear how this would relate to any regulation of gene expression. Since triplexes can exist at neutral pH, their possible involvement in gene expression should now be considered.

**ACKNOWLEDGEMENTS**

This work was supported by the Medical Research Council of Canada with grants to J.S.L. and A.R.M.

**REFERENCES**