THERMODYNAMIC STUDIES ON SYNTHETIC LAC OPERATOR CORE

R.C. Patel* and R. Zupcak
Chemistry Department, Clarkson College of Technology, Potsdam, N.Y. 13676, U.S.A.

S.A. Narang
Division of Biological Sciences, National Research Council of Canada
Ottawa, Canada K1A OR6

Received November 7, 1978

SUMMARY. The nine base pairs long central region of the lac operator gene forms a stable double helix. A comparison of melting temperatures with other biologically useful oligonucleotides indicates the importance of specific base sequence. Binding constants measured with ethidium bromide (1.7 x 10^5 M⁻¹), tyrosine (4.0 x 10^3 M⁻¹), and glutamine (1.5 x 10^3 M⁻¹), are interpreted in terms of the involvement of a relatively small number of amino acids in the lac operator-repressor interaction.

INTRODUCTION

In the lactose operon, the operator (O) gene controls the expression of three adjacent genes z, y, a, by very tight (dissociation constant of 10⁻¹¹ to 10⁻¹³ M) and specific binding with the lac repressor protein (1). The understanding of such protein DNA interactions is of paramount importance in other bacterial operons (2) and in enzymes such as restriction endonucleases (3). Despite this, very little physico-chemical information is available on biologically important, short DNA sequences. A number of models have been proposed to explain the lac operator repressor interaction with very little experimental evidence for support (1,2,4). One feature of these models is based on the existence of open regions in the DNA molecule, e.g. the cruciform structure shown in Fig. 1. In this report, we describe UV and fluorescence studies on the following four biologically important oligonucleotides:

(I) 5'-CCGGATCCGG-3' (Linker I, self-complementary)
(II) 5'-AACAGCTGTT-3' (Linker II, self-complementary)

* To whom correspondence regarding this manuscript should be addressed.
Fig. 1 Cruciform structure of the lac operator proposed to explain the specific lac operator-repressor interaction.

(III) 5′-GAGCGGATA-3′ (III U)

3′-CTGCCTAT-5′ (III L)

(IV) 5′-G₄C₄-3′ (self-complementary)

III U and III L represent the upper and lower strands of the central region of the lac operator gene. Linker I and Linker II form self-complementary double helices, provide a general method for inserting specific DNA sequences into cloning vehicles, and have been used recently to insert synthetic lac operator DNA into plasmid PMB9 DNA (5). They are also useful model substrates for studying the mechanism of cleavage of DNA by a number of restriction enzymes.

MATERIALS AND METHODS

All oligonucleotides were chemically synthesized by a recently developed, improved phosphotriester method (6), and their characterization (sequence determination by two dimensional homochromatography (7,8) and purification have been fully described. Three different types of experiments indicate that the central region of the lac operator, corresponding to the open loop region in Fig. 1, forms a very stable duplex, as shown by the UV melting curves in Figs. 2a and 2b. The fluorescence of ethidium bromide, which binds specifically to double helical DNA (9), is greatly enhanced when bound. Experiments were carried out in the presence of a large excess of ethidium bromide, and the data treated on the basis of a modified Benesi-Hildebrand equation (10). In every
Fig. 2 UV melting curves for lac operator duplex (III U + III L) in a medium of 0.25 M ionic strength (Na\(^+\)), pH = 7.0 (sodium cacodylate Buffer), determined at 260 nm (curve a), and 270 nm (curve b). The relative change in fluorescence intensity of ethidium bromide bound lac operator is shown by curve c. (excitation wave length 330 nm, emission wavelength 565 nm.)

In this case, excellent linear plots, reproducible to well within ± 5% over periods of several weeks, were obtained, indicating the formation of 1:1 complexes under the prevailing experimental conditions. Preliminary work on fluorescence lifetime studies carried out with the collaboration of Dr. F. Garland further substantiate the existence of 1:1 complexes. The binding of tyrosine and glutamine was determined from competition experiments:

\[
\begin{align*}
E + D &= ED \\
C + D &= CD
\end{align*}
\]

\[
K_A = \frac{[ED]}{[E][D]} \\
K_C = \frac{[CD]}{[C][D]}
\]

where \(E\) = ethidium bromide, \(D\) = oligonucleotide duplex, \(C\) = competitor tyrosine or glutamine. The following cubic equation was derived:

\[
\begin{align*}
[D]^3(K_A K_C) + [D]^2(K_A + K_C + K_A K_C E_o" + K_A K_C O_o" - K_A K_C D_o") + [D](1 + K_A E_o" + K_C O_o" - K_A D_o" - K_C D_o") &- D_o" = 0
\end{align*}
\]

Subscript "o" refers to the total, analytical concentration of the species in question. A nonlinear least squares program was used to find the best value of \(K_C\) to fit the observed decrease in the ethidium bromide fluorescence intensity using up to nine experimental points for each amino acid, as well as by an interpolation method, and identical results obtained. The standard deviation of both fits were better than ± 10%. In the case of tyrosine, it was possible to
TABLE 1

Comparison of Thermodynamic Parameters of Oligonucleotides with Defined Base Sequence

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Chain Length</th>
<th>No. GC Base Pairs</th>
<th>Ionic Strength^1 (Medium)</th>
<th>T_m(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I)^2</td>
<td>10</td>
<td>8</td>
<td>0.05M (Na^+), 0.25M (Na^+), 0.25M (Mg^2+)</td>
<td>39.5, 54.0, 54.0</td>
</tr>
<tr>
<td>(II)^2</td>
<td>10</td>
<td>4</td>
<td>0.25M (Na^+)</td>
<td>43.5, 48.0</td>
</tr>
<tr>
<td>(III)</td>
<td>9</td>
<td>5</td>
<td>0.25M (Na^+)</td>
<td>40.0, 45.0</td>
</tr>
<tr>
<td>(IV) G_4C_4</td>
<td>8</td>
<td>8</td>
<td>1.0M (Na^+)</td>
<td>49.5</td>
</tr>
<tr>
<td>(V) A_9GCU_4</td>
<td>10</td>
<td>2</td>
<td>0.05M (Na^+)</td>
<td>25.0 (11)</td>
</tr>
<tr>
<td>(VI) A_8CG + CGU_B</td>
<td>10</td>
<td>2</td>
<td>0.05M (Na^+)</td>
<td>16.0 (11)</td>
</tr>
<tr>
<td>(VII) A_5G_3 + C_3U_5</td>
<td>R</td>
<td>3</td>
<td>0.05M (Na^+)</td>
<td>24.0 (11)</td>
</tr>
</tbody>
</table>

TABLE 2

Comparison of Binding Constants for Synthetic Oligonucleotides^1

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Binding Constant (M^-1)</th>
<th>-ΔG° (kcal mol^-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I)^2 + ethidium bromide</td>
<td>1.3 x 10^5</td>
<td>7.02</td>
</tr>
<tr>
<td>(II)^2 + ethidium bromide</td>
<td>1.2 x 10^5</td>
<td>6.97</td>
</tr>
<tr>
<td>(III) + ethidium bromide</td>
<td>1.7 x 10^5</td>
<td>7.18</td>
</tr>
<tr>
<td></td>
<td>+ tyrosine</td>
<td>4.0 x 10^3</td>
</tr>
<tr>
<td></td>
<td>+ glutamine</td>
<td>1.5 x 10^3</td>
</tr>
<tr>
<td>(IV) G_4C_4 + ethidium bromide</td>
<td>2.8 x 10^5</td>
<td>7.48</td>
</tr>
</tbody>
</table>

make use of the decrease in the intrinsic fluorescence of bound tyrosine as an independent method of determining K_C. Since the fluorescence intensity changes were much smaller than for ethidium bromide, we were only able to obtain an estimate of K_C = 2 x 10^3 M^-1 from this experiment. This value is not greatly different from the more accurate determination from competition experiments. (Table 2).

1. All experiments were carried out at a pH of 7 (cacodylate buffer).
2. These oligonucleotides form self complementary double helices.
The oligonucleotides were saturated with ethidium bromide, and changes in the fluorescence intensity recorded as a function of temperature (Fig. 2c). A very steep change in the fluorescence intensity is found in each case in the temperature range corresponding to the melting of the double helix.

RESULTS AND DISCUSSION

Table 1 summarizes data from typical melting curves determined under a variety of conditions, and also presents information from the literature (11) on additional oligonucleotides. The data in Table 1 show that the melting temperature ($T_m$) increases with increasing ionic strength, and, for Linker II and lac operator, with the presence of Mg$^{2+}$. Despite the relatively short chain lengths of the four experimental oligonucleotides, highly stable duplexes are formed.

Table 1 also indicates that the stability of double helical segments, as reflected by $T_m$ values, is markedly sequence dependent, as opposed to A-T, G-C content alone. For example, decamers V and VI, each with two G-C base pairs, have $T_m$ values in 0.05 M Na$^+$ medium which differ by 9°C. Linker I, with four more G-C base pairs than Linker II, has a $T_m$ value in 0.25 M Na$^+$ medium which is higher than that of Linker II by 10.5°C. In the presence of Mg$^{2+}$, this difference decreases to 6°C. Linker II, with one less G-C base pair than lac operator, has a higher $T_m$ value in 0.25 M ionic strength medium both in the presence and absence of Mg$^{2+}$. Sequence IV, comprised totally of G-C base pairs, has a $T_m$ value in 1.0 M Na$^+$ medium which is 4.5°C less than that of Linker I (which has the same number of G-C base pairs) in only 0.25 M Na$^+$ medium.

It is evident from these experimental results that G-C base pair effects for these oligonucleotides are not additive, and that stability of a given oligonucleotide is a function of its defined base sequence. Due to a critical lack of data on such compounds, it is difficult to describe the melting curves quantitatively on the basis of statistical mechanical formulations (12). However, using the model of an all or none transition (12) as a first approximation, it was possible to calculate a relatively constant enthalpy change of $\sim$ 5 kcal mol$^{-1}$ per base pair for all four oligonucleotides, indicating that entropy changes are probably more important in determining variations in duplex stability.
The relatively high melting temperature of the lac operator core is to be compared with the value of 67°C (in 0.15 M NaCl, 0.015 M sodium citrate) determined for lac operator that is 27 base pairs long (13). Clearly, a very large contribution to duplex stability can be attributed to the lac operator core. It can be argued, in principle, that under actual in vivo conditions, long range interactions along the DNA helix could alter the operator structure leading to different requirements for recognition by the repressor. This seems unlikely, however, in view of the experimental finding that the minimum number of base pairs necessary for recognition is 17 (14).

The values for the ethidium bromide binding constant listed in Table 2 are of the same order of magnitude observed by LePecq and Paoletti (9) for the binding of ethidium bromide to calf thymus DNA (approximately $10^5$ M$^{-1}$ at sodium ion concentrations of 0.05-0.25 M). Further comparisons may be made to the values calculated for the binding constants of the dye proflavine to calf thymus DNA (approximately $10^5$ M$^{-1}$ at a sodium ion concentration of 0.2 M) and the G-C specific drug actinomycin D to calf thymus DNA (approximately $10^6$ M$^{-1}$) (16). The decrease in the binding constants from ethidium bromide to glutamine can be understood on the basis of intercalation and base stacking interactions (17,18) for ethidium bromide and tyrosine, as compared with hydrogen bonding interaction for glutamine (18).

In mapping studies of lac repressor, Files and Weber have found that tryptic digestion of repressor under non-denaturing conditions readily removes 59 amino acids from the amino-terminal end of the polypeptide change (19). Longer digestion removes an additional 20 or more amino acids from the carboxy-terminus, leaving a highly trypsin-resistant core molecule. The lac repressor tetrameric structure and inducer-binding activity are retained by the tryptic core. Most significant, however, is the loss of operator-binding capacity as the amino-terminus of repressor is degraded. Many or all of the possible trypsin cleavage sites in the amino-terminal region are available to attack by the enzyme, indicating that this part of the polypeptide chain is exposed to the environment.
Müller-Hill has also noted the importance of the amino-terminus of lac repressor (1). He has concluded that the operator binding region is divided at least into two subregions: 1) a region of specific DNA recognition between residues 50 and 58, and 2) a region of unspecific and specific DNA recognition between residues 5 and 50. Miller's work with mutant repressor proteins substantiates these findings (20). Tyrosine 17 and glutamines 18, 54, and 55 have been found to be essential in the binding of repressor to operator; they cannot be replaced by any other than the original amino acid. Thus, tyrosine and glutamine are very probable amino acids which make contact with the base pairs of lac operator. The exact free energy of interaction of a given amino acid with lac operator has not been available so far. The results in Table 2 show that for tyrosine and glutamine, \( \Delta G^\circ \) is \( \sim -4 \) kcal mol\(^{-1} \), indicating that the total number of amino acids involved in the lac operator-repressor interaction, with a total \( \Delta G^\circ \) of \( \sim -18 \) kcal mol\(^{-1} \), is probably small (perhaps between 4 - 6). This is particularly significant if one considers that the tetrameric lac repressor binds to the operator.

Acknowledgments

Support of this work by NIH grant No. 1 RO1 GM24523-01 is gratefully acknowledged. We thank Mr. Harold Trimm for technical assistance.

References