Volume 336, number 2, 293–298 FEBS 13417 December 1993

Different flexibility of the upstream regulatory regions of two differently expressed pea *rbcS* genes studied by theoretical evaluation of DNA distortion energy and cyclization kinetics

Stefano Cacchione[a], Pasquale De Santis[b], Maria Savino[a,*]

[a]*Dipartimento di Genetica e Biologia Molecolare* and [b]*Dipartimento di Chimica, Università di Roma 'La Sapienza', P. le A. Moro 5, Roma, Italy*

Received 24 September 1993

Different superstructural features of the upstream regulatory regions of the two pea genes *rbcS-3A* and *rbcS-E9* have been derived using a theoretical method, developed in our laboratory a few years ago, which evaluates the DNA distortion energy from a matrix of the deviations of the 16 possible dinucleotides from the standard B conformation. The theoretical analysis, which predicts different flexibilities of the regulatory regions of the two genes, is in satisfactorily good agreement with experimental evaluations from gel electrophoretic mobility and cyclization kinetics, suggesting a possible model to explain the largely different transcription efficiencies of the two genes.

Sequence dependent DNA superstructure; DNA distortion energy calculation; Cyclization kinetics; *rbcS* gene

1. INTRODUCTION

The intrinsic ability of B-DNA to translate the deterministic fluctuations of base sequence in superstructural elements, is generally considered of relevance in determining its physicochemical as well as its biological properties. Although the physical origin of DNA curvature is still a matter of debate, sound evidence has been accumulated of a major contribution of the nearest-neighbour differential interactions within the dinucleotide steps [1–5]. We have previously shown that DNA curvature can be predicted by integrating the slight local deviations in terms of roll and tilt angles, obtained on the basis of energy conformational calculations for the 16 dinucleotide steps [2]. This method was successfully applied to predict the curvature of numerous synthetic as well as natural DNAs [2,4], in surprisingly good agreement with experimental measurements, mainly based on gel electrophoretic retardations.

The obtained results encouraged us to explore the superstructural features of DNA regulatory regions, where they could have a relevant role either in specific interactions with proteins or in favouring protein–protein interactions.

In this paper, we analyse the superstructural features of the regulatory regions of two pea genes which code for the small subunit of the enzyme ribulose-1,5-bisphosphate carboxylase, *rbcS-3A* and *rbcS-E9*. The largely different transcription efficiency of these two genes has been connected, on the basis of extensive genetic and biochemical analyses, to the upstream region from −410 to the startsite of transcription [6,7]. In particular, analysis of the *rbcS-3A* gene has uncovered three independent upstream regions, which contain light responsive elements (LREs), centered at about −150, −220 and −370, which bind to the regulatory protein GT1 [7,8]. The presence of multiple regulatory elements may enable *rbcS* genes to adapt to different environment conditions and strongly suggests that a multiproteic complex regulates gene transcription.

We have previously studied the superstructural features of the LRE positioned from −166 to −112 in different *rbcS* genes, and found a correlation with the different transcription efficiencies, suggesting that LREs different flexibility could be relevant in the interactions with GT1 [9–11].

In this paper we analyse the whole region from −410 to the startpoint of transcription of the two genes *rbcS-3A* and *rbcS-E9*, since the thermodynamic and kinetic aspects of the transcriptional complex should depend on the superstructural features of the regulatory region as a whole, as well as on those of the individual regulatory sequences. In order to characterize the two DNA regions, 410 bp fragments with *Hin*dIII protruding ends have been obtained from the regulatory regions of *rbcS-3A* and *rbcS-E9* (see Fig. 1), and studied both by theoretical evaluation of DNA distortion energies and by means of cyclization kinetics.

2. MATERIALS AND METHODS

410-3A was obtained from an *Hin*dIII/*Hin*dIII fragment (−402; −8 from the transcription startsite) from the clone pUC18-3A/E9 [12].

*Corresponding author. Fax: (39) (6) 444 0812.

410-E9 was obtained from an *Eco*RI/*Eco*RV fragment (−1100; +24) from the clone pBR325-E9 [13], subsequently digested with *Hin*fI (−406). The two fragments, having 5′ protruding ends, were filled with Klenow polymerase, ligated with 12mer *Hin*dIII linkers (Boehringer), digested with *Hin*dIII (cutting site on *rbcS-E9* at −8), and ligated in the *Hin*dIII site of pUC18. The two recombinant plasmids were used to transform *E. coli* JM83 competent cells. Sequences were confirmed by the Sanger dideoxy sequencing approach modified for double stranded DNA.

1 μg of 410-3A and 410-E9 were dephosphorylated by treatment for 40 min with calf intestinal alkaline phosphatase at 0.02 unit/pmol 5′ ends of DNA. The reaction mixtures were then heated to 70°C for 15 min and extracted two times with phenol and four times with diethyl ether. DNAs were subsequently labeled with ^{32}P by incubation with 100 μCi/μg [γ^{32}-P]ATP and T4 polynucleotide kinase. Unincorporated label was removed by spun column cromatography over 1 ml syringes packed with Sephadex G_{50} (Pharmacia). Cyclization experiments were done at 20°C at a DNA concentration of 3.7×10^{-10} M in 60 μl of a solution 10 mM Tris, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM ATP. T4 DNA ligase (USB) was opportunely diluted and added to the reaction mixtures, followed by manual mixing. At specific time intervals, 10 μl aliquots were removed and quenched with EDTA, pH 8 (final concentration 50 mM) and heat (65°C, 10 min). Samples were electrophoresed on 4% NuSieve 3:1 Agarose gels (FMC). After the run, gels were dried and autoradiographed. The radioactivity of each band was measured by scanning the autoradiographs on a LKB densitometer.

3. RESULTS AND DISCUSSION

To predict the circularization probability of the two examined sequences we used a recently developed theoretical method [14,15], which allows the evaluation of DNA circularization distortion energy in terms of the

a

410-3A

```
  1                                                                   100
5' AGCTTGCGAGCTTGGACAGGAACAAATGTTACCCATACATAAAAGATATTTGTGAAGTAACAGTCACAAAATTCCATGAGGCCAACATACTACAATTGAA
3'      ACGCTCGAACCTGTCCTTGTTTACAATGGGTATGTATTTTCTATAAACACTTCATTGTCAGTGTTTTAAGGTACTCCGGTTGTATGATGTTAACTT

HindIII

101                                                                   200
  TTTTCATGGATACAATTCTTACAAAATAAAAATATCGACATAACCACCATCACACATTTACACTCTTCACATGAAAAGATAAGATCAGTGAGGTAATATC
  AAAAGTACCTATGTTAAGAATGTTTTATTTTTATAGCTGTATTGGTGGTAGTGTGTAAATGTGAGAAGTGTACTTTTCTATTCTAGTCACTCCATTATAG


201                                                                   300
  CACATGGCACTGTCCTATTGGTGGCTTATGATAAGGCTAGCACACAAAATTTCAAATCTTGTGTGGTTAATATGGCTGCAAACTTTATCATTTTCACTAT
  GTGTACCGTGACAGGATAACCACCGAATACTATTCCGATCGTGTGTTTTAAAGTTTAGAACACACCAATTATACCGACGTTTGAAATAGTAAAAGTGATA


301                                                                   400
  CTAACAAGATTGGTACTAGGCAGTAGCTAAGTACCACAATATTAAGACCATAATATTGGAAATAGATAAATAAAAACATTATATATAGCAAGTTTTAGCA
  GATTGTTCTAACCATGATCCGTCATCGATTCATGGTGTTATAATTCTGGTATTATAACCTTTATCTATTTATTTTTGTAATATATATCGTTCAAAATCGT


401      410  HindIII

  GAAGCTCGCA     3'
  CTTCGAGCGTTCGA 5'
```

b

410-E9

```
  1                                                                   100
5' AGCTTGAATCATTAATAAAACTTTCCACTTTTATTTAATTTTTTCAATTCAGTTTCATCATCAAAGCAAATGTTTCCCTGCCATCTATCTGTCAACACTA
3'      ACTTAGTAATTATTTTGAAAGGAGAAAATAAATTAAAAAAGTTAAGTCAAAGTAGTAGTTTCGTTTACAAAGGGACGGTAGATAGACAGTTGTGAT

HindIII

101                                                                   200
  ACATCGAATGTACTTATCTCATTAGTTTAAATTATTGTTTGATCATGTTTAATCCTTATACTGTTGTTAGTTTTTTCAGTTAGCTTAGTGGGCATCTTAC
  TGTAGCTTACATGAATAGAGTAATCAAATTTAATAACAAACTAGTACAAATTAGGAATATGACAACAATCAAAAAAGTCAATCGAATCACCCGTAGAATG


201                                                                   300
  ACGTGGCATTATTATCCTATTGGTGGCTAATGATAAGGTTAGCACACAAAACTTTTCAATCTTGTGTGGTTAATATGACTGCAAAGTTTATCATTTTCAC
  TGCACCGTAATAATAGGATAACCACCGATTACTATTCCAATCGTGTGTTTTGAAAAGTTAGAACACACCAATTATACTGACGTTTCAAATAGTAAAAGTG


301                                                                   400
  AATCCAACAAACTGGTTCTAGGCAGTGGCCAACTACCACAATTTTAAGACCATAATATTGGAAATAGAAAAATCAATACATTATATAAAGCAAGTTTTAG
  TTAGGTTGTTTGACCAAGATCCGTCACCGGTTGATGGTGTTAAAATTCTGGTATTATAACCTTTATCTTTTTAGTTATGTAATATATTTCGTTCAAAATC


401      410  HindIII

  TACAAGCTCA     3'
  ATGTTCGAGTTCGA 5'
```

Fig. 1. Nucleotide sequences of the fragments (a) 410-3A and (b) 410-E9.

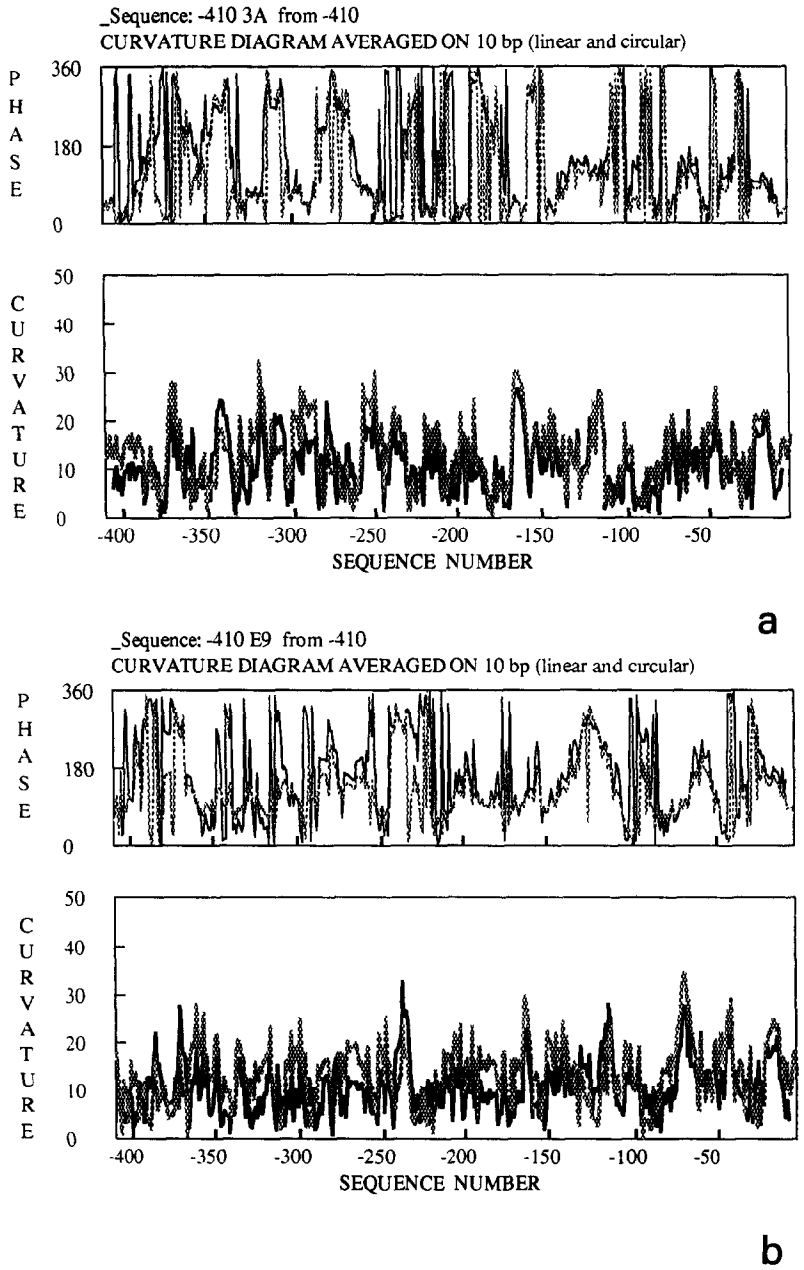

Fig. 2. Theoretical curvature profiles of the DNA fragments reported in Fig. 1. (a) 410-3A; (b) 410-E9. The curvature vector is reported as modulus and phase, averaged each 10 bp. The dark line refers to linear DNA, while the grey line refers to circular DNA.

nucleotide sequence as well as the structure of the circular DNA, which is characterized by the lowest distortion with respect to the linear form.

Fig. 2 illustrates the theoretical curvature diagrams of 410-3A and 410-E9 DNA tracts (see Fig. 1). The curvature modulus defines the local deviations of the double helix axis, whereas the direction of such deviations is indicated in the corresponding phase diagram with respect to the dyad axis of the first base pair of the sequence. When integrated along the chain, they produce the superstructures of the linear as well as the circular forms as shown in the stereoviews of Fig. 3a and b. It is worth noting that the superstructures shown in Fig. 3 represent the hypothetical unperturbed forms in the absence of thermal fluctuations. Although Olson et al. [16] have recently shown the importance of thermal fluctuations on the solution properties of curved DNAs, the validity of our representation relies mainly on the internal comparison of different DNA segments, which allows an easy localization of more bent tracts, as well as the visual comparison between different DNAs. Furthermore, the mathematical method used to obtain the corresponding most stable circular DNAs assures the maximum invariance of the local structures compatible with the circular form, and allows a reliable evaluation of the circularization energy, since the influ-

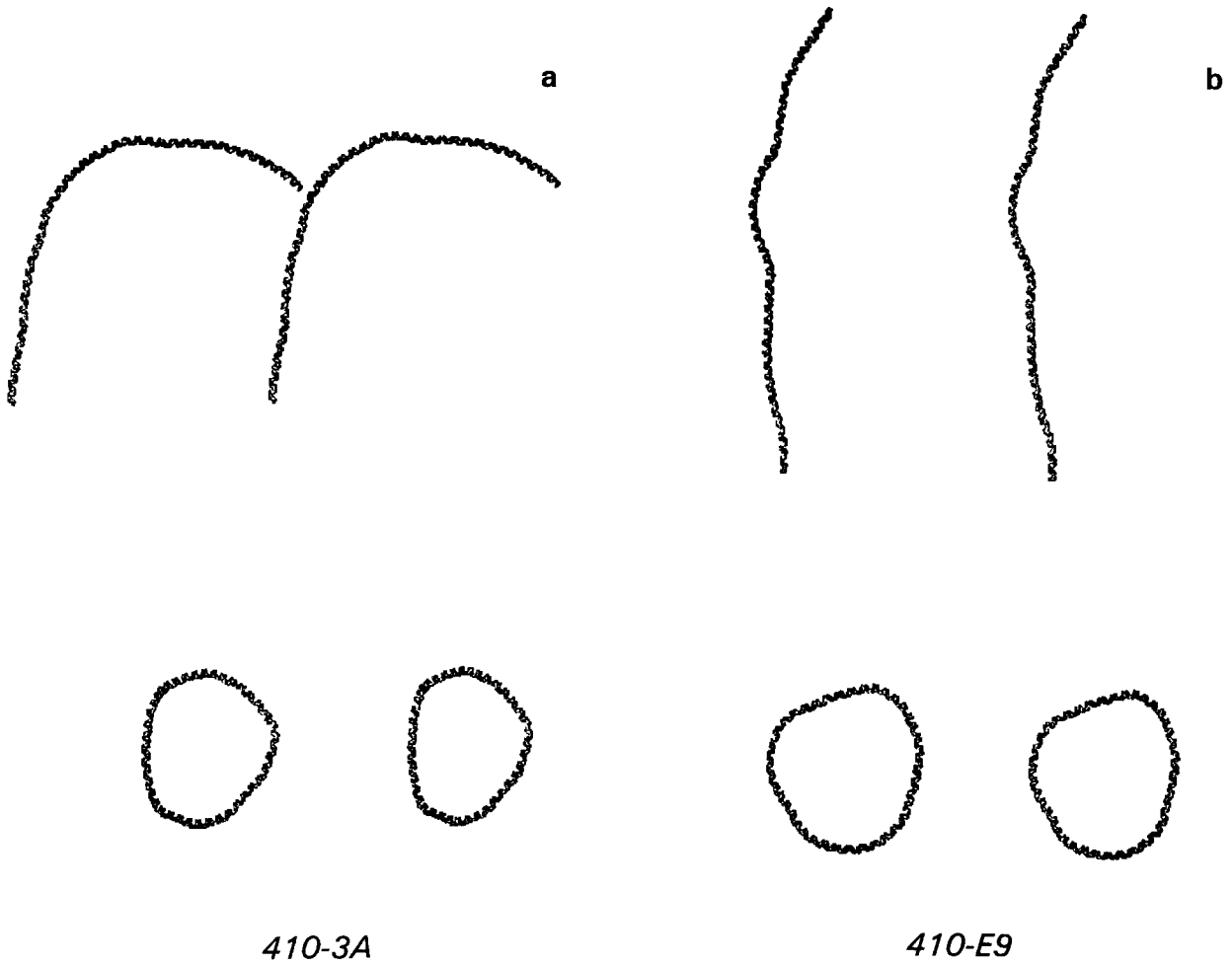

Fig. 3. Stereoviews of the DNA fragments (a) 410-3A and (b) 410-E9, in the linear (top) and circularized form (bottom).

ence of thermal fluctuations is assumed to be similar in the two forms.

The superstructural features of the regulative regions from −410 to the start-site of transcription, appear largely different in the two cases in spite of the high homology of the sequences up to −200 [17]; in particular, 410-3A is characterized by the presence of an higher number of curved tracts than 410-E9, whose stereoview corresponds approximately to that of a straight DNA.

The theoretical differential circularization energy values, ΔE, are calculated in terms of that of an equivalent straight DNA tract, $\Delta E°$. Adopting the average values of the persistence length, $P = 500$ Å, and of the torsional modulus, $C = 3.4 \times 10^{-19}$ erg · cm for a straight DNA [18], it is possible to evaluate the differential circularization energy values in absolute units (De Santis et al., in preparation). These values are reported in Table I; they represent the sequence-dependent activation energy contributions to the activation free energy of the circularization reaction for the same protruding ends.

During the last few years, ligase catalyzed DNA ring closure has been used as an useful approach for the determination of sequence directed bends in the helical axis of B-DNA. In fact, formation of the circular DNA molecules that are substrates for ligation reactions is expected to require large fluctuations in the dimensions of the DNA molecule, since both ends of the molecule must come close together to ligate. The cyclization equilibria are therefore expected to be very sensitive to sequence-directed or protein-induced DNA bending [18].

The widely used factor to define cyclization equilibria is the ring closure probability, J factor, defined as the ratio $J = K_c/K_a$ of the equilibrium constant for cyclization to that for bimolecular association; it can be evaluated experimentally by calculating the ratio of the rate constants for cyclization (k_1) and bimolecular association (k_2) [19,20]. J factor depends on the concentration of one end of a DNA molecule near the other end, and on the angular orientation between the two ends, which is a function of the DNA length and twist.

To study the flexibility of the *rbcS-3A* and *rbcS-E9* regulatory regions, we have compared the J values derived for the 410 bp fragments, having the same length and the same protruding ends. The twist values of the two examined sequences have been considered equal to

the average value of 10.4. The obtained results are reported in Fig. 4a and b. The larger flexibility of the regulatory region of the *rbcS-3A* gene with respect to *rbcS-E9*, is clearly evident from the higher kinetic constant at all the ligase concentrations used (see Table I). The ratio between the kinetic constants, $k_{1\text{-}3A}/k_{1\text{-}E9}$, is equal to the ratio between J_{3A}/J_{E9}, since the dimerization kinetic constants is equal in the two sequences having the same terminals (experimentally tested, not shown).

The ratio between the J factors of the two regulatory sequences allows an experimental evaluation of their activation free energy difference:

$$J_{3A}/J_{E9} = [\exp -(\Delta Gc_{3A} - \Delta Gd)/RT]/[\exp -(\Delta Gc_{E9} - \Delta Gd)/RT]$$

where ΔGc_{3A} and ΔGc_{E9} represent the circularization activation free energy of the two regulative sequences, whereas their dimerization activation free energy ΔGd is considered practically equal.

Thus $RT \ln(J_{3A}/J_{E9}) = \Delta Gc_{3A} - \Delta Gc_{E9} = 0.7$ kcal per mol of cycle, against a theoretical value of 0.6 kcal per mol of cycle (see Table I), since the entropy of activation can be considered equal for the two molecules; it is in fact related to the probability of ends encountering for Gaussian worm-like chains with the same length.

The satisfactory agreement between theoretical prediction and experimental evaluation of DNA flexibility suggests that: (a) the nearest neighbour approximation works satisfactorily in the prediction of DNA flexibility as experimentally tested by cyclization kinetics; (b) the flexibility of DNA sequences, such as DNA regulatory regions, can be easily derived from the reported theoretical method – this parameter can be usefully used in connection with the transcription efficiency; (c) the higher flexibility of *rbcS-3A* with respect to *rbcS-E9* regulatory upstream regions, suggests an important role of DNA superstructural features on the efficiency of the transcription process of these two genes [21–23].

Table I

Theoretical distortion energies and cyclization rate constants

	410-3A	410-E9
$\Delta E/\Delta E°$	0.77	0.90
ΔE (kcal/mol)	4.2	4.8
k_1 ($\times 10^{-3}$ s^{-1})	1.78[a]	0.51[a]
	2.91[b]	0.84[b]
	3.56[c]	1.00[c]
J_{3A}/J_{E9}		3.49[a]
		3.46[b]
		3.56[c]

$\Delta E/\Delta E°$ represents the theoretical distortion energies with respect to a straight DNA 410 bp long; ΔE, energy absolute values, derived from DNA persistence length and torsion module; k_1, cyclization rate constants at various ligase concentration (U/ml, Weiss units: [a]0.083; [b]0.125; [c]0.15); J_{3A}/J_{E9}, ratio between the J factors of 410-3A and 410-E9 at different ligase concentrations.

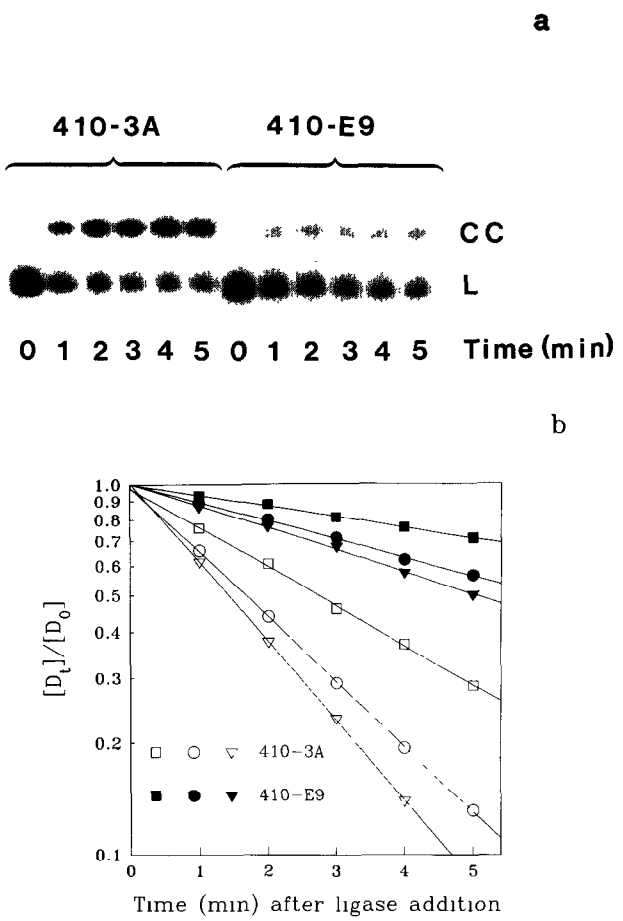

Fig. 4. (a) Autoradiograph showing progressive ligation of the linear 410 bp DNAs (L) to form covalently closed circular DNA (CC). (b) Cyclization kinetics of 410-3A and 410-E9 at various ligase concentrations. Data are plotted as logarithm of fraction of remaining reactant DNA at time t ($[D_0]/[D_t]$) as a function of time. The symbols indicate experiments at different ligase concentrations (Weiss units): squares, 0.083 U/ml; circles, 0.125 U/ml; triangles, 0.15 U/ml.

Acknowledgements: We wish to thank Dr. Nam-Hai Chua and Dr. Chris Kuhlemeier for the clones pUC18/-3A/E9 and pBR325-E9. The financial support of Comitato nazionale per le Biotecnologie e la Biologia Molecolare and of Istituto Pasteur – Fondazione Cenci Bolognetti is gratefully acknowledged.

REFERENCES

[1] Ulanovsky, L.E. and Trifonov, E.N. (1987) Estimation of wedge components in curved DNA. Nature 326, 720–722.
[2] De Santis, P., Morosetti, S., Palleschi, A., Savino, M. and Scipioni, A. (1988) in: Biological and Artificial Intelligence Systems (Clementi, E. and Chin, S. eds.) pp. 143–161, Escom Science Publ., Leiden.
[3] De Santis, P., Palleschi, A., Savino, M. and Scipioni, A. (1988) Biophys. Chem. 32, 305–317.
[4] Cacchione, S., De Santis, P., Foti, D., Palleschi, A. and Savino, M. (1989) Biochemistry 28, 8706–8713.

[5] Calladine, C.R., Drew, H.R. and McCall, M.J. (1988) J. Mol. Biol. 201, 127–137.
[6] Kuhlemeier, C., Fluhr, R. and Chua, N.-H. (1988) Mol. Gen. Genet. 212, 405–411.
[7] Kuhlemeier, C., Cuozzo, M., Green, P.J., Goyvaerts, E., Ward, K. and Chua, N.-H. (1988) Proc. Natl. Acad. Sci. USA 85, 4662–4666.
[8] Green, P.J., Yong, M.H., Cuozzo, M., Kano-Murakami, Y., Silverstein, P. and Chua, N.-H. (1988) EMBO J. 7, 4035–4044.
[9] Cacchione, S., Savino, M. and Tecce, G. (1989) Biophys. Chem. 33, 217–226.
[10] Cacchione, S., Savino, M. and Tufillaro, A. (1991) FEBS Lett. 289, 244–248.
[11] Bordin, F., Cacchione, S., Savino, M. and Tufillaro, A. (1992) Biophys. Chem. 44, 92–112.
[12] Fluhr, R., Moses, P., Morelli, G. and Chua, N.-H. (1986) EMBO J. 5, 2063–2071.
[13] Coruzzi, G., Broglie, R., Edwards, C. and Chua, N.-H. (1984) EMBO J. 3, 1671–1679.
[14] Boffelli, D., De Santis, P., Palleschi, A., Savino, M. and Scipioni, A. (1992) Int. J. Quantum Chem. 42, 1409–1426.
[15] De Santis, P., Fuà, M., Palleschi, A. and Savino, M. (1993) Biophys. Chem. 46, 193–204.
[16] Olson, W.K., Marky, N.L., Jernigan, R.L. and Zhurkin, V.G. (1993) J. Mol. Biol. 232, 530–554.
[17] Fluhr, R., Kuhlemeier, C., Nagy, F. and Chua, N.-H. (1986) Science 232, 1106–1112.
[18] Crothers, D.M., Drak, J., Kahn, J.D. and Levene, S.D. (1992) Methods Enzymol. 212, 3–29.
[19] Shore, D., Langowski, J. and Baldwin, R.L. (1981) Proc. Natl. Acad. Sci. USA 78, 4833–4837.
[20] Shore, D. and Baldwin, R.L. (1983) J. Mol. Biol. 170, 957–981.
[21] Collis, C.M., Molloy, P.L., Both, G.W. and Drew, H.R. (1989) Nucleic Acids Res. 17, 9447–9468.
[22] Gartenberg, M.R. and Crothers, D.M. (1991) J. Mol. Biol. 219, 217–230.
[23] Lavigne, M., Herbert, M., Kolb, A. and Buc, H. (1992) J. Mol. Biol. 224, 293–306.