Biophysical Chemistry 42 (1992) 147–152
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01629

Theoretical prediction of the gel electrophoretic retardation changes due to point mutations in a tract of Sv40 DNA

P. De Santis [a], A. Palleschi [a], M. Savino [b] and A. Scipioni [a]

[a] *Dipartimento di Chimica, and* [b] *Dipartimento di Genetica e Biologia Molecolare, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome (Italy)*

(Received 23 May 1991; accepted 16 July 1991)

Abstract

The changes of gel electrophoretic retardation due to single base substitutions in a 173 bp fragment of Sv40 DNA were predicted by using a theoretical model based on conformational energy calculations. As described in previous papers, this model allows successful prediction of the gel electrophoretic retardation of synthetic as well as natural DNAs reported in literature. The experimental retardations related to 195 point-mutated DNAs were reproduced with a standard deviation of 0.05 comparable with the experimental one of 0.04. This result, which represents a very critical test for the proposed model, indicates that DNA superstructures can be satisfactorily predicted on the simple physical basis of the integration of the nearest-neighbour perturbations in the dinucleotide steps. Thus, cooperative effects appear, in the majority of cases investigated, to play a second order role.

Keywords: Gel electrophoretic retardation; Sv40 DNA; DNA superstructure prediction model

1. Introduction

Since the first hypothesis on the possible existence of an anisotropic sequence dependent bendability of DNA [1] and the initial detection of an actual intrinsic DNA bend [2], curved DNA fragments have been identified in a variety of DNA sequences. The possible biological implications of the virtual and actual DNA curvatures are currently examined in different laboratories; from these investigations the concept that the intrinsic superstructural features of DNA have biological implications appears at present to be widely accepted.

In contrast, the physical origin of the DNA curvature is currently subject still open for debate. Basically, two classes of models have been advanced: the first class derives from Trifonov's original proposal of the "wedge model" [1] which represents the first nearest-neighbour model; the other one, firstly advanced by Crothers et al. [3], identifies the origin of the curvature at the con-

Correspondence to: Dr. P. De Santis, Dipartimento di Chimica, and [b] Dipartimento di Genetica e Biologia Molecolare, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome (Italy).

formational transition between the canonical B-DNA and a non-standard structure characterizing the tracts of repeated A-residues.

In fact, the majority of curved regions of DNA contains A-tracts (or the complementary T-tracts) helically spaced, providing additive effects on the helical axis. Further, A-tracts characterize also the inner regions of induced curvatures in DNAs by cyclization as well as by nucleosome formation [4].

Different authors provided arguments favouring one or the other of the models [1,3,5–15]. A few years ago [16], with the aim to clarify such a problem using the methods of the theoretical conformational analysis, we first introduced a model for predicting the structures of linear macromolecules [17–22], and later a theoretical model for the sequence dependent curvature of DNA. The model is based on integration along the sequence of the deviations from the canonical B-DNA structure relative to the 16 different dinucleotide steps as obtained by conformational energy calculations. The model was subsequently refined [23–26] in order to improve the correlation with the experimental data.

We showed that the smallness of the local deviations allows a useful reduction of the representation of the DNA superstructure in terms of curvature diagram, namely of the complex funtion $\bar{C}(n)$ which quantifies the local deviations of the helical axis (in modulus and phase) after a turn of the double helix along the sequence:

$$\bar{C}(n) = \nu/g \sum_{s=g/2}^{n+g/2} d_s \, e^{2\pi i s/\nu} \qquad (1)$$

where ν is the average periodicity of B-DNA and $d_s = \rho_s - i\tau_s$, the deviation from the B-DNA in terms of the roll (ρ) and tilt (τ) angles at the dinucleotide step, s, along the sequence. The integration grid, usually a tract of 21 or 31 bp corresponding to 2 or 3 turns of double helix, is represented by g. A similar definition of the curvature, as a graphic vectorial summation of tilt and roll angles, was later proposed by Ulanovsky and Trifonov [11] on an empirical basis.

The deviations d (ρ, $-\tau$) for the 16 possible dinucleotide steps are reported below as a Hermitian matrix. Such a form easily assures that the same curvature is obtained when using the sequence of strand I or II, alternately:

d (deg)	A	T	G	C
T	(8.0, 0.0)	(−5.4, −0.5)	(6.8, 0.4)	(2.0, −1.7)
A	(−5.4, 0.5)	(−7.3, 0.0)	(−1.0, 1.6)	(−2.5, 2.7)
C	(6.8, −0.4)	(−1.0, −1.6)	(4.6, 0.0)	(1.3, −0.6)
G	(2.0, 1.7)	(−2.5, −2.7)	(1.3, 0.6)	(−3.7, 0.0)

From the curvature diagrams it was easy to obtain the superstructure of tracts of DNA in good agreement with experiment [16,23–27].

2. Electrophoretic retardation

As a result of these investigations, we found that the gel electrophoretic retardation, which is the most sensitive technique for detecting the presence of bends in DNA, is related to the central dispersion of the integral curvature, σ^2, namely, to the angular dispersion of the local helical axes (practically, of the base pair orientations):

$$\sigma^2 = \langle C(n) \, | \, C(n) \rangle - | \langle C(n) \rangle |^2 \qquad (2)$$

where

$$C(n) = \sum_{s=1}^{n} d_s \, e^{2\pi i s/\nu} \qquad (3)$$

represents the integral curvature; practically, the angle between the first and the nth base pair planes.

The curvature dispersion, σ^2, shows good correlation (correlation coefficient 0.98) with gel electrophoretic retardation for a large collection of biosynthetic multimeric oligonucleotides with different sequences, molecular weights (50–200 bp) and periodicities [24,26,27]. Furthermore, it reproduces very satisfactorily all the experimental data of permutation gel electrophoresis which is the favoured technique currently adopted for localizing the position of the principal bend region in DNA tracts [28].

In fact, current theories on the gel electrophoretic behaviour of linear DNA are based on the reptation model firstly advanced by De

Gennes [29]. Such a model was recently adopted by Levene and Zimm [30], for understanding the anomalous electrophoretic behaviour of bent DNA molecules. It does not, however, appear to be of an easy usage and requires large computers for time simulations of the DNA motions and a priori knowledge of the gel elasticity as well as of the DNA bendability.

Our model is, on the contrary, formally and physically very simple; it is based on the thermodynamic theory of reaction kinetics. The mobility of a DNA chain under an electric field in a tight polyacrylamide gel matrix is controlled by the overcoming of potential energy barriers caused by stochastic deformations of the DNA chain as well as of the gel fibers. Thus, the mobility ratio between a curved and an equivalent straight-chain DNA tract, can be expected to be proportional to $\exp(-\Delta E^{+}/RT)$, where ΔE^{+} is the difference in average deformation energy between curved and straight chain-DNA required for the electrophoretic transport. We adopted for such a difference a quantity proportional to the straightening energy of the curved DNA which can be represented by the dispersion of the curvature σ^2 [26]. Since the relative mobility of straight-chain DNA with respect to its equivalent curved tract can be, to close approximation, represented by the relative molecular weight, R:

$$\log R = k\sigma^2 \tag{4}$$

with k depending on the gel characteristics and temperature.

3. Application to point mutations

Recently, Milton et al. [31], following on their previous paper [32], mutagenized a 60 bp region of a Sv40 tract (1836–2008) in order to obtain all the possible base pair substitutions, and analyzed the changes of the gel electrophoretic mobility resulting from each single mutation. They concluded that "base steps other than ApA are involved in sequence directed DNA bends and that the changes observed due to single base pair substitutions are inexplicable even by the most recent models".

We have calculated the values of σ^2 of all the point mutations relative to the Sv40 tract investigated by Milton et al. [31] and obtained the diagrams illustrated in Fig. 1 following the same drawing style the cited authors used for their experimental data; these are shown, for direct comparison, in the same figure. The degree of correlation is very satisfactory if one takes into account that our model predicts the gel electrophoretic retardation changes induced in a tract of 173 bp by a single base pair substitution. The standard deviation is 0.06 against an experimental standard deviation of 0.04 [31].

We tried to improve the agreement between theoretical and experimental data using a steepest descent optimization but allowing the roll and tilt angles to change in the limited range of ± 1 degree and relaxing the assumed constant DNA periodicity, $\nu = 10.3$. Such changes are fully compatible with the expected accuracy of the conformational calculations. Thus, we obtained a better correlation between experimental and theoretical retardation with a standard deviation 0.05, only 0.01 greater than the experimental standard deviation as estimated by Milton et al. [31]. This slight difference contains the non nearest-neighbour effects as well as those deriving from the more complex nature of the gel electrophoresis as compared with the adopted simple dependence on the dispersion of the curvature, σ^2. Given the complexity of the system, this result probably represents the best one obtainable by a physical model.

Figure 2 shows the correlation between the original roll and tilt angles and their values after the optimization process (standard deviation 0.5°; correlation coefficient 0.98). It should be noted that the average DNA periodicity decreases from the initial value of 10.3 to 10.2 in account of the relatively high AT content (68%). The correlation diagram between the experimental and theoretical retardation factors is illustrated in Fig. 3 for all the 184 point mutations; in addition, also the five double and triple base substitutions investigated by Milton et al. [31] are represented as well as seven supplementary single substitutions (out

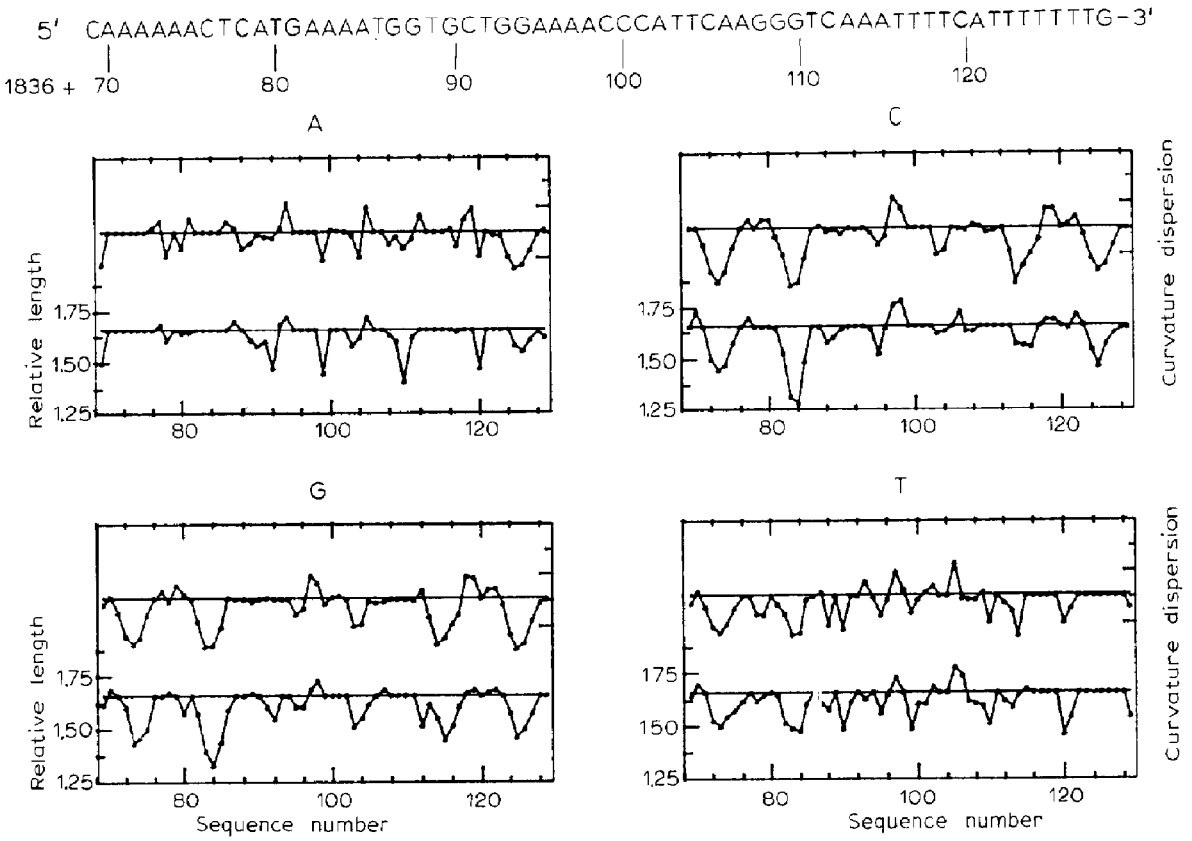

Fig. 1. Comparison between the central dispersion of curvature (σ^2) and the relative length measurements of the single base-pair substitutions within the 60 bp curved region of the Sv40 tract (1836–2008) reported. Each diagram shows the theoretical and the observed data when an adenine (A), cytosine (C), guanine (G) or thymine (T) residue is substituted at each residue position. The experimental data were obtained from the original paper of Milton et al. [31].

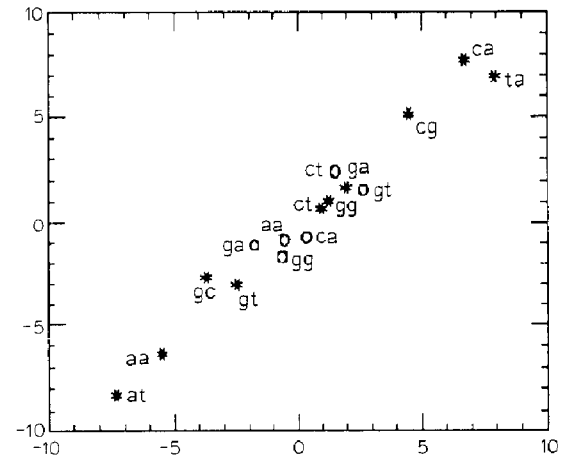

Fig. 2. Correlation diagram between the original roll (∗) and tilt (○) angles and their values after the optimization process (standard deviation 0.5°)

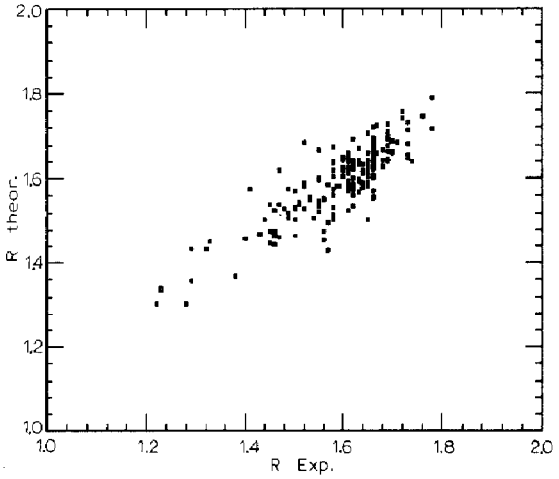

Fig. 3. Correlation diagram between the experimental and theoretical retardations corresponding to the Milton et al. data [31,32].

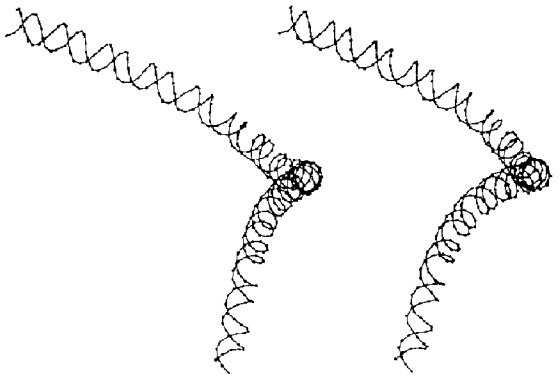

Fig. 4. Stereoprojection of the theoretical superstructure of the Sv40 tract (1863–2008) as fitted along the phosphodiester chains. The base pair 84 is on the plane of figure with its minor groove facing to the curvature.

of the central 60 bp fragment) reported by the same authors in their previous paper [32]. Whilst it is difficult to give detailed analysis of each substitution effect on the electrophoretic mobility (which depends on both phase and position of the base mutation, as well as on the overall superstructure of the DNA tract—see Fig. 4), nevertheless, some basic considerations in terms of curvature can be made.

The curvature diagram, as shown in Fig. 5, indicates that the most relevant decreases in retardation, resulting from substitutions are localized in regions of high local curvature; whereas, the central tract, which is characterized by a lower curvature, gives both increases and decreases in retardation. Further, the tract about the sequence number 70–90 is the most sensitive to substitutions because it is localized in the center of the whole DNA fragment; the changes of the local curvature in this region therefore yield the most pronounced effects in gel electrophoresis. In fact, the permutation gel assay is based on this effect [28].

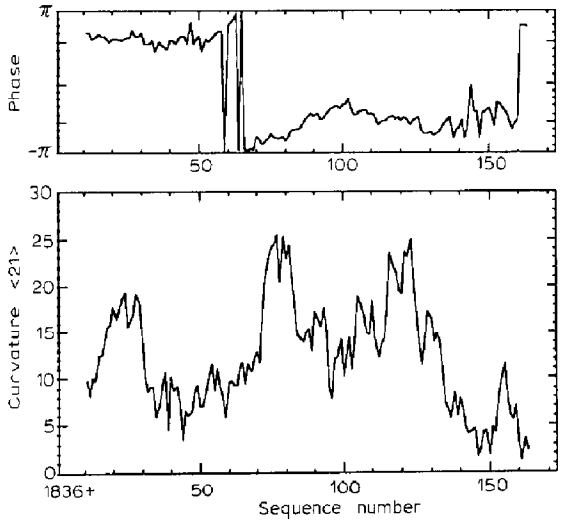

Fig. 5. Curvature diagram (modulus and phase) of the Sv40 tract (bp 1836–2008; integration grid g = 21 bp).

The distribution of the deviations between theoretical and experimental retardation, is practically Gaussian and possibly reflects the distribution of random experimental errors: the standard deviations are in fact, 0.05 and 0.04, respectively. Finally, the majority of the largest deviations from the experimental data is coincident with the largest decreases in retardation and plausibly contains the effects of the non-linearity of the retardation against the curvature dispersion. On the other hand, it could also be the result of cooperative contributions on the curvature due to the repeating AA or TT base steps.

4. Conclusions

Whilst based on the simple nearest-neighbour interactions hypothesis, the proposed theoretical model is capable to translate the deterministic fluctuations of base sequence in pieces of information on DNA superstructures and on their electrophoretic manifestations in surprisingly good agreement with experiment. This is clearly illustrated by the experimental data presented in this paper, which appear to be very critical as a validity test for the proposed theoretical model. Such a general pattern of agreement between theoretical and experimental data convincingly proves the dominant role of nearest-neighbour interactions at the origin of DNA curvature and allows to adopt the theoretical calculations with a large degree of confidence as a valid alternative to the experiments for the detection and localization of superstructural features of DNA. A similar conclusion was very recently reached by Dalma-Weiszhausz et al. [33] about the induced curvature in DNAs by the CAP protein.

Finally, it is worthy of discussion that the large majority of the point mutations decreases the curvature of this biologically relevant DNA tract, suggesting that the curvature could be evolutionary selected.

Acknowledgement

This work was supported by CNR under the project "Chimica Fine II".

References

1 E.N. Trifonov, Nucl. Acids Res. 8 (1980) 4041–4053.
2 J.C. Marini, S.D. Levene, D.M. Crothers and P.T. Englund, Proc. Natl. Acad. Sci. USA 79 (1984) 7664–7668.
3 H.-S. Koo, H.-M. Wu and D.M. Crothers, Nature 320 (1986) 501–506.
4 H.R. Drew and A.A. Travers, J. Mol. Biol. 186 (1985) 773–790.
5 V.B. Zhurkin, M.P. Lysov and V.J. Ivanov, Nucl. Acids Res. 6 (1979) 1081–1090.
6 W. Ross and A. Landy, J. Mol. Biol. 156 (1982) 523–530.
7 P.J. Hagerman, Proc. Natl. Acad. Sci. USA 81 (1984) 4632–4636.
8 N.B. Ulyanov and V.B. Zhurkin, J. Biomol. Struct. Dynam. 2 (1984) 361–369.
9 C.-S. Tung and S.C. Harvey, Nucl. Acids Res. 14 (1986) 381–387.
10 S. Dieckmann, EMBO J. 6 (1987) 4213–4217.
11 L. Ulanovsky and E.N. Trifonov, Nature 326 (1987) 720–722.
12 A.M. Burkhoff and T.D. Tullius, Nature 331 (1988) 455–457.
13 S. Dieckmann and E. von Kitzing, in: DNA bending and curvature, eds. W.K. Olson, M.H. Sarma and M. Sundaralingam, Structure and Expression, Vol. 3. (Adenine Press, Guiderland, NY, 1988) pp. 57–67.
14 T.D. Tullius and A.M. Burkhoff, in: DNA Bending and curvature, eds. W.K. Olson, M.H. Sarma and M. Sundaralingam, Structure and Expression, Vol. 3: (Adenine Press, Guiderland, NY, 1988).
15 T.E. Haran and D.M. Crothers, Biochemistry 28 (1989) 2763–2767.
16 P. De Santis, S. Morosetti, A. Palleschi and M. Savino, in: eds. E. Clementi and S. Chin, Structures and Dynamics of Nucleic Acids, Proteins and Membranes (Plenum Publishing, New York, NY, 1986) pp. 31–49.
17 P. De Santis, E. Giglio, A.M. Liquori and A. Ripamonti, Nuovo Cimento 26 (1962) 616–618.
18 P. De Santis, P. Giglio, A.M. Liquori and A. Ripamonti, J. Polym. Sci. *Part A*, (1963) 1383–1404.
19 P. De Santis, P. Giglio, A.M. Liquori and A. Ripamonti, Nature 206 (1965) 456–458.
20 P. De Santis, R. Rizzo and G. Ughetto, Nature 237 (1972) 94–95.
21 P. De Santis, S. Morosetti, A. Palleschi and M. Savino, Biopolymers 20 (1981) 1727–1739.
22 F.G. Calascibetta, M. Dentini, P. De Santis and S. Morosetti, Biopolymers 14 (1975) 1667–1684.
23 P. De Santis, S. Morosetti, A. Palleschi, M. Savino and A. Scipioni, in: Biological and artificial systems, eds. E. Clementi and S. Chin (ESCOM Science Publ., New York, NY, 1988) pp. 143–161.
24 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biophys. Chem. 32 (1988) 305–317.
25 P. De Santis, G. Gallo, A. Palleschi, M. Savino and A. Scipioni, J. Mol. Liquids 41 (1989) 291–304.
26 P. De Santis, A. Palleschi, M. Savino and A. Scipioni, Biochemistry 29 (39) (1990) 9269–9273.
27 S. Cacchione, P. De Santis, D. Foti, A. Palleschi and M. Savino, Biochemistry 28 (1989) 8706–8713.
28 H.M. Wu and D.M. Crothers, Nature 308 (1984) 509–512.
29 P.G. de Gennes, J. Chem. Phys. 55 (1971) 572–579.
30 S.D. Levene and B.H. Zimm, Science 245 (1989) 396–399.
31 D.L. Milton, M.L. Casper and R.F. Gesteland, J. Mol. Biol. 213 (1990) 135–140.
32 D.L. Milton and R.F. Gesteland, Nucl. Acids Res. 16 (1988) 3931–3949.
33 D.D. Dalma-Weiszhausz, M.R. Gartenberg and D.M. Crothers, Nucl. Acids Res. 19 (1991) 611–616.