FEBS 18381

FEBS Letters 406 (1997) 69–74

Distribution of sequence-dependent curvature in genomic DNA sequences

Andrei Gabrielian[1], Kristian Vlahovicek, Sándor Pongor*

International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, Padriciano 99, 34012 Trieste, Italy

Abstract The distribution of inherent, sequence-dependent curvature was calculated for a number of prokaryotic (*M. genitalium, H. influenzae, M. jannaschii*), viral (adenovirus 2, equine herpes virus 1), phage (M13, lambda), eukaryotic (*S. cerevisiae*) and mitochondrial genomes as well as *E. coli* and human genomic fragments. The genomic averages are in the range of 6–8°/helical turn and only about 20% of DNA is curved less than 3°/helical turn. The prokaryotes and phages appear to have a consistently higher frequency of curved DNA in their genomes than the other genomes tested. Long, highly curved segments, similar to artificially designed curved DNA, are apparently absent from the genomes. Short, curved segments, differing in G+C content may provide environmentally modulated conformational signals for gene regulation. A WWW-server was constructed for the prediction of curved sites from DNA sequences ⟨http://icgeb.trieste.it/dna/curve_it.html/⟩..

© 1997 Federation of European Biochemical Societies.

Key words: DNA structure; DNA curvature; Genome analysis

1. Introduction

The fact that DNA may contain local, sequence-dependent conformations has profoundly influenced the thinking of biologists in recent years. DNA is no longer considered to be an idealized, featureless double helix but rather a series of local conformations that may add up to macroscopic curvature even without external factors such as DNA-binding proteins [1]. Inherent bending of DNA is usually associated with periodically repeating sequence motifs such as A-tracts [2] or GC-type elements [3]. In this respect, it is customary to speak about A-type and non-A-type DNA curvature that differ in their sensitivity to metal ions [3] and in the direction of bending [4]. Given the differential response of curved motifs to environmental factors, it is plausible to speculate that these motifs may act as 'environmental sensors' whose conformational transitions act as regulatory signals. The aim of the present work is to obtain preliminary information on how inherently curved DNA segments are distributed in genomic sequences and to see if the currently available complete genomes differ in this respect.

Prediction of DNA curvature is based on the geometry of the individual dinucleotide steps (roll, twist and tilt angles). There are several methods for the calculation of the DNA trajectory that may give different results, especially for longer DNA segments [5–9]. However, for shorter segments, like those involved in local bending, essentially all of these are suitable, so we chose a simple and visually meaningful index, the angle of basepairs one helical turn apart (schematically shown on the model of the d(CGCGAATTCGCG) double helix [10], Fig. 2). This value is expected to be zero for ideally straight B-DNA and positive for curved segments.

Here we show, through the analysis of complete genomes, that prokaryotic genomes (*H. influenzae*, *M. jannaschii*, *M. genitalium*) appear to contain more curved segments than the human, yeast and viral genomic DNA tested. Long, uniformly curved segments, such as can be easily produced with artificially designed repetitive sequence motifs, seem to be absent in genomic DNA. However, shorter, curved segments differing in G+C contents are ubiquitous.

2. Methods

There are a number of published geometric parameter sets determined on the basis of X-ray data [11], theoretical geometry calculations [7], gel mobility analysis [12] as well as NMR measurements [13] of which we used the latter. The magnitude of DNA curvature was calculated with the BEND algorithm of Goodsell and Dickerson [5,14]. In this algorithm, the normal vectors of successive basepairs are added up vectorially and the angle between two averaged normal vectors 31 basepairs (~3 helical turns) apart is used as the indicator of curvature. We presented the values of the curvature as the deflection angle per 10.5 residue helical turn (1°/bp=10.5°/helical turn). The genomic sequences were taken from the sequence databases and from the literature, as indicated. The distribution of curvature is presented as histograms with a bin-width of 1.5°/helical turn, i.e. the percentage of segments with a curvature between 0 and 1.5°/helical turn is plotted at 0°/helical turn, etc. The average distributions (Fig. 2) are arithmetic averages of the groups presented in Fig. 1A-D, respectively.

3. Results and discussion

Even though curvature values determined by gel mobility analysis strongly depend on the experimental conditions [15] and on the length of the tested molecule [16], there are a number of sequence motifs whose curvature characteristics were established by several research groups, and we used these to test the chosen method of calculation. The results in Table 1 show that there is a good over-all agreement between the curvature index (angular deflection per helical turn) with the experimental results. The calculated values for curved DNA is above 9°/helical turn, the straight sequence motives give values below 4–5°/helical turn. There are exceptions to this rule (i.e. wrong predictions); e.g. the $(tctctaaaaaatatataaaaa)_n$ motif, which is highly curved according to electrophoretic mobility and circulatization kinetics experiments [17], gives a conspicuously low value. Therefore this calculation — like all predictions — has to be considered approximate, so absolute threshold values may not be defined. Nevertheless, the correlation with experimental curvature values in these and in previously reported examples [14] is satisfactory. Roughly

*Corresponding author.
E-mail: pongor@icgeb.trieste.it

[1]*Permanent address:* Institute of Molecular Biology, Russian Academy of Sciences, Vavilov St. 32, 117984 Moscow, Russia.

0014-5793/97/$17.00 © 1997 Federation of European Biochemical Societies. All rights reserved.
PII S0014-5793(97)00236-6