The global intrinsic curvature of archaeal and eubacterial genomes is mostly contained in their dinucleotide composition and is probably not an adaptation

Enrique Merino and Alejandro Garcia-Rubio*

Institute for Biotechnology, National University of Mexico, Cuernavaca, Apelo 510-3, Morelos 62250, Mexico

Received January 5, 2000; Revised March 29, 2000; Accepted April 24, 2000

ABSTRACT

Until now, the genomic DNA of all eubacteria analyzed has been hyper-curved, its global intrinsic curvature being higher than that of a random sequence. In contrast, that rule failed for archaea or eukaryotes, which could be either hypo- or hyper-curved. The existence of the rule suggested that, at least for eubacteria, global intrinsic curvature is adaptive. However, the present results from analyzing 21 eubacterial and six archaeal genomes argue against adaptation. First, there are two eubacterial exceptions to the former rule. More significantly, we found that the dinucleotide composition of the genome alone (which lacks all sequence information) is enough to determine the genome curvature. Additional evidence against adaptation came from showing that the global curvature of bacterial genomes could not have evolved under either of two complementary models of curvature selection: (i) that curvature is selected locally from unbiased variability; (ii) that curvature is established globally through the selection of a curvature-altering mutational bias. We found that the observed relationship between curvature and dinucleotide composition is incompatible with model (i). We also found that, contrary to the predictions of model (ii), the dinucleotide compositions of bacterial genomes were not statistically special in their curvature-related properties (when compared to stochastically generated dinucleotide compositions).

INTRODUCTION

In its minimum energy state DNA is slightly curved. This intrinsic axial curvature is sequence dependent, so at each position along a DNA molecule there is a characteristic degree of curvature. On average the curvature of DNA is 3.5 degrees per helix turn (d.p.t.). In many cases local variations in intrinsic curvature have great biological meaning. Different studies have demonstrated the importance of DNA local curvature in transcription (1–4), replication (5–9), recombination (10–12) and chromatin structure (13–16).

Extensive experimental and theoretical studies have led to algorithms that predict the intrinsic curvature of a DNA segment from its nucleotide sequence (17–21). These methods and the availability of extensive genome sequences have made possible the study of intrinsic curvature as a global property of the genome of the species. This is done by computing the intrinsic curvature at each point of a genome and then taking some statistical measure of those values, such as their average or their frequency distribution, in order to represent the whole genome.

Recent works by Gabrielian et al. (22) and our group (23), pioneering this approach, found that the curvature distribution and average curvature of real genomes differ significantly from those of a random sequence. More importantly, these works showed that these measures vary significantly from one species to another. The most intriguing observation made at that time was that eubacterial genomes were all hyper-curved (their average curvatures being >3.5 d.p.t., which is the curvature of a random sequence), while archaeal and eukaryotic genomes could be either hypo- or hyper-curved. This observation suggested that the DNA of eubacteria could have distinct structural requirements, perhaps related to the fact that in this group, as opposed to the other two, DNA packaging does not rely on histones.

Why do genomes have the particular curvatures they have? This is an essential question. But, before trying to find a functional, adaptive explanation for curvature per se, it is reasonable to try to exclude the possibility that the curvature of a genome is a secondary consequence of some of the many requirements that make of a genome a non-random sequence. In a previous work (23) we tested whether the curvature of a genome arises from its nucleotide composition (the A+T:G+C ratio). We found that after randomly permutating a genome (which maintains the nucleotide composition) the resulting genome was closer in curvature to a random sequence (whose A+T:G+C ratio equals 1) than to the original genome. This result established that the species-specific curvature of a genome is not a trivial consequence of its mononucleotide composition.

In this work we have tested whether dinucleotide composition dictates the curvature of the genome. It is well known that DNA curvature is primarily due to base-stacking effects
between nucleotides in consecutive steps of the DNA ladder. Besides that, we had three other reasons for focusing on the dinucleotide composition. The first was that since the nucleotide composition cannot explain the curvature of a genome, the dinucleotide composition is the next simplest hypothesis. The second, that the dinucleotide composition is an important structural characteristic of a genome (24,25). So much so, that a closely related measure, the dinucleotide relative abundance, has been called ‘the genome’s signature’ (26). The dinucleotide relative abundance remains approximately uniform along the genome (when measured over spans >50 000 bp long), but varies significantly from one species to another (25,27). Furthermore, the relative abundances of trinucleotides or longer polynucleotides appear to result from the relative dinucleotide abundances (24). The third reason to focus on the dinucleotide distribution of a genome was that, unlike mononucleotides, the probabilities of any two dinucleotides in a short DNA segment are not independent, because any nucleotide in a sequence is, at the same time, the second nucleotide of a dinucleotide and the first nucleotide of the next dinucleotide. This fact creates a Markov chain of conditional probabilities. Although the signal fades out rapidly, this property could be important for DNA curvature which, due to the flexibility of DNA, is only meaningful over relatively short segments.

In theory, if the curvature of a genome has been selected, this could have happened at two different levels. On the one hand, selection could define the curvature of a genome through the accumulation of individual mutations produced without any bias [model (i)]. This ‘standard’ selection mechanism can target specific DNA regions. Localized regions of intrinsic curvature (i.e. those found upstream of particular promoters) must be due to this. On the other hand, selection could define curvature by selecting those mechanisms producing mutational bias that alter DNA composition in such a way that DNA tends to adopt an adaptively favorable curvature [model (ii)]. This ‘more speculative’ mechanism could be more suitable to explain global, widespread curvature. Here we have examined whether bacterial genomes show signs of adaptation under either of these two models and, more generally, we have investigated whether genome global curvature is adaptive.

**MATERIALS AND METHODS**

**Database sequences**

All nucleotide sequences were obtained from GenBank (http://www.ncbi.nlm.nih.gov/GenBank). These included the complete genomes of six Archaeobacteria and 21 Eubacteria: Archaeobacteria: *Aquifex pyrophilus* (ApPy); *Archaeoglobus fulgidus* (AE000782); *Methanobacterium thermautotrophicum* (AE000666); *Methanococcus jannaschii* (L77117); *Pyrococcus abyssi* (AL096836); *Pyrococcus horikoshii* (PyroH). Eubacteria: *Aquifex aeolicus* (AE000657); *Bacillus subtillis* (AL009126); *Borrelia burgdorferi* (AE000783); *Campylobacter jejuni* (AL111618); *Chlamydia pneumoniae* (AE001363); *Chlamydia trachomatis* (AE001273); *Chlamydophila pneumoniae* (AE002161); *Escherichia coli* (U00096); *Haemophilus influenzae* Rd (L42023); *Helicobacter pylori* (AE000511); *H pylori* J99 (AE001439); *Mycobacterium tuberculosis* (AL123456); *Mycoplasma genitalium* (L43967); *Mycoplasma pneumoniae* (U00089); *Neisseria meningitidis* (AE002098); *Rickettsia prowazekii* (AF235269); *Synechocystis sp.* PCC6803 (AB001339); *Thermotoga maritima* (AE000512); *Treponema pallidum* (AE000520); *Ureaplasma urealyeticum* (AF222894). As a control DNA with selected curvature we used the minicircle DNA of the *Leishmania tarentolae* kinetoplast (S67135).

**Curvature prediction**

To evaluate the theoretical DNA curvature of sequences from whole genomes we used the nearest-neighbor algorithm of Goodsell and Dickerson (28). First, the algorithm computes the DNA curvature as the successive accumulation of rotational and spatial displacement. It then evaluates the normal vector of each base pair and averages these values over a 10 bp interval. Finally, curvature is calculated as the angle between averaged normal vectors 31 bp apart. The reported value of curvature is the deviation angle per 10.5 nt, expressed in d.p.t. In some cases local curvature values at every nucleotide were grouped into histogram profiles in order to analyze the distribution of these values. The contribution matrix we used was the trimer matrix of the nucleosome position model of Satchwell et al. (14). This combination of algorithm and matrix was chosen because in a comparative study it was the most accurate for predicting the curvature of well-characterized curved DNA sequences (28). However, have also used the model of Calladine et al. (29) and corroborated that the method does not affect our conclusions.

**Nucleotide compositions and construction of Markovian genomes**

Virtual genomes with predefined mono-, di- or trinucleotide compositions are said to be Markovian genomes of class Mkv-1, Mkv-2 or Mkv-3, respectively. Although the following description refers to Mkv-2 genomes, the other Markovian genomes were made similarly. The dinucleotide composition of a genome contains the frequency of each of the 16 dinucleotides on both DNA strands, divided by twice the length of the sequence. The balance index of the composition is defined as 16 times the geometric mean of the probabilities of the 16 dinucleotides. Mkv-2 genomes are Markovian chains grown by successively adding at position $n + 1$ a new terminal nucleotide, A, C, G or T, chosen according to the transition probabilities $X \rightarrow A$, $X \rightarrow C$, $X \rightarrow G$ or $X \rightarrow T$, where $X$ represents the nucleotide at position $n$. These transition probabilities are equal to the frequencies of $XA$, $XC$, $XG$ or $XT$, where $X$ represents the nucleotide at position $n$. The transition probabilities are equal to the frequencies of $XA$, $XC$, $XG$ or $XT$ in the dinucleotide composition divided by the sum of the four frequencies. The size of all Markovian genomes was $5 \times 10^5$ nt. At this size, any two Markovian genomes based on the same composition are indistinguishable in average curvature and curvature distribution, so no duplicates were made.

In some experiments Markovian genomes were based on stochastically generated dinucleotide compositions (aleatoric genomes). The generation of these dinucleotide compositions was as follows. We chose at random the 16 needed transition probabilities, built a 500 000 nt long Mkv-2 chain, generated its complementary strand and, finally, obtained the dinucleotide composition of the two strands combined. This practical approach ensured that these dinucleotide compositions have all the symmetries that arise in natural compositions from both DNA strands having very similar compositions.
To statistically assess how rare the curvature of a bacterial Mkv-2 genome is, a distribution of average curvatures was obtained from 200 aleatoric genomes whose balance index did not differ by more than 0.025 units from the balance index of the Mkv-2 genome. The Z score probability for the curvature of the genome was calculated against that distribution. Values of Z score >1.96 (P < 0.05) were considered as an indication of not belonging to the distribution.

Evolution of artificial genomes

Evolution of random sequences for increased curvature was as follows. The process started with a 50 000 nt long random sequence. After making 12 copies of it, successive cycles of the following three steps were applied: (i) introduction of variation by random point mutations at a predefined mutation rate; (ii) selection of the six most curved variants; (iii) generation of 12 new genomes by random recombination of the six selected ones. The evolution process lasted 3700 cycles. Mutations were completely unbiased and the mutation rate per cycle was two changes per genome. To assess the adaptive process, every 100 cycles the curvature of the best genome was determined.

RESULTS

A preliminary explanation

We must first take a short detour to introduce some concepts and nomenclature used in this work.

Most arguments in this work will relate to the predicted average curvature of different genomes. All genomes belong to one of the following three classes.

1. Real genomes. These include six archaeal and 21 eubacterial genomes and the kinetoplast DNA of L.tarentolae.

2. An ‘evolved genome’. This genome is the result of an in silico evolution experiment, which will be described later.

3. Markovian genomes. These genomes are built through a Markovian process to have a predefined mono-, di- or trinucleotide composition (and are hence called Mkv-1, Mkv-2 or Mkv-3 genomes).

All the nucleotide compositions considered (upon which the Markovian genomes were built) belong, by their origin, to one of the following four categories.

1. A perfectly balanced mononucleotide composition. Markovian genomes based on it are called random genomes. Random genomes have the interesting property of also being perfectly balanced in their di-, tri- and polynucleotide compositions.

2. The mono-, di- or trinucleotide compositions of real genomes. Markovian genomes based on these compositions are referred to, for example, as ‘E.coli Mkv-1’ (meaning a Markovian genome based on the mononucleotide composition of the E.coli genome).

3. The dinucleotide composition of the ‘evolved genome’, which is only used to generate the ‘evolved Mkv-2 genome’.

4. A very large set of dinucleotide compositions generated through a stochastic process (see Materials and Methods). Markovian genomes based on such compositions will be called ‘aleatoric genomes’.

Other than having a predefined composition, Markovian genomes are as random as possible. It is an important fact that all Markovian genomes based on a given composition have precisely the same average curvature, in spite of being unrelated in sequence. For example, all random genomes have a curvature of 3.5 d.p.t. and all E.coli Mkv-2 genomes have a curvature of 3.79 d.p.t. Consequently, it is possible to assign a curvature value to a given composition (the curvature of any Markovian built on it). The reverse is not true; we cannot associate a composition to a curvature because many different compositions could lead to the same curvature.

It is worth noting that a Mkv-1 of a species is equivalent to what in a previous work we called a permuted genome of that species.

The curvature of real genomes is contained in their dinucleotide compositions

For each species in this study we built a Mkv-1, a Mkv-2 and a Mkv-3 genome. We then compared the predicted curvatures of the real and the three derived Markovian genomes. Figure 1 exemplifies the case for E.coli. As can be seen, the Mkv-2 and Mkv-3 genomes are not very different from the real genome. The similarity is not restricted to the average or the mode of the distributions, but comprises the full shape of the curvature distribution histograms. In contrast, the Mkv-1 genome deviates much more from the real genome and actually resembles more the random genome. The Mkv-2 and Mkv-3 genomes are almost indistinguishable, but Mkv-3 is minimally closer to the real genome. Trinucleotide compositions have four times more information than dinucleotide compositions; the similarity between the Mkv-2 and Mkv-3 genomes would suggest that most of the extra information is actually redundant. Due to their similarity to Mkv-2 genomes, Mkv-3 genomes will not be considered in the rest of this work.

The results obtained for all the archaeal and eubacterial species analyzed were similar to those for E.coli. The average curvatures of the real, Mkv-1 and Mkv-2 genomes are compared in Figure 2. A strong correlation between the Mkv-2 and real genomes was found ($r^2 = 0.92$). In comparison, the correlation between the Mkv-1 and real genomes was poor ($r^2 = 0.42$). Since the real bacterial genomes and their Mkv-2 genomes share no other information, these results indicate that
the dinucleotide composition of a bacterial genome contains most of the information required to specify its global intrinsic curvature. For curvature to extend over several helical turns it is necessary that DNA-bending dinucleotides are phased with helical-turn periodicity (10.5 bp). It is thus surprising that an essentially random sequence mimics the curvature of a real genome. This suggests that the curvature of bacterial genomes is mostly random. The contrast with a kinetoplast DNA (included as a control DNA that has arguably been selected for increased curvature) should be noted. In this case the average curvature of the real and the Mkv-2 genomes were very different (3.99 versus 4.89 d.p.t.; see genome 11 in Fig. 3).

The curvatures of Mkv-2 genomes resemble those of real genomes at short and long window sizes

Throughout this work curvature was assessed over a standard window of 31 bp (equal to three helical turns). An anonymous referee noted that ‘One might expect that if the curvature is random and not the result of adaptation, the curvature (in terms of d.p.t.) would decrease with increasing window size. However, if functionally important nonrandom curvature prevails the curvature values could remain approximately constant for various window sizes.’ Indeed, we believe this would be a prediction of model (i). Thus, we evaluated the curvature of all archaeal and eubacterial genomes (and their Mkv-2 genomes) using window sizes of 2, 3, 4, 5 and 10 helical turns. The results are summarized in Table 1. As predicted, the curvature per helix turn of real genomes decreases with window size, suggesting that random curvature prevails in them. Moreover, the correlation coefficient between real and Mkv-2 genomes was very high at all window sizes, even at 106 bp, indicating that Mkv-2 genomes resemble the real ones regardless of window size. In the case of extreme genomes (like H. pylori), this indicates that given an adequate dinucleotide composition, a Markovian sequence can generate long fragments which on average have a significant absolute curvature (in degrees, not d.p.t.).

Not all eubacterial genomes are hypo-curved

As a result of including many new genomes in this study, we found two exceptions to the previous conjecture that eubacterial genomes are more curved than a random genome (M. tuberculosis and the marginal case of P. pallidum, with average curvatures of 3.02 and 3.45 d.p.t., respectively; see genomes 7 and 28 in Fig. 3). The number of eubacterial species with a hyper-curved genome still heavily outweighs those with hypo-curved genomes (18 versus two, if the two H. pylori genomes are counted as one; Fig. 3, open circles). In comparison, among the Archea four genomes were hyper-curved and two were hypo-curved (Fig. 3, open triangles).

Model (i) does not predict a relationship between dinucleotide composition and curvature

Observing that the kinetoplast has probably been selected for curvature and that it is the only ‘genome’ whose curvature is not contained in its dinucleotide composition, we hypothesized that a dissociation between curvature and dinucleotide composition could be a hallmark of selection for curvature. To test the hypothesis, we created a completely random genome and evolved it for higher curvature by a process akin to a genetic algorithm. The process was chosen because it is a plausible mimic of model (i) of how natural selection could alter the curvature of a genome. The process implied successive rounds of unbiased mutation followed by selection of the most curved variants (see Materials and Methods). After 3700 generations the process reached an equilibrium which arose from two opposing forces: selection, which tries to increase curvature, and mutational pressure, which, being perfectly balanced, tries to bring curvature down to 3.5 d.p.t. (the value of a random sequence). The resulting genome (referred to as ‘the evolved genome’) had a curvature much larger than that of any of the real genomes (5.44 d.p.t.). Surprisingly, the evolved
genome contained almost identical proportions of the four nucleotides and also had very similar proportions of the 16 dinucleotides (balance index 0.956). When a Mkv-2 genome of the evolved genome was generated, its average curvature was only 3.72 d.p.t., 1.72 d.p.t. less than the curvature of the evolved genome (see genome 29 in Fig. 3). Thus, it is evident that the dinucleotide composition of the evolved genome does not contain the information necessary to define its global intrinsic curvature. This exercise shows that the fixation of point mutations by a strong selection for increased curvature does not lead to a significant change in the dinucleotide composition of the genome and that, in particular, it does not imprint in the dinucleotide composition the curvature of the genome.

The dinucleotide compositions of bacterial genomes are not special with regard to the curvature they imply

As stated above, any composition has an associated curvature. Here we tried to see if the dinucleotide compositions of bacterial genomes (eubacterial genomes in particular) are outstanding with regard to their associated curvature, because that would be a sign of curvature selection according to model (ii). The idea was to build many Markovian genomes based on a stochastically generated dinucleotide composition (aleatoric genomes) to see if the curvatures of the bacterial Mkv-2 genomes are common among them. However, since the dinucleotide composition of aleatoric genomes may be arbitrarily unbalanced, while the natural dinucleotide composition may not, we had to take into account a balance index. This index is defined as 16 times the geometric mean of the frequencies of the 16 dinucleotides (it equals 1 for a perfectly balanced composition and approaches 0 for a very unbalanced one). For this exercise we created $>6 \times 10^6$ aleatoric genomes. This number was required to populate (with at least 200 points) the range with a balance index $>0.95$, since balanced compositions are much more rare than unbalanced ones. When the curvatures of the aleatoric genomes were plotted against their balance index, a shape of this sort was expected: Markovian genomes can deviate from a curvature of 3.5 d.p.t. only by departing from perfect balance. As can be seen, all the bacterial Mkv-2 genomes (and almost all their closely matched real genomes) lie within this cone shaped area. Furthermore, when compared to aleatoric genomes of similar balance index, all but one of the bacterial Mkv-2 genomes lie within $\sigma = 1.96$, which corresponds to 95% of all points (the area enclosed within continuous lines in Fig. 3). However, since we are considering 27 Mkv-2
genomes, precisely one exception would be expected if the distribution is random. The exception is the \textit{N.meningitidis} Mkv-2, which lies slightly outside the \( \sigma = 1.96 \) region (Fig. 3, genome 22). Two other rare cases correspond to the \textit{Synechocystis} and \textit{H.pylori} Mkv-2 genomes (\( P = 0.05 \)). Since for every bacterial Mkv-2 genome there is a good chance of finding among aleatoric genomes of comparable balance index one with similar curvature, we must conclude that the dinucleotide compositions of bacterial genomes are not special with regard to the curvature they imply and, at least under these criteria, do not show signs of selection for curvature.

**DISCUSSION**

The first finding of this work was that the dinucleotide composition of bacterial genomes contains most of the information necessary to determine their curvature. As is clearly shown by the existence of exceptions (the kinetoplast and the evolved genomes), this observation is not a trivial consequence of the methodology used to predict DNA curvature. The programs rely on experimentally determined data on the angular and linear displacements between dinucleotides. Yet, it cannot be known \textit{a priori} to what extent the dinucleotide composition of a genome can determine its curvature, because, in principle, it is not only the dinucleotide composition but also the dinucleotide organization along the genome that should be considered. Important properties that determine the curvature of a genome could be lost in the overall dinucleotide composition. At the most local level, the angular and linear displacements between dinucleotides can either combine constructively or cancel each other out, depending on their order. At the global level, there could be regional variations in the dinucleotide composition of the genome. What is surprising in our finding is that the aforementioned considerations need only explain the small percentage of the difference in curvature that was not explained by the dinucleotide composition. This is even more surprising because for significant curvature to extend over long DNA stretches, curvature-inducing dinucleotides have to occur with some kind of near-helical periodicity. The dinucleotide composition accounted for 89% of the curvature difference in the archaeal genomes (with a linear correlation of 0.91) and 67% of the curvature difference in the eubacterial genomes (with a linear correlation of 0.89) (data not shown.). The worst case was \textit{H.pylori}, where only 59% of the curvature difference was explained. Of course, real genomes are not random, so the assumption that they can be modeled by Markov chains is, at best, a gross approximation.

The second finding of this work was that two eubacterial genomes are less curved than a random genome. These create two exceptions to the previous conjecture that eubacterial genomes are hyper-curved. Now we know of two hypo- and 13 hyper-curved eubacterial genomes. Thus, the original observation is not an absolute; it is only a tendency. However, this tendency is now supported by a large number of genomes and we believe that it still requires an explanation.

The third finding is that if curvature is selected according to model (i), no correlation is expected between dinucleotide composition and curvature. The \textit{in silico} evolution of an initially random genome showed that, after selection had gradually fixed thousands of mutations, the resulting evolved genome had a very high curvature, 5.45 d.p.t., while its Mkv-2 genome was very close to a random genome, with a curvature of only 3.72 d.p.t. Although quite different dinucleotide compositions may coincide in curvatures close to 3.5 d.p.t., in this case the similarity in curvature between the random and the Mkv-2 genomes is the result of their similarity in dinucleotide composition: the evolved genome has a balance index very close to 1. It is most likely that the kinetoplast has been subjected to a model (i) kind of selection, because it too shows a big difference in curvature compared to its Mkv-2 genome. In contrast, all bacterial genomes show a good correlation between dinucleotide composition and curvature, which suggests that their curvatures did not result from a model (i) kind of selection.

Our fourth finding was that the dinucleotide compositions of bacterial genomes show no signs of selection for curvature. From our other findings we knew that the dinucleotide compositions of bacterial genomes are unbalanced and that this leads to their specific curvatures. We also knew that selection for curvature according to model (ii) is only one of the many possible causes for the imbalance in their dinucleotide compositions. However, we judged that the other causes would not make the dinucleotide composition special with regard to curvature specification. So, if the dinucleotide compositions of real genomes were improbable in that particular regard, they would be reflecting a selected curvature-altering mutational bias, which is what model (ii) states. Instead, we found that the curvature implied in the dinucleotide composition of bacterial genomes is not uncommon among similarly balanced aleatoric genomes. Thus, bacterial genomes do not appear to have been selected for curvature according to model (ii) either.

Among the bacterial genomes, \textit{H.pylori} deserves special comment. (Two strains of \textit{H.pylori} were analyzed but they are so similar that what applies to one applies to the other.) The \textit{H.pylori} genome is particular in three ways. First, it is by far the most curved bacterial genome. Second, for no other genome was its Mkv-2 genome so different in curvature (a difference of 0.5 d.p.t. in \textit{H.pylori}, compared to a difference of 0.3 d.p.t. in \textit{C.jejuni}, the next most different case). Third, when compared to aleatoric genomes of similar balance index, the curvature of \textit{H.pylori} Mkv-2 was relatively uncommon (it is among the 5% rarest; by these criteria only \textit{N.meningitidis} is more rare). Without being a statistical proof, the evidence could be used to argue that the curvature of \textit{H.pylori} has been selected. However, we have tried to come up with some adaptive reason for the extreme curvature of \textit{H.pylori}, without finding any reason that singles out this species. It is possible that \textit{H.pylori} looks like a special case because not enough genomes have been yet analyzed.

Our results do not deny that DNA curvature is important for particular, well-established bacterial functions, such as activation or repression of some promoters (3,30,31) and the binding of some proteins to DNA (32). It is important to bear in mind that global intrinsic curvature is an average. The lack of biological significance of this global property and the importance of curvature in particular processes can be reconciled if we assume that the latter only involve a small percentage of the genome. It reasonable to assume that most curvature selection has occurred in the non-coding regions, because most of the functions known to be affected by DNA curvature are related to these regions and because the coding regions are strongly compromised by their function. In a previous work we showed...
that non-coding regions, which usually represent <10% of the genome, contribute little to the global average (23). Furthermore, their curvature tends to be correlated with, and only slightly higher than, the genome average, indicating that even among the non-coding DNA, the sequences whose curvature has been specifically selected could be a minority. On the other hand, by scanning the genomes along their length we have tried to identify distinguishable curvature patterns or regions with special curvatures. However, this has been complicated by the fact that even random sequences can attain a broad range of local curvatures (see Fig. 1) and that the local curvature of bacterial genomes varies wildly, changing widely even within short DNA spans (<300 bp) (data not shown; but see 33).

Given that species differ in their global curvature, it is difficult to accept that this has no effect on their biology. In theory, global curvature should influence DNA supercoiling, the formation of the bacterial nucleoid and perhaps the integration of global transcriptional states such as the growth phase, osmoregulation and thermoregulation and others (34). However, it is likely that most of these dynamic changes in DNA are primarily actuated by interaction with proteins, in which case DNA flexibility (bendibility) is even more important than intrinsic curvature (35–37). Many pleiotropic proteins, including CAP, IHF, HU and HN-S, are known to function by bending DNA. On the other hand, σ54-dependent promoters provide examples where intrinsic curvature and protein-induced curvature happen to be alternative adaptations to the same problem (38). We believe that DNA flexibility and protein interactions allow bacterial species to accommodate their particular global intrinsic curvature with minimal cost to their overall fitness.

Since a near-helical periodicity of some dinucleotides is required for curvature to extend over a DNA of some length, it is interesting that in analyzing complete genomes Herzfel et al. (39) found periodicities of mono- and dinucleotides with the required period which appear to be distinctly related to DNA folding (and not to the coding of required period which appear to be distinctly related to DNA is interesting that in analyzing complete genomes Herzel et al. required for curvature to extend over a DNA of some length, it and that the local curvature of bacterial genomes varies wildly, changing widely even within short DNA spans (<300 bp) (data not shown; but see 33).

ACKNOWLEDGEMENTS

The authors thank Drs Enrique Morett and Juan Carlos Almagro and the two anonymous referees for useful comments to improve this manuscript. We would also like to thank Ricardo Ciria, Abel Linares, Juan Manuel Hurtado and Alma Martinez for computer support and Shirley Ainsworth for bibliographic assistance. This work was partially supported by UNAM-PAPIIT grant IN230798.

REFERENCES