

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

Identifying structured regions in *E. coli*DNA

ArticleInfo		
ArticleID	:	3633
ArticleDOI	:	10.1186/gb-2000-1-2-reports0056
ArticleCitationID	:	reports0056
ArticleSequenceNumber	:	30
ArticleCategory	:	Paper report
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2000-5-20 Received : 2000-5-20 OnlineDate : 2000-7-19
ArticleCopyright	:	BioMed Central Ltd2000
ArticleGrants	:	

Rachel Brem

Abstract

A new method for identifying biologically relevant sequences by their DNA structure has been described.

Significance and context

It is now becoming possible to predict some features of DNA structure. But many computational methods for this purpose focus on just bending, or stacking stability, or flexibility - that is, each program is restricted to a single structural feature. Pedersen *et al.* have designed a new approach that uses five of these single-feature programs simultaneously. If a region of DNA is given a high score in all five programs, the authors hypothesize that the region is biologically significant. The authors report and analyze these putatively significant regions in the genes, promoters and non-coding regions of 18 prokaryotic genomes. The new methodology is important, in that its signal-to-noise ratio may be very much greater than that in individual programs: it may pick out biologically relevant sequences where other methodologies cannot.

Key results

Pedersen *et al.* list 20 putatively significant regions of 'extreme structure' - that is, regions predicted to be more significantly structured than controls - in the genome of *Escherichia coli*. Only one of these - an operon containing the uncharacterized *rhsE* gene - has been previously identified. The authors also cluster all *E. coli* genes with respect to bending score, stacking stability score, and so on, as scored by the programs. At least 8 of the resulting 11 clusters are enriched for genes involved in specific functions, such as respiration. (There is no control for significance level in this calculation.) Lastly, Pedersen *et al.* study the differences in bending, stacking stability, and other parameters between coding and non-coding DNA across all genomes, relative to shuffled controls. Although trends do not stand out with strong significance in these data, the authors determine that intergenic DNA containing promoters is more curved, less flexible and less stable than coding DNA.

Methodological innovations

The authors use previously documented programs that score di- or tri-nucleotides via empirical parameters trained on the following types of data: DNaseI cutting frequencies, which report flexibility; nucleosome binding, which reports flexibility; disparity of positions in X-ray crystal structures of DNA bound to proteins, which reports deformability; quantum-mechanical energy calculations, which report stability; and mobility on polyacrylamide gel electrophoresis, which reports curvature. Pedersen *et al.* apply each program to each di- or tri-nucleotide in a genome of interest, then identify significant 1000 bp regions as those containing many di- or tri-nucleotides given high scores by all five programs. Similar calculations on shuffled genomes provide a control, which establishes the probability of finding high-scoring regions by chance.

Conclusions

The authors speculate that several of their 20 predicted regions of 'extreme structure' in the *E. coli* genome may be positions of kinks in supercoiled DNA. They also speculate, on the basis of results from their 11 clusters of *E. coli* genes, that functionally related genes might have similar DNA structure. And their finding that promoter DNA is less stable and more curved is consistent with biochemical hypotheses: during transcriptional initiation, the double helix needs to unwind easily, and it is also believed to wrap around the RNA polymerase molecule.

Reporter's comments

The methodology in this paper is sound and potentially important, but it is hard to evaluate the results fully because they contain few positive controls. The next step should be experimental verification of the authors' 20 putatively significant DNA regions. Then can Pedersen *et al.* can make a convincing case that their new tool makes truly useful predictions.

Table of links

[Journal of Molecular Biology](#)

References

1. Pedersen AG, Jensen LJ, Brunak S, Staerfeldt H-H, Ussery DW: A DNA structural atlas for *Escherichia coli*. J Mol Biol. 2000, 12: 907-930. 0022-2836