Computational Prediction of Gene Regulation by DNA Conformation

Mick Noordewier[1] **Peter Morrison**[1] **Craig Benham**[2]
`Mick_O_Noordewier@sbphrd.com` `Peter_H_Morrison@sbphrd.com` `Benham@camelot.mssm.edu`

[1] SmithKline Beecham Pharmaceuticals, 1250 S. Collegeville Rd, PO Box 5089, Collegeville, PA 19426-0989, USA
[2] Mt. Sinai Medical School, 1 Gustav L Levy Place, NY, NY 10029-6574, USA

Keywords: superhelical, destabilization

1 Introduction

DNA within living organisms contains regions of superhelicity, which imposes stress on the molecule. B-form duplex DNA can relieve such stress through localized strand separation at specific regions in the DNA. Computational methods have been developed to predict the locations and extent of destabilization based solely on DNA sequence (2). The results of these analyses agree precisely with experimental determinations of the extents and locations of denatured regions, as found by nuclease digestion, both *in vitro* and *in vivo* (5). In this report we utilize these methods to predict the destabilization properties of other sequences on which biological experiments have not been performed.

2 Methods and Results

Analysis of genomic sequences reveals that predicted destabilized regions are found to be closely associated with several specific types of DNA regulatory regions. The strong associations found between predicted stress-destabilized sites and certain regulatory regions suggests that stress-induced duplex destabilization may be involved in their mechanisms of action. *Iin vivo* experimental results support this conclusion in several cases (5). The most strongly destabilized sites have been found in the 3' flanks of genes (3). This pattern is detected in prokaryotes, eukaryotic viruses, yeast and humans (2). An analysis of 26 yeast genes found that promoter and terminal regions were destabilized, but the region encoding the primary transcript was not. Several coordinately regulated families of prokaryotic genes exhibit characteristic patterns of destabilization (4).

We have characterized the presence of a strong bias to havefor destabilized chromatin localized upstream and proximal to gene start sites in prokaryotes. The concept that superhelicity may regulate gene expression in prokaryotes has been supported by numerous studies that have modified gyrase and toposiomerase activities in bacteriaorganisms . These proteins act to modulate the torsional strain of DNA and thereby impact alter the propensity for DNA unwinding (5). However tThere is evidence, however, to suggest that the proteins are directed to bind to specific destablized denatured or single-stranded sequences that may subsequently further stabilize the unwound (or open) DNA conformation (6). Prokaryotes display discernable patterns of regulated gene expression when they are subjected to changing environmental conditions that result in modulating chromosome torsional strain (5) (e.g. oxygen deprivation and osmotic stress). Such coordinated regulation is an example of how the innate sequence of the genome could allow organisms to rapidly respond to external stimuli through the winding/unwinding of gene regulatory regions of the chromosome.

Our preliminary analysis of the entire *E. coli* sequence demonstrates that a large proportion of genes have a region of destabilization immediately upstream to the gene start site while little or no such correlation is seen downstream of the gene terminus or within the coding sequence. This same pattern of destabilization is observed in the portion of the *B. subtilius* chromosome examined as well. The absence of complete concordance would be anticipated due to the presence in prokaryotes, of operons, in which co-expressed proteins down stream from the promoter would lack their own transcription regulatory sites. Therefore we examined S. cerevisiae, a lower eukaryote that lacks operons. In this organism we found a significantly stronger correlation of destabilized sites immediately upstream of gene starts, in contrast to there their absencse within coding sequences, in further support of our regulatory model. Additionally, this also demonstrates its wide distribution throughout evolution. Finally we have evidence to suggest that this analytic tool has potential for greatly enhancing our ability to rapidly analyze large genomes.

References

[1] Aranda, A., Perez-Ortin, J.E., Benham, C.J. and Olmo, M.L.D., Analysis of the structure of a natural alternating d(TA)n sequence in yeast chromatin, *Yeast*, 13(4):313–326, 1997.

[2] Benham, C.J., Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci, *Proc. Natl. Acad. Sci. USA*, 90(7):2999–3003, 1993.

[3] Benham, C.J., Computation of DNA structural variability–a new predictor of DNA regulatory regions, *Comput. Appl. Biosci.*, 12(5):375–381, 1996.

[4] Benham, C.J., Duplex destabilization in superhelical DNA is predicted to occur at specific transcriptional regulatory regions, *J. Mol. Biol.*, 255(3):425–434, 1996.

[5] Sheridan, S.D., Benham, C.J. and Hatfield, G.W., Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence, *J. Biol. Chem.*, 273(33):21298–21308, 1998.

[6] Werten, S., Langen, F.W., Schaik, R.v, Timmers, H.T., Meisterernst, M. and Vliet, P.C.v. d, High-affinity DNA binding by the C-terminal domain of the transcriptional coactivator PC4 requires simultaneous interaction with two opposing unpaired strands and results in helix destabilization, *J. Mol. Biol.*, 276(2):367–377, 1998.