**DNA STRUCTURE IN HUMAN RNA POLYMERASE II PROMOTERS.**

PEDERSEN ANDERS GORM1, BALDI PIERRE2, CHAUVIN YVES2, BRUNAK SOREN1

1Center for Biological Sequence Analysis, The Technical University of Denmark, building 208, DK-2800 Lyngby, Denmark.

2Net-ID, Inc., 4225 Via Arbolada, suite 500, Los Angeles, CA 90042, USA

Keywords: DNA structure, human RNA polymerase II promoters, hidden Markov models, DNA bendability, structural profile, transcriptional start

The fact that DNA three-dimensional structure is important fortranscriptional regulation begs the question of whether eukaryoticpromoters contain general structural features independently of whatgenes they control. We present an analysis of a large set of human RNApolymerase II promoters with very low sequence similarity. Thesequences, which include both TATA-containing and TATA-less promoters,are aligned by hidden Markov models (HMMs). Using three differentmodels of sequence-derived DNA bendability, the aligned promotersdisplay a common structural profile with bendability being low in aregion upstream of the transcriptional start point and significantlyhigher downstream. Investigation of the sequence composition in the tworegions shows that the bendability profile originates from thesequential structure of the DNA, rather than the general nucleotidecomposition. Several trinucleotides known to have high propensity formajor groove compression are found much more frequently in the regionsdownstream of the transcriptional start point, while the upstreamregions contain more low-bendability triplets. Within the regiondownstream of the start point, we observe a periodic pattern insequence and bendability, which is in phase with the DNA helicalpitch. The periodic bendability profile shows bending peaks roughly atevery 10 bp with stronger bending at 20 bp intervals. Theseobservations suggest that DNA in the region downstream of thetranscriptional start point is able to wrap around protein in a mannerreminiscent of DNA in a nucleosome. This notion is further supported bythe finding that the periodic bendability is mainly caused by thecomplementary triplet pairs CAG/CTG and GGC/GCC, which previously havebeen found to correlate with nucleosome positioning. We present modelswhere the high bendability regions position nucleosomes at thedownstream end of the transcriptional start point, and consider thepossibility of interaction between histone-like TAFs and this area. Wealso propose the use of this structural signature in computationalpromoter finding algorithms.