A Systematic Analysis of Stress Induced DNA Duplex Destabilization (SIDD) Sites in the *E. coli* Genome: Implications of SIDD Analysis for Promoter and Operon Prediction in Prokaryotes

Huiquan Wang[1] and Craig J. Benham[2]

Keywords: SIDD, promoter, operon prediction, transcription regulation, DNA supercoiling

Introduction

DNA structure and topologically driven structural transitions have been suggested to play important roles in regulating gene expression (1). Stress induced DNA duplex destabilization (SIDD) analysis exploits the known structural and energetic properties of DNA to predict the sites which are susceptible to become separated under superhelical stress (2, 3). Experimental results show that this analysis is quantitatively accurate in predicting transcriptional regulatory regions, matrix/scaffold attachment sites and replication origins (4, 5). Here we report a systematic analysis of the SIDD profile of the *E. coli* genome using a new algorithm specific for long genomic DNA sequences (6).

Results

1. Less than 7% of the *E. coli* genome has the propensity to be destabilized at the physiological superhelical densities (Figure 1).
2. Sites with high destabilization potential are statistically significantly associated with divergent and tandem intergenic regions, but not with convergent intergenic regions, and they strongly avoid coding regions (Figure 2).
3. More than 80% of the intergenic regions containing experimentally characterized promoters are found to overlap these SIDD sites (Figure 3).
4. A large majority of SIDD sites overlap long tandem intergenic regions, suggesting a potential role of SIDD sites in defining operon boundaries (Figure 4).
5. Strong SIDD sites are also found in the 5' upstream regions of genes regulating stress responses in *E. coli*, suggesting a possible link between their locations, the degrees of their destabilization, and the functioning of these genes (Table 1).

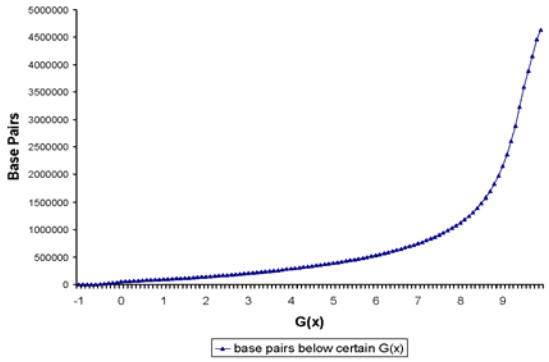

Figure 1. The cumulated G(x) distribution, the number of base pairs destabilized below the specified value.

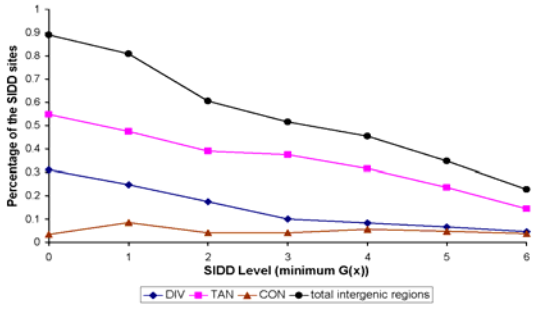

Figure 2. Percentage of SIDD sites at each level overlapping intergenic regions in E. coli genome

---

[1,2] UC Davis Genome Center, University of California, One Shields Avenue, Davis, CA 95616. Email: hqwang@ucdavis.edu; cjbenham@ucdavis.edu

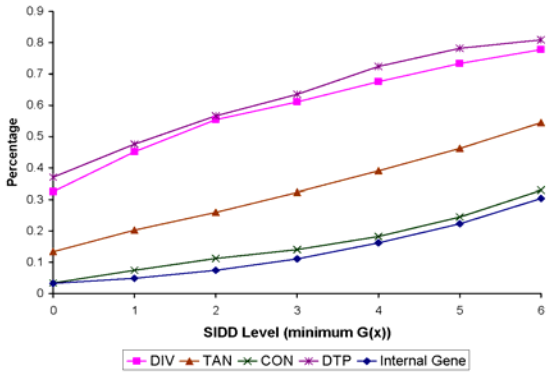

Figure 3: Percentage of DIV, TAN, CON, DTP and internal gene regions overlapping SIDD sites at each level in E. coli genome

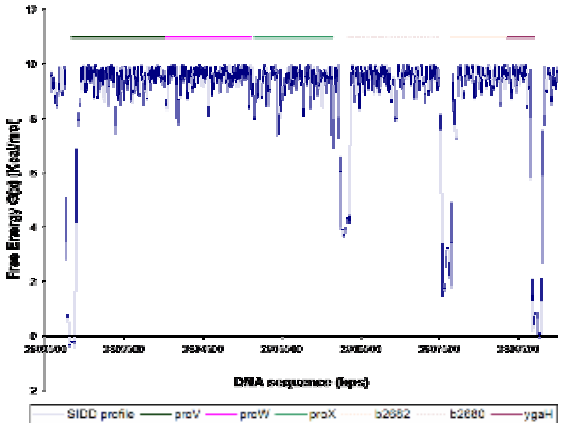

Figure 4. Strong SIDD sites are located at the boundaries of the proU operon. The genes are given above the graph and transcribe directly.

Table I. Global transcriptional regulators for stress responses with their 5' upstream regions overlapping the highly destabilized SIDD sites

Stress	SIDD	Gene	Function
Osmotic, nutrition starvation, cold	0	rpoS	Sigma S (sigma 38) factor of RNA polymerase, major sigma factor during stationary phase
Same as above	2	gyrA	DNA gyrase, subunit A, typeII topoisomerase
Same as above	0 0	hupA hupB	DNA-binding protein HU-alpha (HU-2), plays a role in DNA replication and in rpo translation
Same as above	1	H-NS	Transcriptional regulator, DNA-binding protein HLP-II, increases DNA thermal stability
Same as above?	1	crp	Transcriptional regulator, cyclic AMP receptor protein (cAMP-binding family), interacts with RNAP
Aerobic/anaerobic	0	fnr	Transcriptional regulator of aerobic, anaerobic respiration, osmotic balance (cAMP-binding family)
Aerobic/anaerobic	0	narX	Sensory histidine kinase in two-component regulatory system with NarL, regulation of anaerobic respiration and fermentation, senses nitrate/nitrite
Osmotic shock	0	ompR	response regulator in two-component regulatory system with EnvZ, regulates ompF and ompC expression (OmpR family)

References

1. Hatfield GW and Benham CJ, 2002. DNA Topology-Mediated Control of Global Gene Expression in Escherichia coli, *Annu. Rev. Genet*. **36,** 175-203.
2. Benham CJ, 1992. The energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.* **225**, 835-847.
3. Benham CJ, 1996. Duplex destabilization in superhelical DNA is predicted to occur at specific transcriptional regulatory regions. *J. Mol. Biol.* **255**, 425-34.
4. Sheridan SD, Benham CJ and Hatfield GW, 1998. 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**, 21298-21308.
5. Leblanc BP, Benham CJ and Clark DJ, 2000. An initiation element in the yeast CUP1 promoter is recognized by RNA polymerase II in the absence of TATA box-binding protein if the DNA is negatively supercoiled. *Proc. Natl. Acad. Sci. USA* **97**, 10745-50.
6. Benham CJ and Bi C-P, 2004. The Analysis of Stress-Induced Duplex Destabilization in Long Genomic DNA Sequences *J. Comput. Biol.* to appear