

Strong SIDD (Stress Induced DNA Duplex Destabilization) sites occur at regulatory elements of the genes initiating responses to physiological and environmental stresses in *E. coli*.

Huiquan Wang¹ and Craig Benham²

Keywords: Stress Induced DNA Duplex Destabilization (SIDD), DNA structural motif, transcriptional regulation

1 Introduction.

DNA structure and topologically driven structural transitions have been suggested to play important roles in regulating gene expression. Stress induced DNA duplex destabilization (SIDD) analysis exploits the known structural and energetic properties of DNA to predict the sites which are susceptible to be separated under superhelical stress. It has been shown that predicted SIDD sites preferentially occur at transcriptional regulatory loci containing promoters or terminators, and avoid coding regions [1, 2]. Statistical analysis shows that an intergenic region has an 80% chance of containing a promoter if it is associated with a strong SIDD site; while an intergenic region without a strong SIDD site has only a 24% chance. However, only about 32% of divergent regions and about 13% of the tandem intergenic regions contain the strongest SIDD sites in *E. coli* genome. Here we investigate the gene locations of these strong SIDD sites. We find that they tend to occur at positions where the expression of genes involved in the responses to physiological or environmental stresses are regulated.

2 Methods.

1. SIDD profile and a SIDD site: The SIDD profile of *E. coli* K12 genome is calculated as described in [3]. The extent of destabilization is given by the parameter $G(x)$, the free energy needed to force the base pair at position x to always be open. More destabilized sites have smaller values of $G(x)$. A SIDD site defined as a consecutive set of base pairs where $G(x) < 8$. Further, the SIDD sites are binned into 8 disjoint groups, according to the minimum value of $G(x)$ they attain.

2. Distribution of SIDD sites among gene function categories: The *E. coli* gene classification by gene type and cellular role made by Dr. Riley's lab [4] were used here. Genes were first grouped according to whether they had a SIDD site in their 5' upstream regions within 50 bps of their start site. The genes in each SIDD group were further clustered according to their gene product types defined in the ProtEC database. The distribution among the defined gene type categories of the genes in each SIDD group follows a hypergeometric distribution [5].

3 Results.

^{1, 2} Bioinformatics program in Genome Center, UC Davis, One Shields Avenue, Davis, CA, USA, Email: hqwang@ucdavis.edu, cjbenham@ucdavis.edu

Here we analyzed the association of different gene type/product categories with SIDD sites in *E. coli* genome. Our results indicate:

1. The strongest SIDD sites in *E. coli* genome are statistically associated with the immediate upstream regions of genes encoding regulators, membrane proteins and transport proteins, while the immediate upstream regions of the RNA genes and the gene from external origin are highly stable. (Figure 1)
2. Particularly within the regulator category, the upstream regions of the LysR type family regulators and the two-component systems are commonly highly destabilized. (Figure 2)
3. Similar destabilization characteristics are found in the target genes regulated by LysR regulators, but not by two-component systems.
4. Comparisons of SIDD properties across gene product/type categories suggest that genes respond to physiological or environmental changes tend to have more destabilized upstream regions than others. (Table 1)

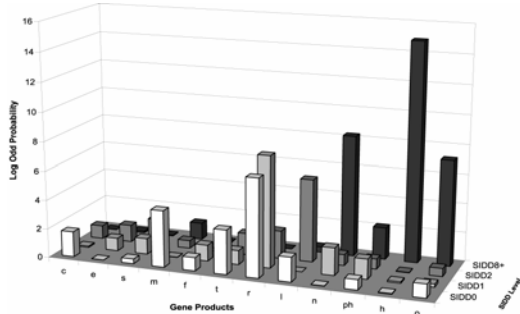


Figure 1. Distribution of SIDD sites in gene product function categories in *E. coli*

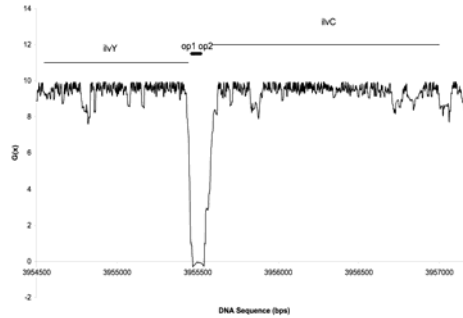


Figure 2. A SIDD site at level 0 located in the regulatory region of the gene pair *ilvY-ilvC*.

Table I. Contrast SIDD properties in the upstream regions of the genes responding to physiological and environmental stimulations and those do not.

SIDD	Gene	Functions	Stress response
0/7	<i>rpoS/rpoD</i>	Sigma factors, transcription factor	Yes/No
2/6	<i>gyrA/topA</i>	DNA topoisomerases	Yes/Yes?
0/4	<i>hupA/himA</i>	DNA binding proteins	Yes/No?
1/6	<i>crp/lrp</i>	Global transcriptional regulators	Yes/No
0/7	<i>proV/glnQ</i>	Transport proteins	Yes/No
0/8	<i>ompC/fadL</i>	Membrane proteins	Yes/No

4 References.

- [1] Benham, C.J. 1993. Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proc Natl Acad Sci U S A.*, **90**(7): p. 2999-3003.
- [2] Wang, H., Noordewier, M. and Benham, C.J. 2004. Stress-Induced DNA Duplex Destabilization (SIDD) in the Escherichia coli Genome: SIDD Sites are Closely Associated with Promoters. *Genome Res.*: **14**(8):1575-1584.
- [3] Benham, C.J. and Bi, J. 2004. The analysis of stress-induced duplex destabilization in long genomic DNA sequences. *J Comput Biol.*:**11**(4):519-43.
- [4] Serres, M.H., Goswami, S. and Riley, M. 2004 GenProtEC: an updated and improved analysis of functions of Escherichia coli K-12 proteins. *Nucleic Acids Res.*, **32 Database issue**: p. D300-2.
- [5] Sokal, R.R. and Rohlf, F.J. 1995. *Biometry: The principles and practice of statistics in biological research*. Third Edition ed., New York: W.H.Freeman & Company.