

## **WebSIDD: Server for Prediction of the Stress-induced Duplex Destabilized Sites in Superhelical DNA**

Chengpeng Bi and Craig J. Benham

UC Davis Genome Center, University of California, One Shields Ave., Davis CA 95616

### **ABSTRACT**

**Summary:** WebSIDD is a web-based service designed to predict locations and extents of stress-induced duplex destabilization (SIDD) that occur in a double-stranded DNA molecule of specified base sequence, on which a specified level of superhelical stress has been imposed. The algorithm calculates the approximate equilibrium statistical mechanical distribution of a population of identical molecules among the accessible states. The input to the program is a DNA sequence, and its output is the calculated transition probability and destabilization energy of each base pair in the sequence. The structural and energy parameters used in the calculation are all determined experimentally. This method has illuminated the roles of SIDD properties in the regulation of diverse biological processes, including transcriptional initiation and termination, and the eukaryotic nuclear scaffold attachments that partition chromosomes into domains. WebSIDD should prove useful for the analysis of SIDD sites in genomic sequence.

**Availability:** <http://genome.bme.ucdavis.edu/sidd/>.

**Contact:** [cjbenham@ucdavis.edu](mailto:cjbenham@ucdavis.edu)

DNA is constrained into topological domains *in vivo*, typically several kilobases in length, consisting of closed loops within a chromosome that are formed by periodic attachments of the chromatin fiber to the nuclear matrix (Alberts *et al.*, 2002). The topological constraint on a closed-loop domain is precisely equivalent to that on a circular molecule; in both cases the linking number  $Lk$  is fixed. This value is regulated *in vivo* by a variety of processes involving transient strand breakage and religation. In this

way the actual linking number  $Lk$  can be varied from its relaxed value  $Lk_o$ , so a linking difference  $\alpha = Lk - Lk_o$  is imposed. This phenomenon is called DNA superhelicity. DNA superhelicity, which is closely regulated *in vivo*, can induce the formation of locally unpaired regions at defined sites within DNA molecules. Nuclease digestion experiments have shown this local denaturation to occur at specific regulatory regions. The initiation of replication in both prokaryotes and yeast has been shown to require the presence at a precise position of a site that is susceptible to superhelical strand separation (Kowalski and Eddy, 1989; Huang and Kowalski, 1993). When the base sequence of this site is altered, replication occurs *in vivo* only if the susceptibility to stress-induced denaturation at the correct position is retained. SIDD sites have also been shown to occur at chromosomal attachment regions (Benham *et al.*, 1997). These attachments are known to augment transcription, and to form barriers between independently regulated domains (Bode *et al.*, 1996). Sites susceptible to DNA duplex destabilization also occur at binding sites for other molecules such as transcription factors and other regulators. In several cases the regulatory proteins require locally denatured DNA to bind (Rothman-Dees *et al.*, 1998). Here, we provide a publicly available online tool, termed WebSIDD, which performs computation of the stress-induced DNA duplex destabilization (SIDD) and predicts the SIDD sites on a DNA molecule on which negative superhelicity has been imposed. The algorithm is based on a statistical mechanical SIDD analysis procedure that has been presented earlier (Benham, 1992). A detailed description of how to use the web software is also available online.

### **WebSIDD options**

WebSIDD uses an approximate approach that finds all states whose free energy does not exceed a specified threshold. The free energy associated to each state is comprised of three terms: the chemical energy for separation of strands, the torsional energy for rotation of the single strands within denatured regions, and the residual supercoiling free energy. The chemical energy is associated with each state of base-pairing of the DNA, and is computed based upon either copolymeric or near neighbor energetics. WebSIDD calculates two properties (transition probability and destabilized energy in each base pair)

that describe the destabilization experienced by the input sequence at a specified stress level.

The WebSIDD program has a simple HTML interface that enables the user to control the input and output. To new users, four samples are provided in order to give a quick tour of how SIDD profiles are calculated. To carry out WebSIDD computing, the user can type or copy and paste a DNA sequence, and name it. The legal character set is {A, C, G, T, 0-9}. The output names are generated according to the sequence name. A circular DNA is assumed. If a linear DNA is specified, the program will add 50 G/C to the end and then handle it the same way as a circular molecule. The maximum DNA length is 10 kb. The larger the sequence size, the longer the computing time. If one needs to analyze longer or whole genomic DNA sequence, then sends the sequence to the authors.

There are two types of opening energies: either copolymeric or near neighbor. The copolymeric type assigns the same free energy value to every AT and a different value to every GC base pair. The near neighbor type assigns entropies and enthalpies to each of the ten different neighbor types, as measured by Klump (Steger, 1994). Both thermodynamic data were measured experimentally. It is up to the user which one is preferred. More low energy states will be found and thus more time spent if near neighbor energetics is chosen. The profile output can be either graphic or text or both. An energy cutoff is chosen to predict the potential SIDD sites. The user can also control other output, for example, showing the input sequence and statistics of low energy states.

There are other parameters that can be varied by the user if the advanced WebSIDD is activated. The default values of temperature and salt concentration are 37°C (310 K) and 0.01 M respectively, as they are the experimental settings used in the mung bean nuclease digestion procedure by which stress-induced strand opening is most accurately assessed *in vitro* (Kowalski *et al.*, 1988). However, we leave the possibility for the user to specify other conditions. We note, however, that the algorithm takes substantially longer to execute when a high temperature is selected. The default value of superhelix density is  $\sigma = -0.055$ . The current SIDD version only allows negative superhelicity as occurs *in vivo* under normal physiological situations. The number of states satisfying any given threshold increases approximately exponentially with the absolute value of  $\sigma$ . The

default threshold energy is 12.0 kcal/mol with which the program executes efficiently while providing high accuracy. In practice this threshold should not be decreased below 10.0 kcal/mol. The maximum size of the open region is set in between 200 and 250 bp. Three important energy parameters were set as their defaults: the quadratic coefficient ( $K = 2220RT/N$ ), the torsional stiffness ( $C = 1.91 \text{ kcal/mol/rad}^2$ ) and the initial energy ( $a = 10.16 \text{ kcal/mol}$ ). These values were optimally estimated based on experimental data (Kowalski *et al.*, 1988).

The program provides a detailed profile including the base position, transition probability and destabilized free energy. The profile can be downloaded. Graphic profiles are also generated on-the-fly in GIF format. Given an energy cutoff, the program outputs possible locations and extents of SIDD sites. Proteins play important roles in DNA packaging and conformational transitions. It might be helpful to understand the dynamical process of DNA structural transitions if we could incorporate the DNA-protein interactions to the SIDD model in the future.

## **ACKNOWLEDGEMENTS**

This work was supported in part by grants DBI 99-04549 from the National Science Foundation, and RO1-HG01973 from the National Institutes of Health, and by additional support from the Diversa Corporation.

## **REFERENCES**

- Alberts,B., Bray,D., Lewis,J., Roberts,K. and Watson, J. D. (2002) *Molecular Biology of the Cell*. Page 216-223. Garland, NewYork.
- Benham,C.J. (1992) Energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.*, **225**, 835-847.
- Benham,C.J. (1993) Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proc. Natl. Acad. Sci. USA*, **90**, 2999-3003.
- Benham,C.J., Kohwi-Shigematsu,T. and Bode,J. (1997) Stress-induced duplex destabilization in chromosomal scaffold/matrix attachment regions. *J. Mol. Biol.*, **274**, 181-196.

- Bode,J., Stengert-Iber,M., Kay,V., Schlake,T. and Dietz-Pfeilstetter,A. (1996) Scaffold/Matrix-Attached regions: topological switches with multiple regulatory functions. *Crit. Rev. Eukaryot. Gene Expr.*, **6**, 115-138.
- Huang,R.Y. and Kowalski,D. (1993) A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. *EMBO. J.*, **12**, 4521-4531.
- Kowalski,D. and Eddy,M.J. (1989) The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *E. coli* replication origin. *EMBO. J.*, **8**, 4335-4344.
- Kowalski,D., Natale,D. and Eddy,M. (1988) Stable DNA unwinding, not breathing, accounts for single-stranded specific nuclease hypersensitivity of specific A+T-rich regions. *Proc. Nat'l Acad. Sci. USA*, **85**, 9464-9468.
- Rothman-Dees,L.B., Dai,X., Davydova,E., Carter,R. and Kazmierczak,K. (1998) Transcriptional regulation by DNA structural transitions and single-stranded DNA-binding proteins. *Cold Spring Harbor Symp. Quant. Biol.*, **63**, 63-73.
- Steger,G. (1994) Thermal denaturation of double-stranded nucleic acids. *Nucleic Acids Res.*, **22**, 2760-2768.