BIOINFORMATICS APPLICATIONS NOTE Vol. 20 no. 9 2004, pages 1477–1479
DOI: 10.1093/bioinformatics/bth304

WebSIDD: server for predicting stress-induced duplex destabilized (SIDD) sites in superhelical DNA

Chengpeng Bi† *and Craig J. Benham**

UC Davis Genome Center, University of California, One Shields Avenue, Davis, CA 95616, USA

Received on May 8, 2003; revised on June 11, 2003; accepted on July 29, 2003
Advance Access publication May 6, 2004

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 is imposed. The algorithm calculates the approximate equilibrium statistical mechanical distribution of a population of identical molecules among its accessible states. The user inputs the DNA sequence, and the program outputs the calculated transition probability and destabilization energy of each base pair in the sequence. As options, the user can specify the temperature and the level of superhelicity. The values of all structural and energy parameters used in the calculation have been experimentally measured. WebSIDD should prove useful for finding SIDD-susceptible sites in genomic sequences, and correlating their occurrence with locations involved in regulatory and pathological processes. This strategy already has illuminated the roles of SIDD in diverse biological regulatory processes, including transcriptional initiation and termination, and the eukaryotic nuclear scaffold attachments that partition chromosomes into domains.

Availability: http://orange.genomecenter.ucdavis.edu/benham/sidd/index.html

Contact: cjbenham@ucdavis.edu

INTRODUCTION

DNA is constrained into topological domains *in vivo*, typically several kilobases in length, consisting of either a circular molecule or closed loops within chromosomes 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 of *Lk* may be changed by processes involving transient strand breakage and religation. In this manner, *Lk* can be changed from its relaxed value Lk_0, which imposes a linking difference $\alpha = Lk - Lk_0$ (also called DNA superhelicity) on the domain. DNA superhelicity is closely regulated *in vivo* by enzymatic and other processes. It can drive the formation of locally unpaired regions at defined sites within DNA molecules (Kowalski *et al*., 1988). These commonly occur at specific regulatory regions (Benham, 1993).

The initiation of replication in both prokaryotes and yeast has been shown to require the presence of a site at a precise position 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. Stress-induced duplex destabilization (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. SIDD has also been shown to be involved in transcriptional regulation. The $ilvp_G$ promoter of *Escherichia coli* is activated by an IHF binding-induced transmission of destabilization from the binding site to the −10 region of the promoter (Sheridan *et al*., 1998). 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 for binding (Rothman-Denes *et al*., 1998).

Here, we present WebSIDD, a publicly available online tool that calculates the SIDD properties of a user-specified DNA sequence, on which a standard level of negative superhelicity has been imposed. The algorithm is based on a statistical mechanical SIDD analysis procedure that has been presented elsewhere (Benham, 1992; Benham and Bi, 2004). All thermodynamic parameter values used in WebSIDD are taken from experimental measurements; there are no free parameters. Yet, it makes quantitatively accurate predictions of the locations and extents of separation as functions of imposed

*To whom correspondence should be addressed.

†Present address: Laboratory of Human Molecular Genetics, Children's Mercy Hospital and Clinics, Kansas City, MO 64108, USA.

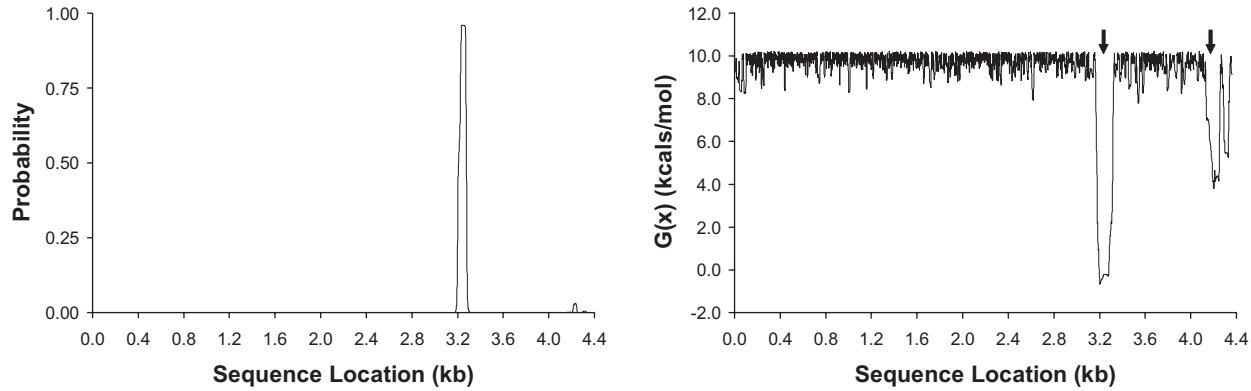

Fig. 1. The probability profile (left) and the SIDD destabilization energy profile (right) are shown for pBR322 DNA sequence, as calculated by WebSIDD using near neighbor energetics. These results are in quantitatively precise agreement with experimental measurement (Kowalski *et al*., 1988). The energy profile is more informative than the probability profile, as it shows that destabilization is confined to specific sites, with most of the DNA sequence remaining virtually as stable as in an unstressed molecule. The SIDD sites are the terminator (left arrow) and the promoter (right arrow) of the β-lactamase gene.

superhelicity. This makes the WebSIDD algorithm useful for finding SIDD sites in other molecules, on which experiments have not been performed.

WebSIDD uses an approximate approach that finds all states whose free energies do not exceed a specified threshold. The free energy associated with each state is comprised of three terms: the base pair-specific energy of strand separation, the torsional energy for rotation of the single strands within denatured regions and the residual supercoiling free energy. The separation energy can be assigned by the user to be either co-polymeric, assigning one energy value to every AT and a different value to every GC base pair, or to include near neighbor effects. Near neighbor energetics assign entropies and enthalpies to each of the 10 different neighbor types, as measured by Klump (Steger, 1994). More low-energy states are found, and thus more time is required, if near neighbor energetics are chosen. The results are quite similar in both cases (Benham and Bi, 2004).

WebSIDD calculates two properties for each base pair in the sequence—its transition probability $p(x)$ and its destabilized energy $G(x)$. Examples are shown in Figure 1. $G(x)$ corresponds to the incremental free energy needed to force the base pair at position x always to be open. It is a more useful parameter, because it also finds sites that are fractionally destabilized. That is, the energy needed to open such a site is reduced by superhelicity, but not to the point where it opens with high probability, as occurs at the right-most site in the example. Fractional destabilization can be biologically important, as it may render a site susceptible to opening by some other process, such as protein binding.

WebSIDD OPTIONS

The WebSIDD program has a simple HTML interface. Four samples are provided in order to give new users a quick view of how SIDD profiles are calculated. A detailed description of how to use the Web software is also available online. The user can either type in or copy-and-paste the DNA sequence to be analyzed. A size limit of 10 kb is imposed. The legal character set is A, C, G, T (case insensitive), and 0–9. So only completely sequenced regions should be analyzed, but the line numbers need not be stripped out. Output names are generated that correspond to the user-specified 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 in the same manner as a circular molecule. The default values of temperature and salt concentration are 37°C (310 K) and 0.01 M, respectively, as these are the environmental conditions used in the mung bean nuclease digestion procedure by which stress-induced strand opening is most accurately assessed *in vitro* (Kowalski *et al*., 1988). The default stress level is superhelix density $\sigma = \alpha/Lk_0 = -0.055$, which corresponds to a midrange superhelicity found in plasmids extracted from bacteria. The other required energy parameters are assigned their experimentally measured values. These are the supercoiling free energy K, the torsional stiffness C of interstrand winding within separated regions, and the nucleation free energy a. The temperature dependencies of all free energy parameters are known, enabling calculations at different temperatures to be performed (Steger, 1994; Bauer and Benham, 1993). Two other parameters used in the calculation are the threshold energy θ for inclusion of a state, and the maximum length $l_{\max}$ of open regions encountered. The default values of these parameters are $\theta = 12.0$ kcal/mol and $l_{\max} = 250$ bp, respectively.

The program provides detailed profiles, including the position, identity, transition probability and destabilized free energy of each base pair. These profiles can be downloaded. Graphic profiles are also generated on-the-fly in GIF format. The profile output can be selected as either graphic or text, or both. The user can also control other outputs, e.g. showing the

input sequence and the statistics of occurrence of low-energy states.

WebSIDD allows the user to change some default settings. The current version only allows negative superhelicity, as occurs *in vivo* under normal physiological situations in all except thermophilic organisms. The temperature may be varied, although the algorithm takes substantially longer to execute (and may lose accuracy) when a high temperature is selected. The ionic strength at present cannot be changed, because the dependences of all energy parameters on this quantity are not known.

DISCUSSION

The purpose of the WebSIDD calculations is simply to provide insight into the locations of SIDD-susceptible sites so that correlations with regulatory and other interesting types of regions can be assessed. Based on observed correlations, experiments can be designed to determine what role, if any, SIDD plays in the mechanisms of activity of those sites. This strategy has been successfully employed in several investigations to date. However, it is not the purpose of WebSIDD to quantitatively accurately mimic *in vivo* conditions. These are not well understood at present, and are likely to be much more complex than are the assumptions of the present calculations.

ACKNOWLEDGEMENTS

This work was supported in part by grants DBI 99-04549 from the National Science Foundation, 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*. Garland, NewYork.

Bauer,W.R. and Benham,C.J. (1993) The free energy, enthalpy and entropy of native and of partially denatured closed circular DNA. *J. Mol. Biol.*, **234**, 1184–1196.

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. and Bi,C.-P. (2004) The analysis of stress-induced duplex destabilization in long genomic DNA sequences. *J. Comput. Biol.* (in press).

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.

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. Natl Acad. Sci., USA*, **85**, 9464–9468.

Rothman-Denes,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.

Sheridan,S., Benham,C.J. and Hatfield,G.W. (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.

Steger,G. (1994) Thermal denaturation of double-stranded nucleic acids. *Nucleic Acids Res.*, **22**, 2760–2768.