

Detecting Z-DNA Forming Regions in the Human Genome

Jie Xiao¹
XIAO0020@ntu.edu.sg

Peter Dröge²
pdroge@ntu.edu.sg

Jinming Li^{1,*}
jml@ntu.edu.sg

¹ Division of Structural and Computational Biology, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551 Singapore

² Division of Genetics and Genomics, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551 Singapore

* To whom correspondence should be addressed

Abstract

In this work, we developed a novel computing tool *Z-Catcher* for detecting potential Z-DNA forming regions (ZDRs) in human DNA sequences. The program is capable of predicting ZDRs in the entire human genome, and the different Z-DNA forming potentials can be selected as the cutoff values for ZDR prediction. Using *Z-Catcher*, we constructed the first human ZDR map and analyzed the ZDR distribution in the human genome.

Keywords: Z-DNA forming regions, free energy, negative supercoiling density, transcriptional start sites

1 Introduction

In 1979, the first single-crystal structure of a DNA fragment was resolved. Unexpectedly it revealed a left-handed structure which surprised the biological academia, since they had focused on the right-handed B-DNA in the preceding 25 years. This structure has a special zigzag arrangement of the sugar-phosphate backbone which its name Z-DNA is named after. Since Z-DNA is a transient structure, it is very difficult to localize the ZDRs (Z-DNA forming regions) in the human genome by experimental methods. However, with the well formularized algorithm for calculating the free energy required in a B-to-Z transition, potential Z-DNA forming regions (ZDRs) can be predicted. To date, there is an online Z-DNA prediction program *Z-Hunt* available [1,2]. However, *Z-Hunt* is not applicable to a genome-wide study since it only accepts query sequences shorter than 1 Mbp. Therefore, we developed the *Z-Catcher* which is capable of handling sequences of very large size. Moreover, *Z-Catcher* enables the users to predict ZDRs with different Z-DNA forming potentials.

2 Method and Results

The program *Z-Catcher* reads a query sequence as dinucleotide blocks and assigns either an Anti-Syn- or a Syn-Anti- conformation to each dinucleotide. The free energy required for stabilizing a Z-DNA fragment is the sum of the energy requirement for stabilizing all dinucleotides. For a Z-DNA forming region the free energy released from the relaxation of the DNA helix should be sufficient to balance the free energy consumption of a B-to-Z transition [1,2,3]. A DNA fragment has a higher Z-DNA forming potential when it requires less free energy for performing a B-to-Z transition, so that the relaxation of a helix with greater supercoiling density (σ) could meet this energy requirement [3]. In *Z-Catcher*, a threshold supercoiling density (σ_0) is set for selecting ZDRs with different Z-DNA forming potentials. *Z-Catcher* only returns the

ZDRs which are able to flip into Z-DNA when its required σ is equal to or greater than the threshold σ_0 . Figure 1 is a diagram of the algorithm used in the *Z-Catcher*. The *Z-Catcher* in Perl script can be downloaded from http://vhp.ntu.edu.sg/zdna/Z_Catcher.zip.

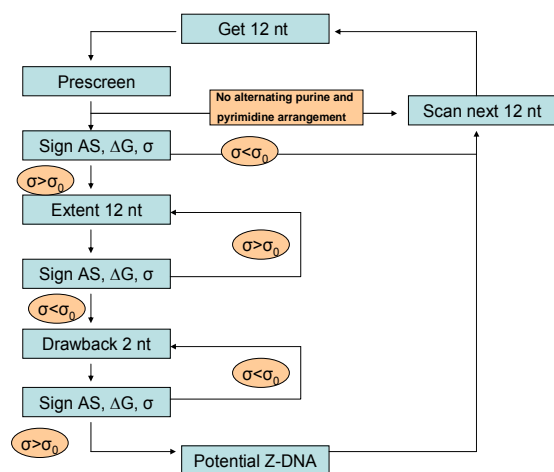


Figure 1 The diagram of the algorithm used in *Z-Catcher*

We scanned the entire human genome for potential ZDRs using *Z-Catcher*, and analyzed their distributions. Our results suggest that ZDRs are widely but not evenly distributed throughout the chromosomes. ZDRs are clustered in the upstream regions of the transcriptional start sites (TSSs), as well as in the regions that are 100 bps immediately downstream of the TSSs, see Figure 2.

3 Discussions

In this study, we developed a ZDR prediction program *Z-Catcher*. It accepts query sequences without size limitation, and therefore enables genome-wide large scale ZDR predictions. The performance of *Z-Catcher* was affirmed by comparing the predictions with the experimental observations. Our analysis suggested that potential ZDRs are enriched in the flanking areas around the human TSSs, especially in the [-600, 100] regions. Genes equipped with potential Z-DNA forming sequences might have a higher transcriptional activity. Our data suggest that the formation of Z-DNA may act as an active factor in the regulations of the transcriptional activities, and this type of regulation might be a widespread mechanism in the human genome.

References

- [1] Ho, P.S., et al., A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. *Embo. J*, 5(10): 2737-2744, 1986
- [2] Schroth, G.P., P.J. Chou, and P.S. Ho, Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J Biol. Chem.*, 267(17): 11846-11855, 1992
- [3] Frank-Kamenetskii, M.D. and A.V. Vologodskii, Thermodynamics of the B-Z transition in superhelical DNA. *Nature*, 307(5950):481-482, 1984.

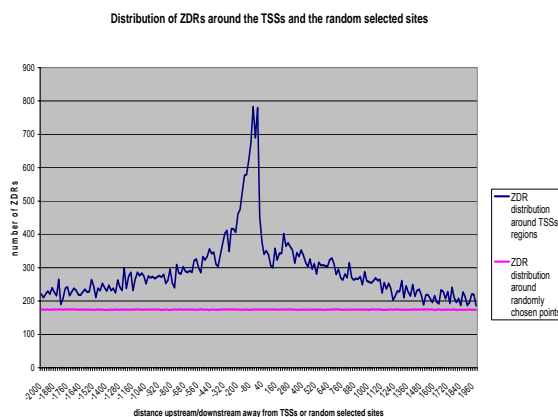


Figure 2 Distribution of ZDRs around TSSs and the randomly selected sites.