DNA conformational energy from molecular
dynamics simulations:
Toward the understanding of the indirect readout
mechanism in protein-DNA recognition

Marcos J. Araúzo-Bravo
marara@bse.kyutech.ac.jp
Shigeori Takenaka
staketcm@mbox.nc.kyushu-u.ac.jp

Shandar Ahmad
shandar@bse.kyutech.ac.jp
Hidetoshi Kono
kono@apr.jaeri.go.jp
Akinori Sarai
sarai@bse.kyutech.ac.jp

Satoshi Fujii
fujii@takenaka.cstm.kyushu-u.ac.jp
Nobuhiro Go
go@apr.jaeri.go.jp

1 Department of Biosciences of Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka, 820-8502, Japan
2 Department of Chemistry & Biochemistry, Kyushu University, Fukuoka, 812-8581, Japan
3 Neutron Research Center and Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute, Kyoto, 619-0215, Japan
4 Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, 630-0192, Japan

Keywords: protein-DNA recognition, indirect readout, molecular dynamics, DNA conformation

1 Introduction

Protein-DNA recognition plays an essential role in the regulation of gene expression. Regulatory proteins are known to recognize specific DNA sequences mainly by way of direct readout through base-amino acid contact and indirect readout through DNA conformation. In previous works [1], we have developed methods to derive statistical potentials from a database of protein-DNA complexes, and used them to quantify the specificities of direct and indirect readouts in protein-DNA recognition and to predict target binding sites.

However, the statistical potentials derived from structural databases suffer from the problems associated with a limited quantity of data and possible bias. To overcome these problems, we have performed molecular dynamics of DNA containing all possible tetranucleotides (three base pair steps), and used the trajectory data to calculate the potentials of mean force for conformational parameters of DNA.

2 Method

To calculate the DNA conformation energy we need to generate the force fields matrices F. This is done by approximating the energy of each dimer six-dimensional conformation fluctuation $\Delta \Theta$ by a harmonic function $E = \Delta \Theta^T F \Delta \Theta$.

The data for generating the force field were obtained from DNA molecular dynamics simulations of individual tetranucleotides. The six dimer conformation coordinates $\Theta$ were calculated using the 3DNA software package [2]. An iterative process was performed to reject the steps with fluctuations.
Figure 1: Scatter plot in the Shift-Slide plane derived from the trajectory of molecular dynamics for AAAA tetranucleotide. The ellipse is the projection of the six-dimensional equi-potential on the Shift-Slide plane.

Figure 2: Individual Z-scores of each tetranucleotide step of the DNA sequence of the 1a02.pdb (top) and 1ign.pdb (bottom) complexes.

greater than three standard deviations, as shown in Fig. 1. Then, the force field matrices were obtained by inverting the covariance matrix of the conformation fluctuations.

In order to test the method, our previously used dataset of 62 non-redundant protein-DNA complexes was employed [1]. The interaction energies and Z-scores were calculated as described in [3]. Examples of the Z-score at each base step of the DNA bound by the NFAT/Fos/Jun protein (1a02 PDB file) and by the Rap1 protein (1ign PDB file) are depicted in Fig. 2.

Currently we are analyzing the effectiveness of the molecular dynamics potentials for free DNA conformations.

3 Discussion

The molecular dynamic data allow us to estimate the conformational energy of DNA for any sequences and to calculate Z-score as a measure of specificity. The analysis will also bring the possibility of broader studies providing new insight into the relationship between specificity and structure in the process of protein-DNA recognition, which would lead to prediction of specific protein-DNA binding sites.

The work of first author was supported by a JSPS grant.

References

