Nonlinear Model of the Specificity of DNA-Protein Interactions and Its Stability

D Dwiputra\textsuperscript{1}, W Hidayat\textsuperscript{1,2}, R Khairani\textsuperscript{1}, F P Zen\textsuperscript{1,2}

\textsuperscript{1} Theoretical Physics Laboratory, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung Jl. Ganesha no. 10 Bandung, 40132, Indonesia
\textsuperscript{2} Indonesian Center of Theoretical and Mathematical Physics (ICTMP), Indonesia

E-mail: donny.dwiputra@itb.ac.id

Abstract. Specific DNA-protein interactions are fundamental processes of living cells. We propose a new model of DNA-protein interactions to explain the site specificity of the interactions. The hydrogen bonds between DNA base pairs and between DNA-protein peptide groups play a significant role in determination of the specific binding site. We adopt the Morse potential with coupling terms to construct the Hamiltonian of coupled oscillators representing the hydrogen bonds in which the depth of the potentials vary in the DNA chain. In this paper we investigate the stability of the model to determine the conditions satisfying the biological circumstances of the DNA-protein interactions.

1. Introduction
The molecular interactions between deoxyribonucleic acid (DNA) and proteins constitute the fundamental basis of all biological regulations. Both interact in realizing an extremely efficient transfer of information between them to provide various biological functions such as gene regulation, transcription, and replication. These processes are essentially realized by proteins that interact with specific DNA sequences. A traditional example concerning the dynamic interplay between DNA and proteins is the metabolism of lactose in the bacterium \textit{Escherichia coli} as a carbon source. Here an operator region of the DNA, called \textit{lac} operon, consists of structural genes: \textit{lacY}, \textit{lacZ}, and \textit{lacI} that can be expressed by RNA polymerase protein to metabolize lactose or repressed by \textit{lac repressor} which turns off expression of the \textit{lac} operon when lactose is not present \cite{1}. RNA polymerase binds to a specific promoter sequence of the \textit{lac} operon to transcribe the genes and synthesize the messenger RNA which serves as blueprint of the metabolizing proteins by ribosomes \cite{2}.

Origins of the specificity in DNA-protein recognition rely on the complex three-dimensional structural features of both macromolecules, i.e. the unique chemical signatures of the base sequences and the sequence-independent DNA shapes \cite{3–5}. The three-dimensional complex inheres numerous molecular bonds which stabilize its structures and regulate the interaction between DNA and binding proteins that is, for instance, electrostatic, van der Waals’, and hydrogen bond interactions. Thus the proteins can “feel” the DNA surface shapes and simultaneously recognize the sequences in virtue of the above interactions. However, a full consideration of the atomistic structure would lead to a cumbersome and expensive molecular simulations. In fact not all of the interactions give significant contribution to the site specificity.
The hydrogen bonds role the specific base readout more significantly than the other interactions, since the transition free energy between best specific binding and nonspecific binding is approximately $16 k_B T$ below the specific binding energy [6] (experimentally, for instance, $17 k_B T$ for *Mnt* and $\approx 16 k_B T$ for *lac* repressor [7]). As a comparison, the nonspecific binding free energy of the repressor protein CI in *λ* virus-infected *E. coli* cells (in vivo) was estimated as only $7 k_B T$ [6]. Concisely, the concept governing specific and nonspecific DNA-binding is as follows: the electrostatic interactions drive between the negatively charged DNA backbone and positively charged amino acids on the protein binding surface drive the nonspecific binding [8], and the hydrogen bonds determine the specific sequence-dependent interactions. The other interactions such as van der Waals’ keep the proteins wandering near DNA together with the electrostatic interactions.

In this paper we present a new model governing the dynamics of the site specific DNA-protein interaction. We assume a simple, untwisted, ladder-like DNA double strand interacting with a regulatory protein through the hydrogen bonding between functional groups of the protein side chains and the bases in major groove of the DNA chain. Here the DNA chain contains inhomogeneous base sequences, which would allow the specific binding. Any environmental effects such as the solution is neglected at this stage.

The notion of signal transmission came from the results of experiments in which the action at a long distance in DNA were studied. The signal comes from regulatory protein and transmitted at high efficiency through the DNA-protein binding process. Many alternative models of the nonlocal action have been proposed [9–11]. We prefer the assumption that the protein molecule induces a local conformational distortion of base pairs causing a breather soliton excitation. This conformal alteration of DNA referred as the transmission of allosteric effects through DNA. The allosteric effects could be carried out by some kind of solitary waves, the possibility of these nonlinear effects to occur in DNA was first scrutinized by Englander et al. [12], and some of the later notable models are given by Peyrard-Bishop [13] and Yakushevich [14]. The dynamics and thermal effects of PB breather in presence of external potential are investigated [15, 16]. Regulatory proteins would increase the Peyrard-Bishop (PB) breather solitons amplitude as shown by Satarić and Tuszyński [17] via a rigorous calculation in nonequilibrium statistical physics. In this paper we show that our model reveal the possibility of PB breather excitation triggered by the binding of regulatory protein by simulations.

This paper presents as follow: in the second section we propose the model Hamiltonian along with the derived equation of motions. In the third section we provide the stability analysis and the computed phase portraits of the system to picture out the overall dynamics and a brief discussion the feasibility of this model in explaining the biological phenomenon. At last we conclude and give some future improvements for the preciseness of the dynamics.

### 2. Model description

Specific interactions of regulatory proteins with DNA can be determined through the hydrogen bonding between the protein side chains and groups of the bases at the DNA chain. Proteins are composed of many amino acids linked together through peptide bonds, hence the side chains interacting to DNA can be the amino acids or peptide bonds. Proteins interact mainly on the major groove of the DNA chain because the backbones are not hindering the way.

Here we model the interaction between one protein and a DNA chain. Our model includes two degrees of freedom, $x_m$, which correspond to the streching of hydrogen bonds between protein and a DNA base, and $y_n$, which correspond to the streching of hydrogen bonds in DNA that connect two bases in a pair. The model is illustrated in Figure 1, here we take the glutamine as an example of the side chain. The indices $n$ and $m$ denotes the protein location in DNA chain and the location of a base pair consecutively. The potential of the hydrogen bonds are approximated by the Morse potential. This potential is not regarding how many bonds exist in
an interaction, the parameters of the potential will do. For simplicity we assume a harmonic coupling due to the stacking between neighboring base pairs with the same coupling constant $k$ along the strands, also we use a common mass $m$ for the base pairs and a common mass $M$ for the protein. Hence, the Hamiltonian for our model is

$$H = H_{\text{DNA}} + H_{\text{prot}} + H_{\text{int}},$$

where the first part contains the Morse potential and the stacking term of DNA in second and third term,

$$H_{\text{DNA}} = \sum_{n} \frac{p_{yn}^2}{2m} + D_z(e^{-\alpha y_n} - 1)^2 + \frac{k}{2}(y_n - y_{n-1})^2.$$

The base nucleotide mass $m$ and momentum $p_{yn}$ are homogenous as mentioned. The Morse potential well depth $D_z$ depends on the strength of the bonding between bases (we place the chain at $z$-direction) and $\alpha$ is the inverse width of the well. The constant $k$ is the harmonic coupling of the longitudinal springs connecting the base pairs. The second part contains the Morse potential between a protein and a nearest base pair of DNA,

$$H_{\text{prot}} = \frac{p_{zm}^2}{2M} + E(e^{-\beta x_m} - 1)^2,$$

while the interaction between those two bonds takes a rather general term,

$$H_{\text{int}} = \sum_{n} \chi \frac{a^a x_m}{2} y_n^{b} f_{mn}.$$

We take a constant the potential well depth $E$ as we assume this to be only protein dependent i.e. independent of DNA site, while $\beta$ is just the inverse width of the well. The coupling constant $\chi$ determine the sensitivity and the strength of the interaction, the value is not specified and could be set to fit the reality. Thus in this model the depths $D_z$ and $E$ define the specificity of DNA-protein interaction, the values are determined by experiments (e.g. in [18]). In this paper we use $f_{mn} = \delta_{mn}$, means the protein interact with a nearest base pair.

We take a restriction of the values $a$ and $b$ in (4) due to the biological requirements influenced by the oscillators initial conditions. We need the existence of an incoming protein to trigger an excitation of a pulse or a breather in DNA. In our model this requires an initial shift in $x_m$ to trigger a shift in $y_n$, or to alter the state. However an initial shift in DNA, $y_n$, should not alter
the state of a resting protein (i.e., \( x_m = 0 \)), as if there is no protein. Otherwise the protein would not steadily bind. These specific requirements involving the initial conditions result a special condition for integers: \( b = 1 \) and \( a > 1 \). This can be understood by considering the equations of motion derived from the Hamiltonian,

\[
\begin{align*}
m\ddot{y}_m &= 2\alpha D_z(e^{-2\alpha y_m} - e^{-\alpha y_m}) + k(y_{m+1} - 2y_m + y_{m-1}) - b\chi x_m^a y_m^{-1}, \\
M\ddot{x}_m &= 2\beta E(e^{-2\beta x_m} - e^{-\beta x_m}) - a\chi x_m^{a-1} y_m^b.
\end{align*}
\]

(5)

First, the equation involving \( \ddot{y}_m \) should contain no \( y_m \) in the interaction term as we want the DNA excitation is affected by the protein only (not by the DNA itself). Second, in the equation involving \( \ddot{x}_m \), an initial shift in DNA should not alter a resting protein, this will require \( x_m y_m \) itself in the interaction term. Hence, the conditions are satisfied with \( b = 1 \) and \( a > 1 \) respectively. Such considerations will result nice and plausible phase portraits. Henceforth, we use the interaction term at lowest power of \( x_m \), \( H_{\text{int}} = \frac{1}{2}\chi x_m^2 y_n \delta_{mn} \).

3. Stability analysis

To discuss the dynamics we need to take a stability analysis of the Hamiltonian system. A fine approach is given in [19, 20], where the Hamiltonian is treated as a Lyapunov function. The method is useful to determine the stability of equilibria in the system and thus, the trajectory can be predicted. All we have to do is find the local extrema and the curvature at these points, this is done by looking the first and second order of the multivariable Taylor expansion at the equilibrium \( (x_0, y_0) \). The stability analysis is calculated within the protein and a nearest base pair. Based on [20], we only need to consider the Morse potentials and the interaction term. Here the harmonic coupling is neglected because it is just longitudinally transferring an amplitude. Consider the potential function

\[
U(x, y) = D_z(e^{-\alpha y} - 1)^2 + E(e^{-\beta x} - 1)^2 + \frac{1}{2}\chi x^2 y,
\]

(6)

where we have dropped dummy indices for brevity. The local minima is in such way that \( \nabla U(x_0, y_0) = 0 \), we get

\[
\begin{align*}
2\alpha D_z(e^{-2\alpha y} - e^{-\alpha y}) - \chi x^2 &= 0, \\
2\beta E(e^{-2\beta x} - e^{-\beta x}) - \chi xy &= 0.
\end{align*}
\]

(7)

The equilibria are in the intersection of these two null lines (Eq. (7)). Next we determine if the equilibria are stable or unstable, to do this we consider the determinant of the Hessian matrix \( H \) which is in the second order of the expansion,

\[
\det H = 2\alpha^2 D_z \left( 2e^{-2\alpha y} - e^{-\alpha y} \right) \left[ 2\beta^2 E \left( 2e^{-2\beta x} - e^{-\beta x} \right) + \chi y \right] - (\chi x)^2.
\]

(8)

This will determine the stability of the equilibrium. The positive curvature, \( \det(H)_{(x_0, y_0)} > 0 \), indicates a stable equilibrium and \( \det(H)_{(x_0, y_0)} < 0 \) for an unstable equilibrium.

For the stability, it is not yet known how to get a generalized formula to get the equilibria for any arbitrary parameters \( D_z, E, \alpha, \beta \) and \( \chi \). However we still can obtain the equilibria by a numerical computation after the parameters of the system is reduced. To do this first we have to treat in continuum limit Eq. (5) and scale it to the form consisting two parameters. The continuum limit assumes that the DNA chain is sufficiently long, this will give a qualitative understanding of the breather formation of DNA. We have

\[
\begin{align*}
\frac{d^2 Y_m}{d\tau^2} &= D_Z(e^{-2Y_m} - e^{-Y_m}) - \frac{1}{2}Y_m^2, \\
\frac{d^2 X_m}{d\tau^2} &= E(e^{-2X_m} - e^{-X_m}) - X_m Y_m,
\end{align*}
\]

(9)
Figure 2. Phase portraits of the systems, (a) with no initial amplitude in $X_m$ and (b) with no initial amplitude in $Y_m$, (c) the trajectory of $X_m$ is leaving the periodic orbit, and (d) the $(X,Y)$ homoclinic orbit.

where $Y_m \equiv \alpha y_m$, $X_m \equiv \beta_m$, $D_z \equiv 2\beta D_z/\chi$, $\mathcal{E} \equiv 2\alpha \beta E/\chi$, and $\tau_1 \equiv t\sqrt{\alpha \chi/m} - z\sqrt{\alpha \chi/(\beta k)}$, $\tau_2 \equiv t\sqrt{\chi/(\alpha M)}$ is the coordinate changes. Here we renormalized the equations consisting only two parameters $D_Z$ and $\mathcal{E}$ because the potential depths $D_z$ and $E$ contribute significantly to the specificity.

We pick $D_Z = 5$ and $\mathcal{E} = 2$ as an example when the hydrogen bonds between the base pair are stronger than between the DNA-protein. Plotting the null lines (7) after adjusting the parameters, we get two equilibria: $(0, 0)$ and $(1.62, -0.19)$. The former is stable and the latter is unstable. The flow of the system tends to be periodic around the stable origin and when reaching the unstable saddle node it gets away (forming a large amplitude). Hence, we have a homoclinic orbit of the $(X,Y)$ plane as figured in Fig. 2d, any initial amplitude inside the closed region of the orbit is stable otherwise it is not.

The simulated phase portraits are in Fig. 2. Figure 2a simulates the trajectory of the system without the presence of initial amplitude $X_m$, it follows that the DNA mode do not alter the DNA-protein binding as we want. Figure 2b simulates the trajectory with an initial amplitude in $X_m$ and $Y_m$ is initially static, showing that the shift of DNA-protein binding trigger the of DNA base pair opening. This excitation reveals the possibility of breather modes in DNA triggered by the binding of protein. Figure 2c is when the protein momentum is sufficiently large it fluctuates and the trajectory goes far away, meaning the protein is leaving the DNA site while the DNA mode is . This process depicts the nonspecific binding of protein, proteins undergo the process of coming and leaving DNA sites until it finds a specific binding site which is at the thermodynamic minimum. The protein keeps oscillating along the process because here we neglect the dissipative aspect of the solution.

So far we have constructed a model for the specific DNA-protein binding, the results presented
here are only an early step toward the understanding of DNA-protein interaction. There are limitations such as the interaction distance. This model only plausible when the protein is sufficiently close to DNA, otherwise the trajectories would blow up. Also the protein never bind steadily without long-time oscillatory motion, to limit the oscillation the dissipative effect due to solution should be considered. Another problem is the protein keeps on triggering the DNA amplitude regardless the specific or nonspecific site. It is remain a conundrum on how the protein exactly excite the DNA opening, referred as DNA bubble, such as whether it is always excite the bubble in nonspecific sites. On the other hand, the stability analysis would result up to four equilibria when $D_Z$ and $E$ is sufficiently small, further investigation of these kind of stability would give interesting results. The analytical solution of the model is in progress.

Acknowledgements
The authors would like to thank Ministry of Research Technology and Higher Education of Republic of Indonesia for research funding Desentralisasi 2015. DD thanks the members of Theoretical Physics Laboratory ITB for hospitality.

References
[1] Wilson C, Zhan H, Swint-Kruse L and Matthews K 2007 Cellular and molecular life sciences 64 3–16
[2] McClure W R 1985 Annual review of biochemistry 54 171–204
[3] Rohs R, Jin X, West S M, Joshi R, Honig B and Mann R S 2010 Annual review of biochemistry 79 233
[4] Halford S E and Marko J F 2004 Nucleic acids research 32 3040–3052
[5] Von Hippel P H and Berg O G 1986 Proceedings of the National Academy of Sciences 83 1608–1612
[6] Williams M C and Maher J L 2010 Biophysics of DNA-protein interactions: from single molecules to biological systems (Springer Science & Business Media)
[7] Gerland U, Moroz J D and Hwa T 2002 Proceedings of the National Academy of Sciences 99 12015–12020
[8] Redding S and Greene E C 2013 Chemical physics letters 570 1–11
[9] Ptashne M 1985 Nature 322 697–701
[10] Wang J C and Giaever G N 1988 Science 240 300–304
[11] Hogan M, Dattagupta N and Crothers D 1979 Nature 278 521–524
[12] Englander S, Kallenbach N, Heeger A, Krumhansl J and Litwin S 1980 Proceedings of the National Academy of Sciences 77 7222–7226
[13] Peyrard M and Bishop A R 1989 Physical review letters 62 2755
[14] Yakushevich L 1989 Physics Letters A 136 413–417
[15] Sulaiman A, Zen F, Alatas H and Handoko L 2012 Physica D: Nonlinear Phenomena 241 1640–1647
[16] Sulaiman A, Zen F, Alatas H and Handoko L 2012 Physica Scripta 86 015802
[17] Satarić M and Tuszyński J 2002 Physical Review E 65 051901
[18] Gao Y, Devi-Prasad K and Prohofsky E 1984 The Journal of Chemical Physics 80 6291–6298
[19] Rink B W and Tuwankotta J M 2000 Stability in Hamiltonian Systems (Universiteit Utrecht. Department of Mathematics)
[20] Verhulst F 1996 Nonlinear differential equations and dynamical systems (Springer Science & Business Media)