Nonlinear dynamics of specific DNA-protein interactions

D Dwiputra[1], W Hidayat[1,2], R Khairani[1] and F P Zen[1,2]

[1] Departement of Physics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung Jl. Ganesha 10, Bandung 40132, Indonesia
[2] Indonesian Center of Theoretical and Mathematical Physics (ICTMP), Indonesia

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

Abstract. Interactions between DNA binding protein and specific base pairs of nucleic acid is critical for biological process. We propose a new model of DNA-protein interactions to depict the dynamics of specific DNA-protein interactions. Hydrogen bonds (H-bonds) are, among the other intermolecular interactions in DNA, the most distinctive in term of specificity of molecular bonds. As H-bonds account for specificity, we only consider the dynamics affected by H-bonds between DNA base pairs and H-bonds connecting protein side chains and DNA. The H-bonds are modelled by Morse potentials and coupling terms in the Hamiltonian of coupled oscillators resembling a coupling between planar DNA chain and a protein molecule. In this paper we give a perturbative approach as an attempt for a soliton solution. The solution is in the form of nonlinear travelling wave having the amplitudes satisfying coupled nonlinear Schrödinger equations and is interpreted as the mediator for nonlocal transmittance of biological information in DNA.

1. Introduction

The dynamics of deoxyribonucleic acid (DNA) and its interaction to proteins is one of the most fascinating conundrum in biophysics because it constitutes the fundamental basis of life. The information transfer between proteins in the process of DNA transcription and replication has to be highly efficient and lossless to maintain the macroscopic life around us over the ages. However, an intact dynamical modelling is difficult due to the complexity of the role of regulatory proteins in DNA such as RNA polymerases. Thus, here we are simplifying the problem by examining only the significant aspects of DNA-protein binding.

The idea of nonlinear solitonic excitations in DNA has established in the past decades [1–3], the notion of the soliton is due to the ability of DNA to maintain a lossless transfer of information along vast distances—which cannot be satisfied with simple linear waves. Of interest here is the breather type of excitation which correspond to DNA denaturation bubbles by Peyrard and Bishop (PB) [2], they developed a statistical approach of the local base pair opening which precede the denaturation process. Later the dynamics and thermal effects of PB breather in presence of external potential are investigated [4, 5], giving the picture on how DNA and external perturbations behave. Regulatory protein would increase the breather amplitude as shown in [6] via nonequilibrium statistical calculations. A further effect of the oscillatory mode of the protein itself is examined on a thermal bath by [7].

Origins of the specificity in DNA-protein recognition rely on the unique chemical signatures of the base sequences and the sequence-independent DNA shapes [8–10]. The proteins can "feel" the DNA surface shapes and simultaneously recognize the sequences in virtue of electrostatic, van der Waals',

and H-bond interactions. However, 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 $k_{\mathrm{B}}T$ below the specific binding energy [11] (experimentally, for instance, 17 $k_{\mathrm{B}}T$ for *Mnt* and ≈16 $k_{\mathrm{B}}T$ for *lac* repressor [12]). 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 $k_{\mathrm{B}}T$ [11].

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 the 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 presence of the solution are neglected at this stage.

2. The dynamical model

Our aim is to examine the dynamics of the binding between regulatory proteins and DNA through the mathematical solution and its properties in agreement with biological observations [13, 14]. Here we would like to focus our attention on the major correspondent of the specific DNA-protein interaction, namely the H-bonds. These bonds are linking between protein side chains and groups of bases inside DNA chain and are responsible for the binding specificity. Proteins are composed of many amino acids which linked together by peptide bonds, practically the side chains which interact with DNA are the amino acids or the peptide bonds. This interaction occurs mainly in DNA major groove and affecting the H-bonds inside DNA nucleotides.

In this paper we model the effect of existing protein to the DNA chain. We conjecture that the existence of regulatory proteins near DNA trigger pulse type excitations of the DNA base pairs, this kind of excitations referred as DNA *bubbles* and are well described by the PB solitonic breathers. These propagating solitons carry important genetic information through the DNA and provide an effective communication between two separated proteins binding DNA in vast distances. Our model includes two degrees of freedom, x_m corresponds to the stretching of H-bonds linking a protein side chain and a side chain, and y_n is the stretching of H-bonds that connect two DNA bases in a pair. The indices are labelling locations in DNA. We simplify the complex DNA-protein to a coupled nonlinear oscillator. Nonlinearity arises while we consider the Morse potential to model a group of H-bonds, the potential permits breaking of the bonds when largely stretched. The stacking between base pairs is assumed harmonic at this stage. Thus, our model Hamiltonian is

$$H = H_{\mathrm{DNA}} + H_{\mathrm{prot}} + H_{\mathrm{int}} \tag{1}$$

where

$$H_{\mathrm{DNA}} = \sum_n \frac{p_{yn}^2}{2m} + D_{(\mathbf{Z})}(\mathrm{e}^{-\alpha y_n} - 1)^2 + \frac{k}{2}(y_n - y_{n-1})^2 . \tag{2}$$

The base nucleotide mass \$m\$ and momentum p_{yn} are homogenous. The Morse potential depth $D_{(\mathbf{Z})}$ is the strength of the bonding between bases (the chain is at **Z**-direction) and α is the inverse width of the well, k is the harmonic coupling of the longitudinal springs connecting the base pairs. H_{prot} contains the Morse potential between a protein and a nearest base pair of DNA,

$$H_{\mathrm{prot}} = \frac{p_{xm}^2}{2M} + E\left(\mathrm{e}^{-\beta x_m} - 1\right)^2, \tag{3}$$

the model is ready to be applied in the case of many proteins, but here we only consider one. Finally, the interaction between these two bonds takes a simple but rather general form,

$$H_{\mathrm{int}} = \sum_n \frac{\chi}{2} x_m^a y_n^b. \tag{4}$$

Here we take a constant potential well depth E as we assume this to be only protein dependent i.e. independent of DNA site; while β is just the inverse width of the well. The coupling constant χ determine the sensitivity and the strength of the interaction, the value is not specified and could be set to fit the reality. In this model the depths $D_{(\mathbf{Z})}$ and E define the specificity of DNA-protein interaction, the values could be determined via experiments (such as in [15]).

To prove that the model is plausible, the qualitative dynamics need to be depicted. This had been done by studying this model stability and phase portraits [16]. The stability was examined with the integer powers $a = 2$ and $b = 1$, the value was taken by assuming the one-way DNA interaction i.e. DNA base pair stretch (y_n) is coupled by protein only (via x_m^2), not by DNA itself. This can be understood by deriving the equations of motion from the Hamiltonian, but this time we will show that we can pick a and b without the assumptions. This will be made clear after we expand the equations for small x and y in the next section.

3. Nonlinear excitations in the DNA-protein interaction

This section will discuss on *how* the interaction occurs, that is, how the protein induces DNA bubbles that precede the total opening of the double helix (i.e. the denaturation process). To explore this aspect, we examine the nonlinear excitations related to the Hamiltonian system.

First we consider the equations of motion derived from the Hamiltonian,

$$\begin{aligned} m\ddot{y}_l &= 2\alpha D_{(\mathbf{Z})}(\mathrm{e}^{-2\alpha y_l} - e^{-\alpha y_l}) + k(y_{l+1} - 2y_l + y_{l-1}) - \tfrac{b}{2}\chi x_m^a y_l^{b-1}\,, \\ M\ddot{x}_m &= 2\beta E(\mathrm{e}^{-2\beta x_m} - e^{-\beta x_m}) - \tfrac{a}{2}\chi x_m^{a-1}\textstyle\sum_l y_l^b\,. \end{aligned} \tag{5}$$

According to the PB approach [17] it is assumed that the oscillations of bases are large enough to be anharmonic, but still insufficient to break the H-bond since the Morse potential plateau is not yet reached. In our model the protein-DNA bond x is also presumed by the same way. Here the base stretching y and the DNA-protein stretching x oscillate around the bottom of the symmetric potential so the following transformation can be safely implemented:

$$Y_n = \epsilon\phi_n \quad \text{and} \quad X_m = \epsilon\psi_m\,, \tag{6}$$

where ϵ is small and $Y \equiv \alpha y,\ X \equiv \beta x$. The continuum approximation is used for now as a first step to construct a solution. We use perturbative approach to obtain a solitonic solution. Inserting this into (5) with continuum limit and expanding in basis ϵ^n, we get (with z_0 replacing m)

$$\begin{aligned} &\phi_{tt} - S\phi_{zz} + V_{(\mathbf{Z})}\left(\phi - \tfrac{3}{2}\epsilon\phi^2 + \tfrac{7}{6}\epsilon^2\phi^3 + O(\epsilon^3)\right) + \tfrac{\mu}{2}\epsilon^{a+b-2}\psi^a|_{(\mathbf{Z}=\mathbf{Z_0})}\phi^{b-1} = 0 \\ (\text{at } \mathbf{Z} = \mathbf{Z}_0) \qquad &\psi_{tt} + W\left(\psi - \tfrac{3}{2}\epsilon\psi^2 + \tfrac{7}{6}\epsilon^2\psi^3 + O(\epsilon^3)\right) + \eta\epsilon^{a+b-2}\psi^{a-1}\textstyle\int \phi^b \mathrm{d}\mathbf{Z} = 0 \end{aligned} \tag{7}$$

where we have dropped the indices for simplicity and defined

$$V_{(\mathbf{Z})} = \tfrac{2\alpha^2 D_{(\mathbf{Z})}}{m},\ W = \tfrac{2\beta^2 E}{M},\ S = \tfrac{K}{m},\ \mu = \chi\tfrac{b\alpha^{2-b}}{2m\beta^a},\ \eta = \chi\tfrac{a\beta^{2-a}}{2M\alpha^b} \tag{8}$$

From here we look straightforward perturbative solution

$$\phi = \phi^{(0)} + \epsilon\phi^{(1)} + \cdots, \quad \psi = \psi^{(0)} + \epsilon\psi^{(1)} + \cdots. \tag{9}$$

The solutions are examined only up to order ϵ to get the nonlinear excitations consisting of second harmonics, which in case of PB model are the breather solitons. For this to be realised, every order of ϵ^n in above equations should contain time or spatial derivatives, thus we use multiple-scale method [18]. Here ψ and ϕ depend on the independent variables $z_0,\ z_1, \ldots$ and $t_0,\ t_1, \ldots$ where $z_m = \epsilon^m z$ and $t_n = \epsilon^n t$ so that the operators $\partial/\partial z$ and $\partial/\partial t$ are expanded as

$$\frac{\partial}{\partial \mathbf{Z}} = \frac{\partial}{\partial z_0} + \epsilon \frac{\partial}{\partial z_1} + \epsilon^2 \frac{\partial}{\partial z_2} + \cdots, \quad \frac{\partial}{\partial t} = \frac{\partial}{\partial t_0} + \epsilon \frac{\partial}{\partial t_1} + \epsilon^2 \frac{\partial}{\partial t_2} + \cdots. \tag{10}$$

This method is actually looking for solutions that variate (oscillate) in independent time and spatial scales, it has the advantage that the solution is accurate up to $z = O(1/\epsilon^m)$ and $t = O(1/\epsilon^n)$. For instance, the term $f(t_1) \propto \exp(i\theta(t_1))$ oscillates slower than the term dependent on t_0. In our model the t_1 term acts as the envelope wave for the t_0 term. This kind of behavior had previously been inspected in [19] for single nonlinear oscillator, it is shown that the multiple-scale method gives an envelope wave which the first harmonic amplitude satisfies nonlinear Schrödinger equation (NLS). In this paper we are generalising the method for a coupled equations.

From equations (5) we chose the integer powers $a + b = 3$ so that the interaction plays the role in $O(\epsilon)$ terms and is not ruining $\phi^{(0)}$ or $\psi^{(0)}$. It can be chosen in general $a + b > 2$, depending on what order ϵ the interaction is wanted to affect. Here we the lowest power state $H_{\text{int}} = \frac{1}{2}\chi x_m^2 y_n$. We use multiple scales up to t_2 and z_1, the asymmetry of this choice will be clear after the equations governing the amplitudes are observed. For simplicity hereafter, we write $z \equiv z_0$, $Z \equiv z_1$, $t \equiv t_0$, $T \equiv t_1$, and $\tau \equiv t_2$. Inserting (9) and (10) into (7), we get

$$\begin{aligned} O(\epsilon^0): \quad \phi_{tt}^{(0)} - S\phi_{zz}^{(0)} + V_{(z)}\phi^{(0)} &= 0 \\ \psi_{tt}^{(0)} + W\psi^{(0)} &= 0 \end{aligned} \tag{11}$$

$$\begin{aligned} O(\epsilon^1): \quad \phi_{tt}^{(1)} - S\phi_{zz}^{(1)} + V_{(z)}\phi^{(1)} &= -2\left(\phi_{tT}^{(0)} - S\phi_{zZ}^{(0)}\right) + \frac{3}{2}V_{(z)}\left(\phi^{(0)}\right)^2 - \frac{\mu}{2}\left(\phi^{(0)}\right)^2 \\ \psi_{tt}^{(1)} + W\psi^{(1)} &= -\psi_{tT}^{(0)} + \frac{3}{2}W\left(\psi^{(0)}\right)^2 - \eta\psi^{(0)}\int\int \phi^{((0))}\mathrm{d}z\mathrm{d}Z \end{aligned} \tag{12}$$

$$\begin{aligned} O(\epsilon^2): \quad \phi_{tt}^{(2)} - S\phi_{zz}^{(2)} + V_{(z)}\phi^{(2)} &= -2\left(\phi_{tT}^{(1)} - S\phi_{zZ}^{(1)}\right) - \left(\phi_{TT}^{(0)} - S\phi_{ZZ}^{(0)}\right) - 2\phi_{t\tau}^{(0)} + 3V_{(z)}\phi^{(0)}\phi^{(1)} \\ &\quad -\frac{7}{6}V_{(z)}\left(\phi^{(0)}\right)^3 - \mu\psi^{(0)}\psi^{(1)} \\ \psi_{tt}^{(2)} + W\psi^{(2)} &= -2\psi_{tT}^{(1)} - \psi_{TT}^{(0)} - 2\psi_{t\tau}^{(0)} + 3W\psi^{(0)}\psi^{(1)} - \frac{7}{6}W\left(\psi^{(0)}\right)^3 \\ &\quad -\eta\left(\psi^{(1)}\int\int \phi^{(0)}\mathrm{d}z\mathrm{d}Z + \psi^{(0)}\int\int \phi^{(1)}dzdZ\right) \end{aligned} \tag{13}$$

We have broken down the nonlinear problem into several inhomogeneous linear differential equations where the solutions as follows,

$$\begin{aligned} \phi^{(0)} &= A_1(Z,T,\tau)\mathrm{e}^{\mathrm{i}\theta} + \text{c. c.}, \qquad \psi^{(0)} = A_2(Z,T,\tau,z)\mathrm{e}^{\mathrm{i}\varphi} + c.c., \\ \phi^{(1)} &= 3|A_1|^2 - \frac{\mu}{V_{(z)}}|A_2|^2_{z=z_0} - \frac{1}{2}A_1^2\mathrm{e}^{2\mathrm{i}\theta} + \frac{\mu}{6V_{(z)}}A_2|^2_{z=z_0}\mathrm{e}^{2\mathrm{i}\varphi} + \text{c. c.}, \\ \psi^{(1)} &= 3|A_2|^2 - \frac{1}{2}A_2^2\mathrm{e}^{2\mathrm{i}\varphi} + \text{c. c.} + \eta\left[\frac{A_2\int A_1\mathrm{d}Z}{\omega^2+2\omega\sqrt{W}}\mathrm{e}^{\mathrm{i}(\theta+\varphi)} + \frac{A_2^*\int A_1 dZ}{\omega^2-2\omega\sqrt{W}}e^{i(\theta-\varphi)} + \text{c. c.}\right]\delta(k). \end{aligned}$$

The envelopes A_1 and A_2, by zeroing the secular terms ($\exp(\pm\mathrm{i}\theta)$ and $\exp(\pm\mathrm{i}\varphi)$) in $O(\epsilon^2)$, meet these conditions:

$$\begin{aligned} \mathrm{i}\frac{\partial A_1}{\partial \tau} + P_1\frac{\partial^2 A_1}{\partial \xi^2} + Q_1|A_1|^2A_1 &= \mu\left[3 + \frac{2\eta}{\omega^2-4W}\right]|A_2|^2A_1, \\ \mathrm{i}\frac{\partial A_2}{\partial \tau} + P_2\frac{\partial^2 A_2}{\partial T^2} + Q_2|A_2|^2A_2 &= \eta\left[3 + \frac{2\eta}{\omega^2-4W}\right]\int |A_1|^2dZ\, A_2, \end{aligned} \tag{14}$$

after we introduce a right moving coordinate where $\xi = Z - v_g T$. Here if we use scale $\epsilon^2 z$ we have to introduce the coordinate moving to left, $\epsilon^2 z + v_g\tau$, to form the NLS. This is contrary to our right moving

coordinate so we cannot use this. The coefficients are $\$P_1 = (S - v_g)/2\omega$, $Q_1 = 4V_{(z)}$, $P_2 = -\left(2\sqrt{W}\right)^{-1}$, and $Q_2 = 4W + 5\mu\eta/(6V_{(z)})$. These are in the form of coupled nonlinear Schrödinger (NLS) equation and the solutions have been widely investigated [20].

We look the solution having the type of bright soliton that is under condition $P_i Q_i > 0$ [20]. The one-soliton solution of A_1 and A_2 when inserted to our full solution will enable a particular breathing mode, this is interpreted as the denaturation bubbles. Further analysis to the solution could bring us to a deeper understanding of the dynamical excitation of DNA bubbles when triggered by a regulatory protein in some specific regions. It is also an interesting question that whether the model can represent the lowering of denaturation free energy as seen from experiments [11, and references therein]. The complete solution and its analysis are in progress.

References

[1] Englander S, Kallenbach N, Heeger A, Krumhansl J and Litwin S 1980 *Proc. Nat. Acad. Sci.* **77** 7222–7226
[2] Peyrard M and Bishop A R 1989 *Physical review letters* **62** 2755
[3] Yakushevich L 1989 *Physics Letters A* **136** 413–417
[4] Sulaiman A, Zen F P, Alatas H and Handoko L T 2012 *Physica D: Nonlinear Phenomena* **241** 1640–1647
[5] Sulaiman A, Zen F P, Alatas H and Handoko L T 2012 *Physica Scripta* **86** 015802
[6] Satarić M and Tuszynski J 2002 *Physical Review E* **65** 051901
[7] Sulaiman A, Zen F P, Alatas H and Handoko L T 2010 *Physical Review E* **81** 061907
[8] Rohs R, Jin X, West S M, Joshi R, Honig B and Mann R S 2010 *Ann. Rev. of biochem.* **79** 233
[9] Halford S E and Marko J F 2004 *Nucleic acids research* **32** 3040–3052
[10] Von Hippel P H and Berg O G 1986 *Proc. Nat. Acad. Sci.* **83** 1608–1612
[11] Williams M C and Maher J L 2010 *Biophysics of DNA-protein interactions: from single molecules to biological systems* (Springer Science & Business Media)
[12] Gerland U, Moroz J D and Hwa T 2002 *Proc. Nat. Acad. Sci.* **99** 12015–12020
[13] Wang J C and Giaever G N 1988 *Science* **240** 300–304
[14] Hogan M, Dattagupta N and Crothers D 1979 *Nature* **278** 521–524
[15] Gao Y, Devi-Prasad K and Prohofsky E 1984 *The Journal of Chemical Physics* **80** 6291–6298
[16] Dwiputra D, Hidayat W, Khairani R and Zen F P 2015 Nonlinear model of specific DNA-protein interactions and its stability (unpublished), submitted at the 6th asian physics symposium
[17] Peyrard M 2004 *Nonlinearity* **17** R1
[18] Jakobsen P 2013 Introduction to the method of multiple scales *Preprint* arXiv:1312.3651
[19] Remoissenet M 1986 *Physical Review B* **33** 2386
[20] Radhakrishnan R and Lakshmanan M 1995 *J. Phys. A: Mathematical and General* **28** 2683

Acknowledgements

The authors would like to thank Ministry of Research Technology and Higher Education of Republic of Indonesia for research funding Desentralisasi 2015 and Desentralisasi 2016. DD thanks the members of Theoretical Physics Laboratory ITB for hospitality.