Frontiers in Bioscience Landmark

Front. Biosci. (Landmark Ed) **2024**; 29(4): 155
https://doi.org/10.31083/j.fbl2904155

Original Research

Dynamics of Open States and Promoter Functioning in the *app*Y_red and *app*Y_green Genetic Constructions Based on the pPF1 Plasmid

Irina Masulis[1,†], Andrey Grinevich[1,†], Ludmila Yakushevich[1,*]

[1]Institute of Cell Biophysics of the Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia
*Correspondence: kind-@mail.ru (Ludmila Yakushevich)
[†]These authors contributed equally.

Academic Editor: Ilaria Baglivo
Submitted: 12 December 2023 Revised: 26 February 2024 Accepted: 29 February 2024 Published: 19 April 2024

Abstract

Background: Although the role of dynamic factors in DNA function still remains unclear, research in this direction is a rapidly developing area of molecular biology. In this work, the genetic constructions *app*Y_red and *app*Y_green, based on the plasmid pPF1 and containing a fragment of *Escherichia coli* (*E. coli*) DNA with predicted promoter-like regions, are considered complex dynamic systems in which local sites of double helix unwinding, called open states, can arise and propagate. The purpose of the article is to show the existence of a connection between the dynamics of open states and the functioning of predicted promoters. **Methods**: We experimentally verified the functionality of the predicted promoters using a reporter vector. Using a reverse transcription reaction, transcription start sites were identified indicating the presence of two divergent promoters, one on each strand. In mathematical studies, a dynamic model was used that described open states as one-soliton solutions (kinks) of a system of nonlinear partial differential equations, and the influence of the torque M_τ on the dynamics of kinks was taken into account. **Results**: Fluorescence analysis of colonies of *E. coli* cells transformed with plasmid constructions showed that one of the two promoters is stronger than the other and that the strength of the promoters depends on the orientation of the fragment under study in the plasmid. On the other hand, using mathematical modeling, the energy profiles of genetic constructions were calculated and the kink trajectories were constructed. In addition, by studying the effect of torsion moment in model studies, we found threshold torque values at which the behavior of kinks changes dramatically: from oscillatory to translational motion. The minimum values of torsion moment required to initiate the transcription process were also assessed. **Conclusions**: A comparative analysis of the results of experimental data and model calculations showed a good correlation between the preferred starting points and the direction of transcription, which in turn confirmed the existence of a relationship between the dynamics of open states and the functioning of promoters.

Keywords: open states; pPF1 plasmid; *E. coli app*Y gene; transcription; mathematical model

1. Introduction

Open states are formed in DNA as a result of local unwinding of the double-stranded helix and play an important role in the transcription process [1]. It is assumed that the dynamic properties of the DNA molecule, in general, and open states, in particular, can determine the basic regulation of transcription [2–5].

One of the DNA dynamic properties is the formation of local or global superspiralized states in the DNA molecule [6]. An important characteristic of superspiralized states is the torsion moment M_τ, which can be caused by various reasons, for example, by the interaction of RNA polymerase with DNA promoter regions [7]. RNA polymerase supercoils DNA while moving during transcription. The resulting torsion stress in DNA can then accumulate and become a barrier causing RNA polymerase to stop, and thus transcription to stop [8]. Direct measurements of the torsion moment performed on single *Escherichia coli* (*E. coli*) DNA showed that in the absence of any other regulators, RNA polymerase is able to generate a torque of ~11 piconewton nanometers (or 11×10^{-21} J) before stopping [9]. It can be assumed that this value determines the maximum value of the torsion moment, above which the transcription process stops. The question of what is the minimum (or threshold) value of the torsion moment required for the implementation of the transcription process has not yet been considered by anyone.

Transcription initiation occurs upon binding of RNA polymerase to DNA promoter regions and depends on the strength of the promoter [1]. It is known that promoter strength depends on the base sequence of the promoter region and can affect transcription-coupled DNA supercoiling [10,11]. However, the relationship between the dynamics of open states in DNA and the strength of promoters, which determines the efficiency of gene transcription and expression, has been poorly studied. This work is devoted to identifying this relationship by studying the dynamics of open states in the genetic constructions *app*Y_red and *app*Y_green (or Red and Green constructions for short) based on the pPF1 plasmid [12] with an integrated

fragment of *E. coli* DNA (positions from -65 to -252 relative to the beginning of the gene, taken as +1), taken from the regulatory region of the *app*Y gene. Interest in this fragment is caused by the fact that, according to computer prediction of promoters based on text analysis of known regulatory regions, this fragment contains multiple potential transcription starts [13].

One of the approaches to studying the dynamics of open states is mathematical modeling [2,5,14–16]. We have previously applied this approach to study the movement of transcription bubbles in the *app*Y_red and *app*Y_green genetic constructions and predicted the preferential direction of the bubble movement [17]. However, our modeling studies did not take into account the torsion moment M_τ and, in addition, there was no experimental confirmation of the predicted behavior of the kinks. In this work, when conducting model studies, we take into account the torsion moment and confirm the predictions with experimental data.

Thus, the main goal of the article, in addition to checking the functionality of the predicted promoters in the studied *E. coli* DNA fragment, is to identify the relationship between the promoters functioning and the dynamics of open states, taking into account the influence of torsion moment.

2. Materials and Methods

2.1 Cloning of the Genomic Regions of Interest

Polymerase chain reaction (PCR)-generated DNA fragment of *app*Y upstream regulatory region (from position -260 up to position -60, respectively, to *app*Y ATG codon) was inserted into the reporter vector pPF1 at the BglII restriction site. Oligonucleotides used for amplification:

*app*Y_For _BglII 5′ GCAAGATCTTTCAGGTGCGTTGTAGTGAG

*app*Y_Rev _BglII 5′ ACAAGATCTTAATCAGGATGATGTGCAT

PCR products were purified in 4.5% Polyacrylamide Gel (PAAG), pPF1 vector, and DNA fragments were subjected to restriction using endonuclease BglII (NEB, Ipswich, MA, USA), separated in 1% agarose and purified using Cleanup Standard kit (Evrogen, Moscow, Russia). T4 DNA ligase (NEB) was used for ligation, the reaction was carried out according to manufacturer protocol (0.1 pmol of DNA fragment and 0.05 pmol of linearized plasmid were taken into reaction in total volume 20 µL). A ligation mixture was used for the transformation of competent *E.coli* K12 MG1655 cells prepared as described in [18]. Bacterial colonies after transformation were selected by fluorescence of *Egfp* and *mCherry*. Primers used for PCR assumed the probability of insertion in opposite orientations. The orientation and quality of inserts were confirmed by sequencing the resulting plasmids using primers specific for *mCherry* and *Egfp* genes (*mCherry* 5′ AGCGCATGAACTCCTTGATG) and *Egfp* 5′ GTCCAGCTCGACCAGGATG). Plasmids corresponding to constructions *app*Y_Red (Red) and *app*Y_Green (Green) (Fig. 1) with confirmed sequence were used for transformation of *E.coli K12* MG1655, purified from cultures using Plasmid Miniprep kit (Evrogen) and stored for subsequent experiments (transformation of *E.coli* K12 MG1655 for transcription starts mapping and colonies fluorescence assay).

2.2 Primer Extension Assay

To determine transcription start sites within *app*Y regulatory region analyzed in this study primer extension reaction was carried out. *E.coli* K12 MG1655 cells transformed by the derivatives of pPF1 plasmid were grown on Lysogeny broth (LB) medium in the presence of kanamycin (40 µg/mL) and harvested at OD_{600} = 0.6. Total RNA was isolated with an ExtractRNA kit (Evrogen). RNA concentration was measured spectrophotometrically using NanoDrop1000 (Thermo Scientific, Wilmington, DE, USA). RNA integrity was estimated in denaturating 1.5% PAAG in the presence of 8M urea. ^{32}P-labeled oligonucleotides were prepared using T4 polynucleotide kinase (NEB) according to manufacturer protocol. A total of 20 µg of RNA and 4 pmol of ^{32}P-labeled oligonucleotide primer were taken for the reaction of reverse transcription with MuLV Reverse Transcriptase (Evrogen). The reaction was performed at 43 °C in conditions specified by the manufacturer. cDNA was precipitated by a 10-fold volume of N-butanol. Reaction products were separated in 6% PAAG and 8M urea and exposed with X-Ray Retina film. Guanine-specific hydrolysis of PCR fragments obtained with oligonucleotide pairs *mCherry**/*Egfp* and *mCherry*/*Eegfp** using Green plasmid as a template was carried out according to [18]. For the construction of plasmid Red, oligonucleotides *mCherry** (5′ AGCGCATGAACTCCTTGATG) and *Egfp** (5′ GTCCAGCTCGACCAGGATG) were used for mapping of Transcription Start Site (TSS) in sense and antisense directions, respectively. For the construction of plasmid Green, *Egfp** and *mCherry** radiolabeled primers respectively reflect direct and antisense transcription.

2.3 Image Acquisition and Processing of Cell Colonies

Transcriptional activity was assessed by the level of fluorescence of transformant colonies when grown on LB agar medium in the presence of 50 µg/mL kanamycin for 14 hours. Fluorescence of *E. coli* cell colonies transformed with plasmid constructions was recorded using a Leica DM 6000B microscope (lot number 289896, Leica Microsystems, Wetzlar,

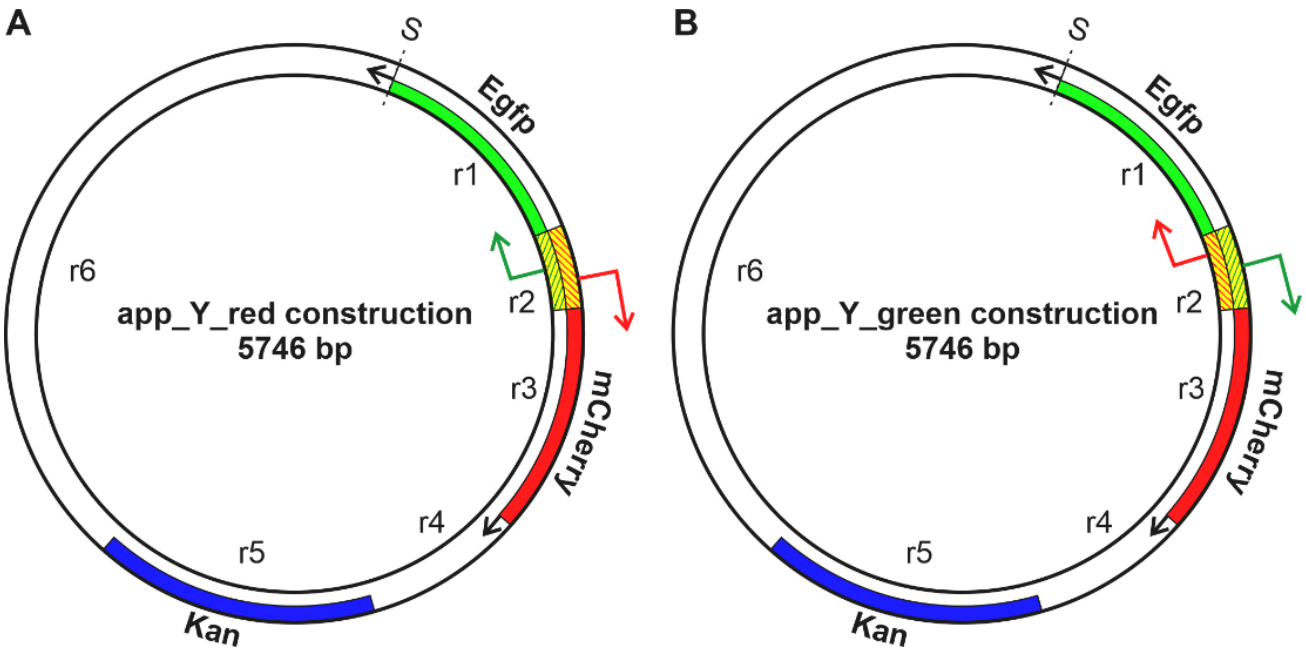

Fig. 1. Schematic representation of two genetic constructions. In the Red (A) and Green (B) constructions, the *mCherry* gene is shown in red, the *Egfp* gene in green, and the Kanamycin gene in blue. A fragment of *Escherichia coli* (*E. coli*) DNA located between the *mCherry* and the *Egfp* genes is shaded. Point S indicates the beginning of the numbering of the nucleotide sequence.

Germany). Images were obtained in the "Fluorescence" mode, and the choice of filters for emission excitation and radiation detection was set at settings "Cubes": "GFP" (excitation in the region 488 $\pm$ 25 nm, emission 500–550 nm) and "N21" (excitation in the region of 515–550 nm, emission in the region of 580–590 nm).

Image processing was carried out using the ImageJ program (ver. 153t, National Institutes of Health, Bethesda, MD, USA) [19]. To perform this, we used the algorithm of threshold selection of cell colonies (whole spots) in the image, followed by the calculation of the average intensity of pixels in the selected areas. Sample sizes were 16 spots for *Egfp* protein and 15 spots for *mCherry*. To correctly estimate the amount of synthesized fluorescent proteins, we took into account the difference in the brightness of the glow of red and green proteins. Thus, according to the data of [20], the luminescence brightness of *Egfp* proteins is 33.54, and that of *mCherry* proteins is 15.84. Then the intensity of fluorescent luminescence can be expressed through the number (N) of proteins and their brightness (α) as follows:

$$I_i = \alpha_i N_i, i = Egfp,\ mCherry. \tag{1}$$

If we go to the new variables describing the amount of proteins in brightness units as follows: $N_i^{'} = \alpha_{mCherry} N_i$, then for the new variables, the following is true:

$$\begin{aligned} N^{'}_{Egfp} &= \frac{\alpha_{mCherry}}{\alpha_{Egfp}} I_{Egfp}, \\ N^{'}_{mCherry} &= I_{mCherry}, \end{aligned} \tag{2}$$

where $\alpha_{Egfp} = 33.54$ and $\alpha_{mCherry} = 15.84$. Such variables allow us to correctly estimate the number of fluorescent proteins on the scale of the intensity of the glow of *mCherry* proteins in colonies of transformed cells.

2.4 Mathematical Model

When conducting model studies, we took into account that the Red and Green constructions contain three genes separated by three intermediate regions. There are six regions in total. To model the internal mobility of each of these regions, a modified Englander model [16] that was based on the assumption that angular displacements of the nitrous bases make the main contribution to the formation of DNA open states was used. The model neglects the contribution of other internal movements involved in the formation of open states, such as transverse and longitudinal movements of nucleotides.

Unlike the model originally proposed by Englander [21], the modified model took into account the angular deviations of the bases in both the coding and complementary chains, and also took into account the effects of dissipation and the action of constant torsion moment M_τ:

$$\begin{aligned} & I_{n,1}\frac{d^2\varphi_{n,1}(t)}{dt^2} - K'_{n,1}[\varphi_{n+1,1}(t) - 2\varphi_{n,1}(t) + \varphi_{n-1,1}(t)] \\ & + k_{n,1-2}R_{n,1}(R_{n,1} + R_{n,2})\ \sin\varphi_{n,1}(t) - k_{n,1-2}R_{n,1}R_{n,2}\sin(\varphi_{n,1}(t) - \varphi_{n,2}(t)) \\ & = -\beta_{n,1}\frac{d\varphi_{n,1}(t)}{dt} + M_\tau, \end{aligned} \tag{3}$$

$$\begin{aligned} & I_{n,2}\frac{d^2\varphi_{n,2}(t)}{dt^2} - K'_{n,2}[\varphi_{n+1,2}(t) - 2\varphi_{n,2}(t) + \varphi_{n-1,2}(t)] \\ & + k_{n,1-2}R_{n,2}(R_{n,1} + R_{n,2})\sin\varphi_{n,2}(t) - k_{n,1-2}R_{n,1}R_{n,2}\sin(\varphi_{n,2}(t) - \varphi_{n,1}(t)) \\ & = -\beta_{n,2}\frac{d\varphi_{n,2}(t)}{dt} + M_\tau, n = 1, 2, \ldots N. \end{aligned} \tag{4}$$

Here $\varphi_{n,i}(t)$ is the angular deviation of the *n*-th nitrogenous base of the *i*-th chain; $I_{n,i}$ is the moment of inertia of the base; $R_{n,i}$ is the distance from the center of mass of this base to the sugar-phosphate backbone; $K'_{n,i} = KR^2_{n,i}$; K is a constant characterizing the stiffness (tensile strength) of the sugar-phosphate backbone; $\beta_{n,i} = \alpha R^2_{n,i}$; α is the dissipation coefficient; $k_{n,1-2}$is a constant characterizing the interaction between complementary bases within the *n*-th pair; M_τ is the constant torsion moment; i = 1, 2; n =1, 2, …N; N is the total number of pairs of nitrogenous bases. It is assumed that the length of the modeled region is sufficiently large, and the effects at the ends of the region can be neglected.

To simplify Eqns. 3,4, the coefficients of the equations were averaged inside the region under consideration:

$$\begin{aligned} & \bar{I}_1\frac{d^2\varphi_{n,1}(t)}{dt^2} - \bar{K}'_1[\varphi_{n+1,1}(t) - 2\varphi_{n,1}(t) + \varphi_{n-1,1}(t)] \\ & + \bar{k}_{1-2}\bar{R}_1(\bar{R}_1 + \bar{R}_2)\ \sin\varphi_{n,1} - \bar{k}_{1-2}\bar{R}_1\bar{R}_2\sin(\varphi_{n,1}(t) - \varphi_{n,2}(t)) \\ & = -\bar{\beta}_1\frac{d\varphi_{n,1}(t)}{dt} + M_\tau, \end{aligned} \tag{5}$$

$$\begin{aligned} & \bar{I}_2\frac{d^2\varphi_{n,2}(t)}{dt^2} - \bar{K}'_2[\varphi_{n+1,2}(t) - 2\varphi_{n,2}(t) + \varphi_{n-1,2}(t)] \\ & + \bar{k}_{1-2}\bar{R}_2(\bar{R}_1 + \bar{R}_2)\ \sin\varphi_{n,2}(t) - \bar{k}_{1-2}\bar{R}_1\bar{R}_2\sin(\varphi_{n,2}(t) - \varphi_{n,1}(t)) \\ & = -\bar{\beta}_2\frac{d\varphi_{n,2}(t)}{dt} + M_\tau, \end{aligned} \tag{6}$$

where

$$\begin{aligned} \bar{I}_i &= I_A C_{A,i} + I_T C_{T,i} + I_G C_{G,i} + I_C C_{C,i}, \\ \bar{R}_i &= R_A C_{A,i} + R_T C_{T,i} + R_G C_{G,i} + R_C C_{C,i}, \\ \bar{K}'_i &= K'_A C_{A,i} + K'_T C_{T,i} + K'_G C_{G,i} + K'_C C_{C,i}, \\ \bar{k}_{1-2} &= k_{A-T}(C_{A,1} + C_{T,2}) + k_{G-C}(C_{G,1} + C_{C,2}), \\ \bar{\beta}_i &= \beta_A C_{A,i} + \beta_T C_{T,i} + \beta_G C_{G,i} + \beta_C C_{C,i}. \end{aligned} \tag{7}$$

Here $C_{j,i} = N_{j,i}/N$ is the concentration of bases of the *j*-th type in the considered region of the *i*-th chain; $N_{j,i}$ the number of bases of the *j*-th type in this region; N is the total number of bases in the region; j = A, T, G, C, i = 1, 2.

Assuming that the solutions of Eqns. 5,6 are sufficiently smooth functions, we rewrote Eqns. 5,6 in the continuum approximation as follows:

$$\bar{I}_1\varphi_{1,tt} - \bar{K}_1^{'}a^2\varphi_{1,zz} + \bar{k}_{1-2}\bar{R}_1\left(\bar{R}_1 + \bar{R}_2\right)\sin\varphi_1 - \bar{k}_{1-2}\bar{R}_1\bar{R}_2\sin(\varphi_{1,} - \varphi_2) = -\bar{\beta}_1\varphi_{1,t} + M_\tau, \tag{8}$$

$$\bar{I}_2\varphi_{2,tt} - \bar{K}_2^{'}a^2\varphi_{2,zz} + \bar{k}_{1-2}\bar{R}_1(\bar{R}_1 + \bar{R}_2)\ \sin\varphi_2 - \bar{k}_{1-2}\bar{R}_1\bar{R}_2\sin(\varphi_2 - \varphi_1) = -\bar{\beta}_2\varphi_{2,t} + M_\tau, \tag{9}$$

where *a* is the distance between the nearest base pairs.

Further, we took into account the features of the distribution of interactions within the DNA molecule: the presence of "weak" hydrogen bonds between nitrogenous bases within complementary pairs and "strong" valence interactions along the sugar-phosphate chains. This property of the DNA molecule made it possible, as a first approximation, to transform two coupled Eqns. 8,9 into two independent equations:

$$\bar{I}_1\varphi_{1,tt} - \bar{K}_1^{'}a^2\varphi_{1,zz} + \bar{k}_{1-2}\bar{R}_1^2\sin\varphi_1 = -\bar{\beta}_1\varphi_{1.t} + M_\tau, \tag{10}$$

$$\bar{I}_2\varphi_{2.tt} - \bar{K}_2^{'}a^2\varphi_{2,.zz} + \bar{k}_{1-2}\bar{R}_2^2\sin\varphi_1 = -\bar{\beta}_2\varphi_{2,t} + M_\tau. \tag{11}$$

Thus, the original problem was divided into two independent problems. The first was related to Eqn. 10, which models the angular deviations of nitrogenous bases in the coding chain. The second problem was with Eqn. 11, which described the angular deviations of nitrogenous bases in the complementary chain.

In a particular case, when the effects of dissipation and the action of a constant torsion moment are small ($\bar{\beta}_1 \cong 0$ and $M_\tau \cong 0$)), Eqns. 10,11 take the form of classical sine-Gordon equations:

$$\bar{I}_1\varphi_{1,tt} - \bar{K}_1^{'}a^2\varphi_{1,zz} + \bar{k}_{1-2}\bar{R}_1^2\sin\varphi_1 = 0, \tag{12}$$

$$\bar{I}_2\varphi_{2.tt} - \bar{K}_2^{'}a^2\varphi_{2,.zz} + \bar{k}_{1-2}\bar{R}_2^2\sin\varphi_1 = 0. \tag{13}$$

One-soliton solutions of these equations—kinks—found by analytical methods have the following form:

$$\varphi_{k,1}(z,t) = 4arctg\{\exp[(\bar{\gamma}_1/\bar{d}_1)(z - v_{k,1}t - z_{0,1})]\}, \tag{14}$$

$$\varphi_{k,2}(z,t) = 4arctg\{\exp[(\bar{\gamma}_2/\bar{d}_1)(z - v_{k,2}t - z_{0,2})]\}. \tag{15}$$

Here, $v_{k,i}$ is the kink velocity in the coding (i = 1) or complementary (i = 2) chains; $\bar{\gamma}_i = \left(1 - v_{k,i}^2/\bar{C}_i^2\right)^{-\frac{1}{2}}$ is the Lorentz factor; $\bar{C}_i = \left(\bar{K}_i' a^2/\bar{I}_i\right)^{1/2}$ is the sound velocity in the i-th chain; $\bar{V}_i = \bar{k}_{1-2}\bar{R}_i^2$; $\bar{d}_i = \left(\bar{K}_i' a^2/\bar{V}_i\right)^{1/2}$ is the size of the i-th kink; $z_{0,i}$ is the coordinate of the i-th kink at the initial moment of time; i is the chain number, i = 1, 2.

The total energy and rest energy of the i-th kink are determined by the following formulas:

$$\bar{E}_i = \bar{\gamma}_i\bar{E}_{0,i}, \tag{16}$$

$$\bar{E}_{0,i} = 8\sqrt{\bar{K}'_i \bar{V}_i}. \tag{17}$$

In the case of low velocities $\upsilon_{k,i} \ll \bar{C}_i$, Eqn. 16 can be rewritten as follows:

$$\bar{E}_i = 8\sqrt{\bar{K}'_i \bar{V}_{i,}} \cdot \bar{\gamma}_i \cong \bar{E}_{0,i} + \frac{\bar{m}_i \, \upsilon_{k,i}^2}{2}, \tag{18}$$

where $\bar{m}_i = \frac{\bar{E}_{0,i}}{2\bar{C}_i^2}$ is the mass of the *i*-th kink. The form of Eqn. 18 explains why DNA kinks are often considered quasiparticles with a certain mass, velocity, and energy.

Section 3.2.1 presents a preliminary analysis of the kink behavior based only on the analysis of the form of the energy profiles of the coding and complementary sequences of the Red or Green constructions.

In the general case ($\bar{\beta}_1 \neq 0,\ M_\tau \neq 0$), Eqns. 10,11 are solved by us numerically using the difference scheme described in [15]. The results obtained are presented in Section 3.2.2 in the form of kink trajectories.

3. Results

3.1 Experiment Studies

*tfa*X/*app*Y intergenic interval represents one of 78 genomic loci of *E.coli* chromosome oversaturated by potential promoters predicted using PlatProm software (http://www.mathcell.ru/model6.php?l=en) [13] that assumes specific structural properties of this region and its intrinsic capacity to respond to DNA conformational rearrangements including transcription-coupled changes in superhelical state. In our previous work, we found that two primary promoters mapped in the regulatory region between positions -260 and -61 with respect to *app*Y ATG are active [12], though only one orientation of the fragment of interest in the reporter vector was studied. Here, we compared transcription activity in two constructions representing the same genomic region placed in inverted positions relative to the same plasmid environment. In both genetic constructions obtained after the transformation of *E.coli* K12MG1655 cells, TSSs were mapped using primer extension reaction with oligonucleotides specific for *EGFP* and *mCherry* genes. Despite the orientation of the insert, the same TSSs are revealed in the direction of the *app*Y gene at the position -81 relative to the initiation codon, as well as in the opposite antisense direction at the position -148 (Fig. 2). Thus, the capacity of the insert to initiate bidirectional transcription remains conserved and provided by two promoters located back-to-back in the region under study.

The nucleotide sequences of the studied *E. coli* DNA fragments in the Red and Green constructions and reviled TSS are shown in Fig. 3A,B, respectively. It turned out that in the region under study, one main promoter on each of the strands works (Fig. 2). For the coding strand in Red and Green constructions, the start of RNA synthesis is at position -81 relative to the start of the *app*Y gene (promoter P1 in Fig. 3C). This prevalent promoter is characterized by almost perfect -10 and -35 consensus hexamers (TATAAA and TTGCAA), though separated by non-optimal spacer 22 bp in length. On the complementary strand, transcription is initiated from position -148 (promoter P2 in Fig. 3C), having the context of -10 and -35 regions TTTAAG and TTGCAA, respectively, and separated by 16 bp. Promoter P2 possessing fewer conservative hexamers -10 and -35 exhibits low activity as compared to P1 (Fig. 2). When the orientation of the studied fragment is changed (Green construction), the promoters maintain their location in the sequence but move from one strand to the other (Fig. 3D).

Transformant colonies and their fluorescence levels are shown in Fig. 4A,D. It can be seen from these figures that in cells transformed with both variants of the plasmid, synthesis of both the *Egfp* protein and *mCherry* is observed. This suggests that transcription initiation is possible both on the coding strand of the *app*Y gene and on its complementary strand. In the case of Red and Green, transcription is observed in both directions, but the transcriptional activity in these two cases is different. From diagrams in Fig. 4B,E, it can be seen that in the case of Red, the brightness of the luminescence of cells containing red proteins is significantly ($p < 0.001$) higher than the brightness of cells containing green proteins. In the case of Green, the opposite picture is observed. The brightness of the glow of red and green proteins is different [20]. Diagrams that take this difference into account are shown in Fig. 4C,F. From these diagrams, it can be seen that in the case of Red, more red protein is synthesized, and in the case of Green, more green.

From the diagrams presented in Fig. 4, it can be observed that in the case of Red construction, the height of the red column in the diagram shown in Fig. 4C is significantly ($p < 0.001$) higher than the height of the green column. In the case of Green construction, the height of the green column shown in Fig. 4F is significantly ($p < 0.003$) higher than the height of the red column. This indicates that in both cases, the P1 promoter is more active than the P2 promoter. In addition, comparing

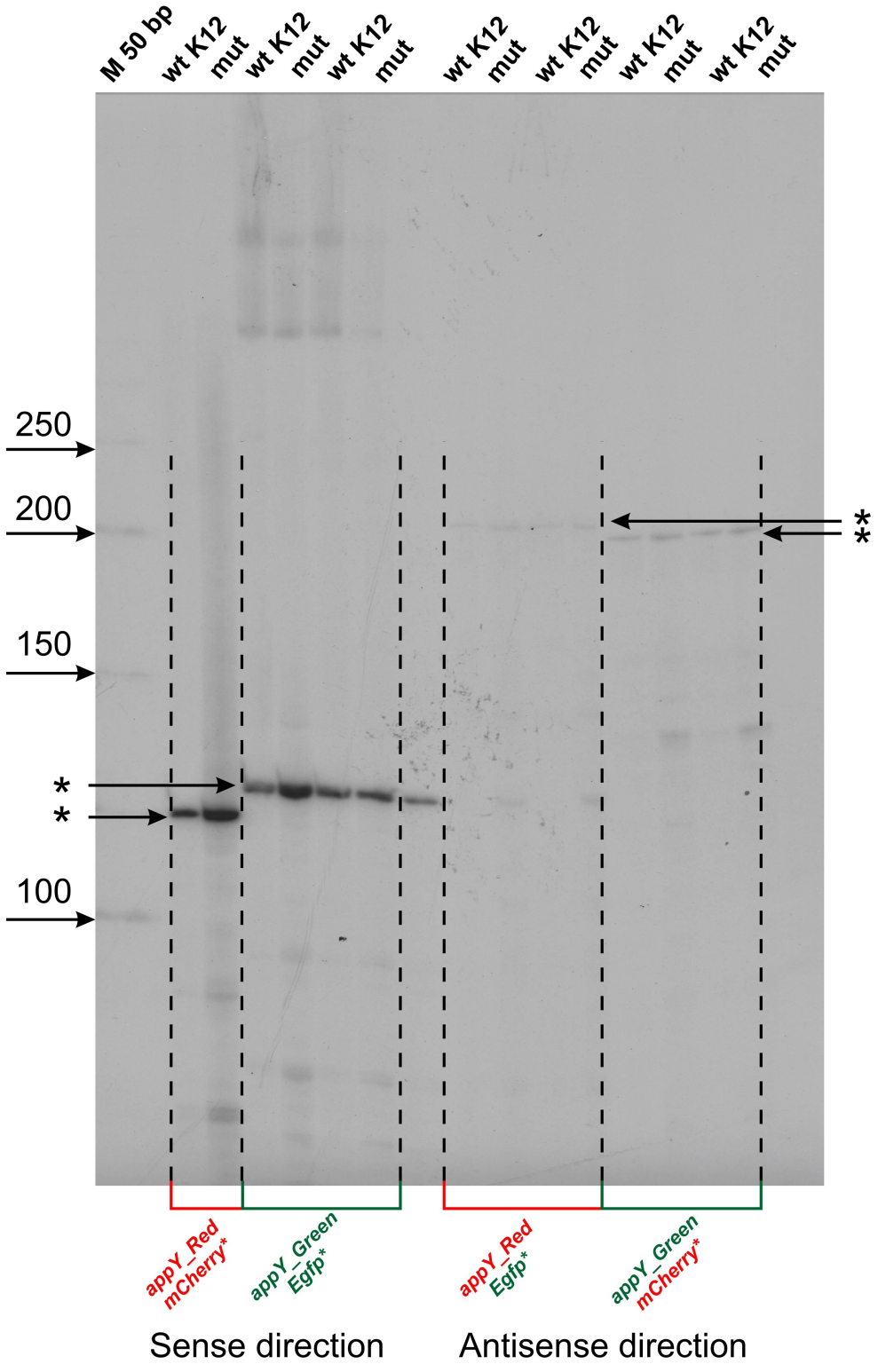

Fig. 2. Mapping of transcription start sites (TSS) within cloned fragments of *app*Y regulatory region. Reverse transcription reaction was carried out with radiolabeled oligonucleotide primers specific for *mCherry* and *Egfp* genes. Products were separated in 6% denaturating PAAG, 8M urea and radioautographed. Gels were calibrated by radiolabeled P32 50 bp DNA ladder (NEB) (M). The lengths of ladder fragments are indicated on the left by arrows. TSSs are marked by asterisks. "M" is marker lane, "wt K12" corresponds to the wild type E.coli MG1655 transformed by plasmids, "mut" corresponds to the host strain of E.coli DPB923 (*E.coli* Genetic Stock Center, CGST 7894).

A

TTCAGGTGCGTTGTAGTGAGTTTATGTTAATAAAAAGCATAGTAAGCGTTGAAAAATGTAACTTTGAAATAAGTTAG
AATAAAAAACAACATACATATAATAATTTAATCTTAAATGAAATTTATTAAAATTTGCAAACTATAATTTTGTGTAT
AAAAATATAAATGCACATCATCCTGATTA

B

TAATCAGGATGATGTGCATTTATATTTTTATACACAAAATTATAGTTTGCAAATTTTAATAAATTTCATTTAAGATT
AAATTATTATATGTATGTTGTTTTTTATTCTAACTTATTTCAAAGTTACATTTTTCAACGCTTACTATGCTTTTTAT
TAACATAAACTCACTACAACGCACCTGAA

Fig. 3. Nucleotide sequences of the studied fragments of *E. coli* and schematic representation of the location of the P1 and P2 promoters in the Red and Green constructions. In the Red (A) and Green (B) constructions, the start points located in the coding chain are shown in red letters, and the start points located in the complementary chain are shown in green letters. The complete nucleotide sequences of the Red and Green constructions are presented in **Supplementary material**. (C) Location of promoters in the Red construction. (D) Location of promoters in the Green construction. The *mCherry* gene is shown in red, the *Egfp* gene is in green, and the *E. coli* region between these two genes is shaded.

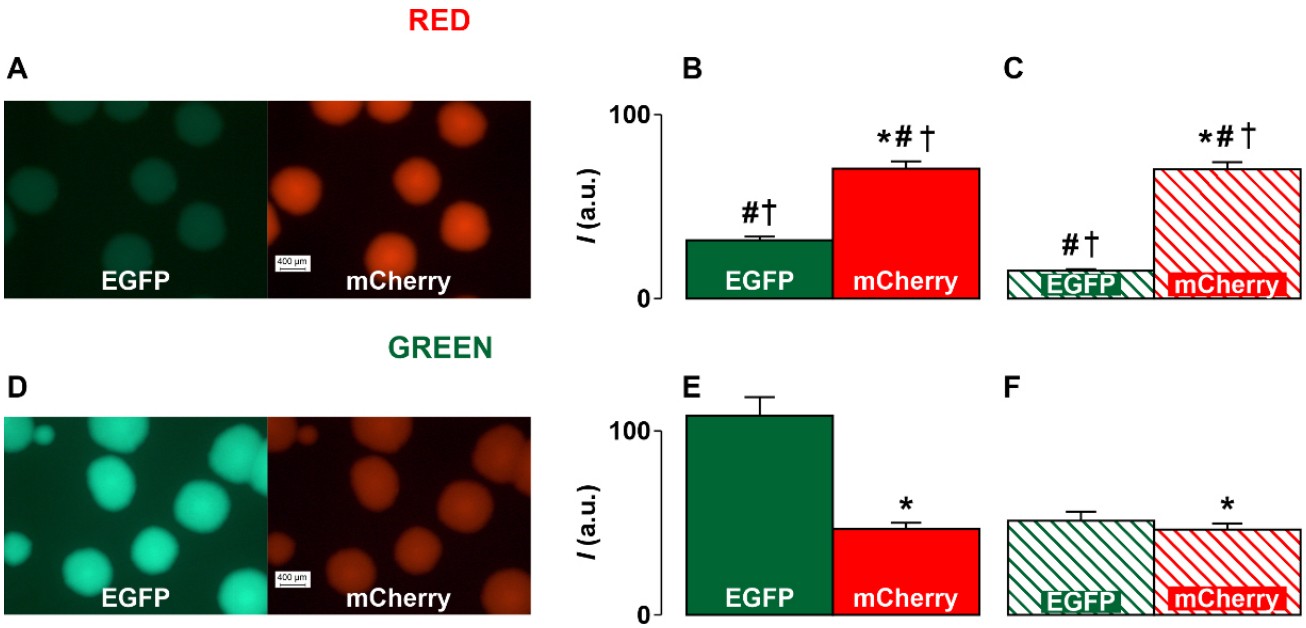

Fig. 4. Fluorescent images of *E. coli* cell colonies and fluorescence intensity diagrams obtained as a result of image processing using the ImageJ program. (A) The images of cell colonies transformed with Red constructions. (D) The images of cell colonies transformed with Green constructions. (B,E) The fluorescence intensity (*I*) diagrams obtained in the case of Red (B) and Green (E) constructions. (C,F) The fluorescence intensity diagrams that take into account the difference in the brightness of the luminescence of cells with red and green fluorescent proteins in the case of the Red (C) and Green (F) constructions. The diagrams show the data as mean and tandard deviation (SD). * is significant differences (Mann–Whitney test, $p < 0.05$) between the fluorescence intensities of *mCherry* and *Egfp* proteins in cell colonies within a single genetic construction. # is significant differences (Mann–Whitney test, $p < 0.05$) in fluorescence intensity of identical proteins between Red and Green constructions. † is significant differences (Mann–Whitney test, $p < 0.05$) in fluorescence intensity of *mCherry* and *Egfp* proteins between Red and Green constructions.

promoter activity in Red and Green constructions, it can be seen that the P1 promoter decreases (Fig. 4C, *mCherry*; Fig. 4F, *Egfp*) and the P2 promoter increases (Fig. 4C, *Egfp*; Fig. 4F, *mCherry*) its functional activity when the orientation of the fragment is changed.

3.2 Model Studies

3.2.1 Energy Profile of the Red or Green Constructions

The energy profiles of the coding and complementary sequences are important characteristics that reflect the physical properties of DNA. An analysis of such profiles makes it possible to draw preliminary conclusions about the behavior of kinks.

In [17], using Eqn. 15, the energy profiles of the coding and complementary sequences of the Red and Green constructions were calculated. It was shown that in both cases, the deepest well corresponds to the region located between the red and green protein genes. This result allowed us to suggest that kinks are most likely to be activated in this region, which, in turn, suggests that promoters and, accordingly, transcription starts are located in this region. Further, we use this result to model the movement of the kink, placing the reference point in the middle of the region located between the red and green protein genes. We also limit our consideration to a small DNA fragment that includes only three regions, namely, the region denoted as r2 and located between the *Egfp* and *mCherry* genes, the region located to the left and denoted as r1, and the region located to the right and denoted as r3. Fig. 5A shows the results of the energy profile calculations obtained for the coding (red solid line) and complementary (green dotted line) sequences of the Red construction. Fig. 5B shows the energy profiles calculated for the Green construction.

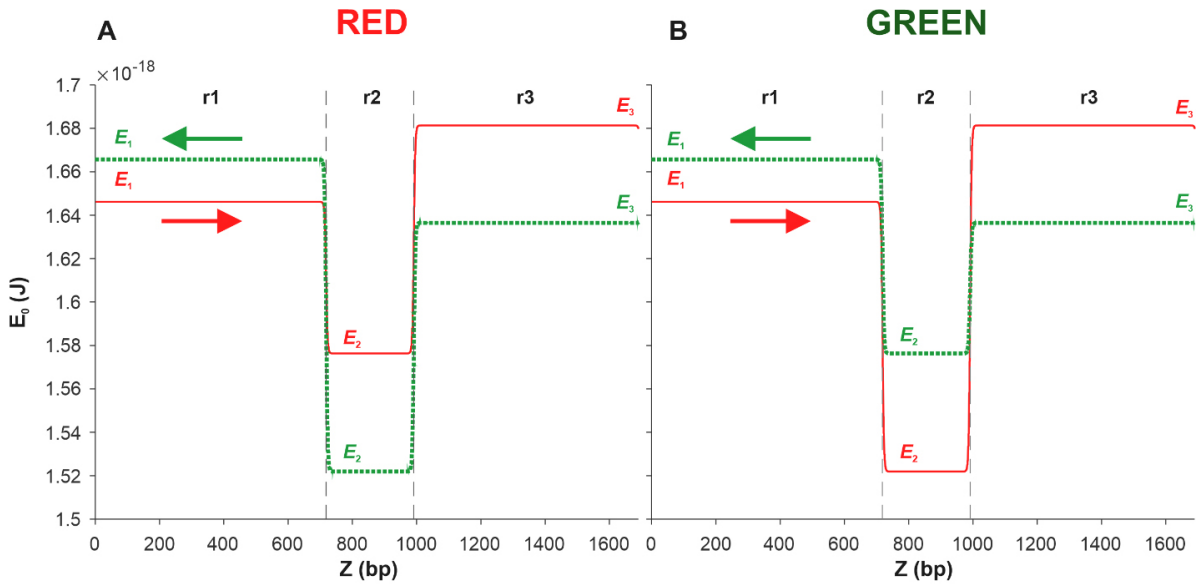

Fig. 5. Energy profiles of the coding (red solid line) and complementary (green dotted line) sequences of two DNA fragments. Each of the fragments consists of three sections r1, r2, and r3 of the pPF1 plasmid, and the *E. coli* fragment is integrated into the r2 section in red (A) and green (B) orientations. E1, E2, and E3 are the minimum energies required to form a kink in the regions r1, r2, and r3, respectively. Arrows indicate the direction from the 3′ to the 5′ end.

From Fig. 5A, it can be seen that in the case of the Red construction, the depth of the well in the energy profile of the complementary sequence, shown by the green dotted line, is greater than that of the coding sequence profile, shown by the red solid line. This suggests that kink formation in the r2 region of the coding sequence is energetically more expensive than in the same region of the complementary sequence.

However, the kink formed in the r2 region of the coding sequence, moving in the direction from the 3′ to the 5′ end, shown by the red arrow, must overcome the barrier with a height of $(E_3 - E_2) = 0.105 \times 10^{-18}$ J. This height is much less than the barrier height, which is necessary to overcome the kink formed in the region r2 of the complementary sequence and move in the direction from the 3′ to the 5′ end, as shown by the green arrow. The height of the latter is $(E_1 - E_2) = 0.144 \times 10^{-18}$ J. These estimates allow us to conclude that the movement of kinks activated in the r2 region of the coding sequence towards the red protein gene is energetically more favorable. This conclusion completely coincides with the results obtained experimentally.

We observe the opposite situation in the case of Green. Fig. 5B shows that in the case of the complementary sequence (green dotted line), the barrier height in the direction of the green protein gene is $(E_1 - E_2) = 0.090 \times 10^{-18}$ J, which is significantly less than the barrier height in the direction of the red protein gene, which is $(E_3 - E_2) = 0.159 \times 10^{-18}$ J. This suggests that the movement of kinks activated in the r2 region of the complementary sequence towards the green protein gene is energetically more favorable and, therefore, preferable. This conclusion also completely coincides with the results obtained experimentally.

Thus, summing up the analysis of energy profiles, it can be stated that the results of the modeling completely coincide with the results of our experiment, according to which, in the case of the Red construction, more red protein is produced, and in the case of the Green construction, more green protein is produced.

3.2.2 Kink Trajectories in the Red and Green Constructions

To clarify the role of the torsion moment M_τ in the dynamics of kinks, we solved the model equations numerically when performing numerical calculations, and the following initial and boundary conditions were used: it was assumed that at $t = 0$, the solutions of the model equations had the form of kinks of the sine-Gordon equations, which were located at points $z_{0,1}$ and $z_{0,2}$ in the center of the region r2, and have zero initial velocities:

$$\varphi_{k,1}(z,0) = 4arctg\{\exp[(\bar{\gamma}_1/\bar{d}_1)(z - z_{0,1})]\}, \tag{19}$$

$$\varphi_{k,2}(z,0) = 4arctg\{\exp[(\bar{\gamma}_2/\bar{d}_1)(z - z_{0,2})]\}. \tag{20}$$

In addition, we assumed that the points $z_{0,1}$ and $z_{0,2}$ are located far enough from the left and right ends of the sequence. Then the boundary conditions at the ends can be approximated as follows:

$$\varphi_{k,1}\left(Z_{\text{left}}, t\right) = \varphi_{k,2}\left(Z_{\text{left}}, t\right) = 0, \tag{21}$$

$$\varphi_{k,1}\left(Z_{\text{right}}, t\right) = \varphi_{k,2}\left(Z_{\text{right}}, t\right) = 2\pi. \tag{22}$$

where Z_{left} and Z_{right} are the coordinates of the left and right ends of the sequence, respectively. Details of the entire numerical calculation scheme are given in [15].

The results of the calculations are presented in Figs. 6,7 in the form of kink trajectories on the {z, t} plane where different trajectories relate to different values of the torsion moment M_τ. When making calculations, the initial kink velocities were taken to be zero, the kink starting point was suggested to be located in the middle of the region r2, and the constant torsion moment was suggested to be directed from the 3′ to the 5′ end.

In the case of the Red construction, the calculated trajectories of the kink movement in the direction of the *mCherry* gene (curves 1, 2, and 3) are presented in Fig. 6A. It can be seen that with an increase in the torsion moment from the value $M_{\tau,1} = 16.25 \times 10^{-23}$ J to the value $M_{\tau,2} = 16.50 \times 10^{-23}$ J, the nature of the kink motion changes significantly: from the damped oscillatory motion of the kink near the right boundary of the r2 region (curve 1) to the translational movement (curve 2). A further increase of the torsion moment, for example, to the value $M_{\tau,3} = 21.25 \times 10^{-23}$ J, does not change the translational nature of the kink motion (curve 3). These results indicate the existence of a threshold value of the torsion moment in the interval $(16.25 \times 10^{-23};\ 16.50 \times 10^{-23})$ J.

The trajectories of the kink movement in the direction of the *Egfp* gene (curves 4, 5, and 6) are shown in Fig. 6B. It can be seen that, as in the previous case, the character of the kink motion changes: from damped oscillatory motion of the kink near the left boundary of the region r2 (curve 4) to translational motion (curves 5 and 6). Thus, it can be argued that in this case, there is also a threshold value of the torsion moment, and only its value is greater than in the previous case, namely it is located in the interval $(21.00 \times 10^{-23};\ 21.25 \times 10^{-23})$ J.

The prethreshold and threshold values of the torsion moment, as well as data on the kink behavior at these values of the torsion moment, are collected in Table 1.

Table 1. Prethreshold and threshold values of the torsion moment calculated in the case of Red construction.

RED	Toward *mCherry* gene	Toward *Egfp* gene
Prethreshold values	16.25×10^{-23} J	21.00×10^{-23} J
Threshold values	16.50×10^{-23} J	21.25×10^{-23} J

From the data in Table 1, it follows:

(1) When the kink movement starts from the middle of the region r2 towards the red protein gene, the threshold value of the torsion moment is less than the threshold value when the kink moves from the same start point towards the green protein gene. It follows that it is the movement of the kink towards the red protein gene that will be preferable, which means that in the case of the Red construction, more red proteins will be produced. This conclusion agrees with the obtained experimental data, as well as with the predictions made in [17].

(2) Calculations showed that the minimum value of the torsion moment required to produce red proteins is in the range $\left(16.25 \times 10^{-23};\ 16.50 \times 10^{-23}\right)$ J. And for the production of green proteins, the minimum value is in the range $(21.00 \times 10^{-23};\ 21.25 \times 10^{-23})$ J.

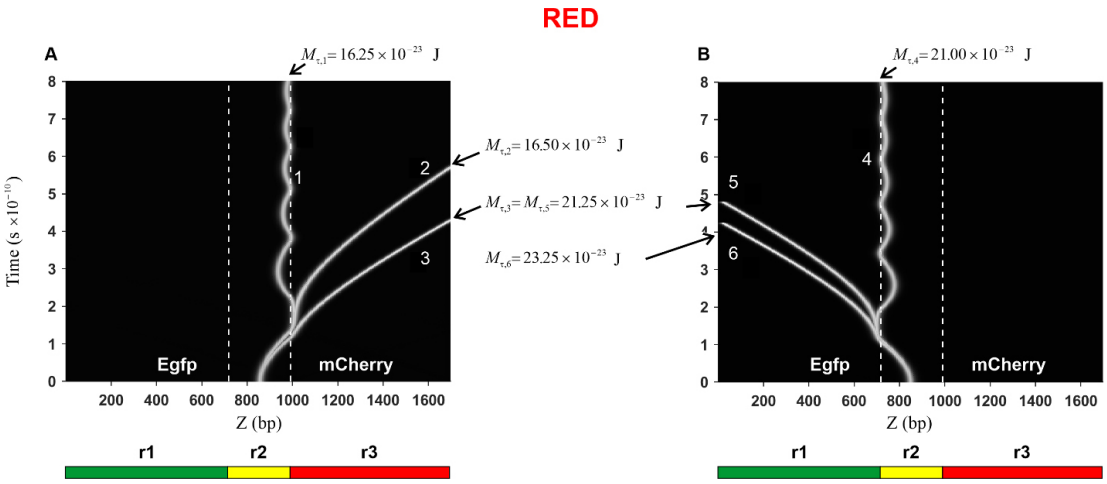

Fig. 6. Kink trajectories calculated in the case of Red construction. (A) Trajectories of kinks propagating in the coding strand. (B) Trajectories of kinks propagating in the complementary strand. Calculations were carried out for the following values of the constant torsion moment M_τ: 16.25×10^{-23} J, 16.50×10^{-23} J, 21.00×10^{-23} J, 21.25×10^{-23} J and 23.25×10^{-23} J.

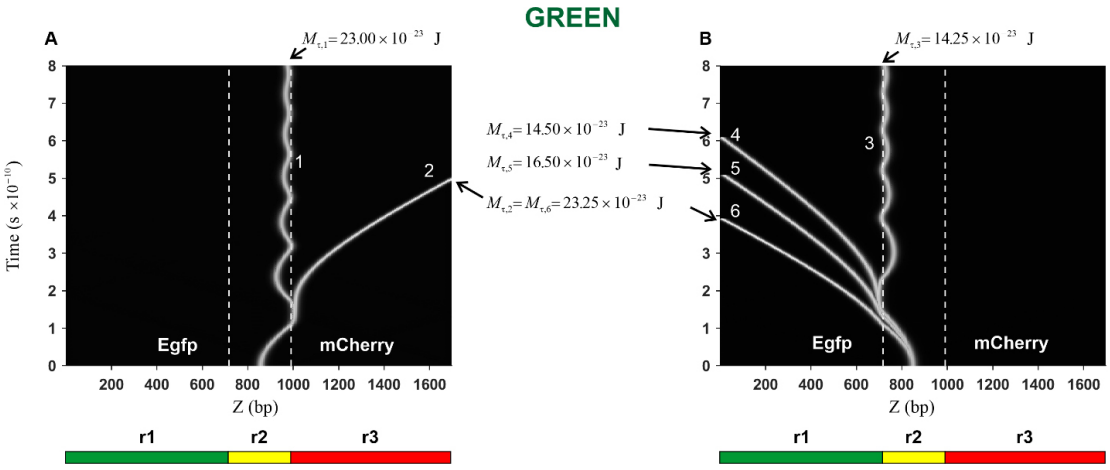

Fig. 7. Kink trajectories calculated in the case of Green construction. (A) Trajectories of kinks propagating in the coding strand. (B) Trajectories of kinks propagating in the complementary strand. Calculations were carried out for the following values of the constant torsion moment M_τ: 14.25×10^{-23} J, 14.50×10^{-23} J, 16.50×10^{-23} J, 23.00×10^{-23} J and 23.25×10^{-23} J.

In the case of the Green construction, the calculated trajectories of the kink movement in the direction of the *mCherry* gene (curves 1 and 2) are presented in Fig. 7A. It can be seen that with an increase in the torsion moment from the value $M_{\tau,1} = 23.00 \times 10^{-23}$ J to the value $M_{\tau,2} = 23.25 \times 10^{-23}$ J, the nature of the kink motion in the direction of the *mCherry* gene changes significantly: from the damped oscillatory motion of the kink at the right boundary of the r2 region (curve 1) to the translational movement (curve 2). This indicates the existence of a threshold value of the torsion moment in the interval $\left(23.00 \times 10^{-23};\ 23.25 \times 10^{-23}\right)$ J.

The trajectories of the kink movement in the direction of the *Egfp* gene (curves 3, 4, 5, and 6) are shown in Fig. 7B. It can be seen that the character of the kink motion dramatically changes: from damped oscillatory motion of the kink (curve 3) to translational motion (curves 4, 5 and 6). Thus, it can be argued that in this case, there is also a threshold value, only its value is less and it is located in the interval $(14.25 \times 10^{-23};\ 14.50 \times 10^{-23})$ J.

The prethreshold and threshold values of the torsion moment, as well as data on the kink behavior at these values of the torsion moment, are collected in Table 2.

Table 2. Prethreshold and threshold values of the torsion moment calculated in the case of Green construction.

GREEN	Toward *mCherry* gene	Toward *Egfp* gene
Prethreshold values	23.00×10^{-23} J	14.25×10^{-23} J
Threshold values	23.25×10^{-23} J	14.50×10^{-23} J

From the data in Table 2, it follows:

(1) When the kink moves towards the green protein gene, the threshold value of the torsion moment is less than the torsion. This conclusion agrees with the obtained experimental data, as well as with the prediction moment value when the kink moves towards the red protein gene. It follows that it is the movement of the kink towards the green protein gene that will be preferable, which means that in the case of the Green construction, more green proteins will be produced [17].

(2) Calculations showed that the minimum value of the torsion moment required to produce green proteins lies in the range $(14.25 \times 10^{-23};\ 14.50 \times 10^{-23})$ J. And, for the production of red proteins, the minimum value is in the range $(23.00 \times 10^{-23};\ 23.25 \times 10^{-23})$ J.

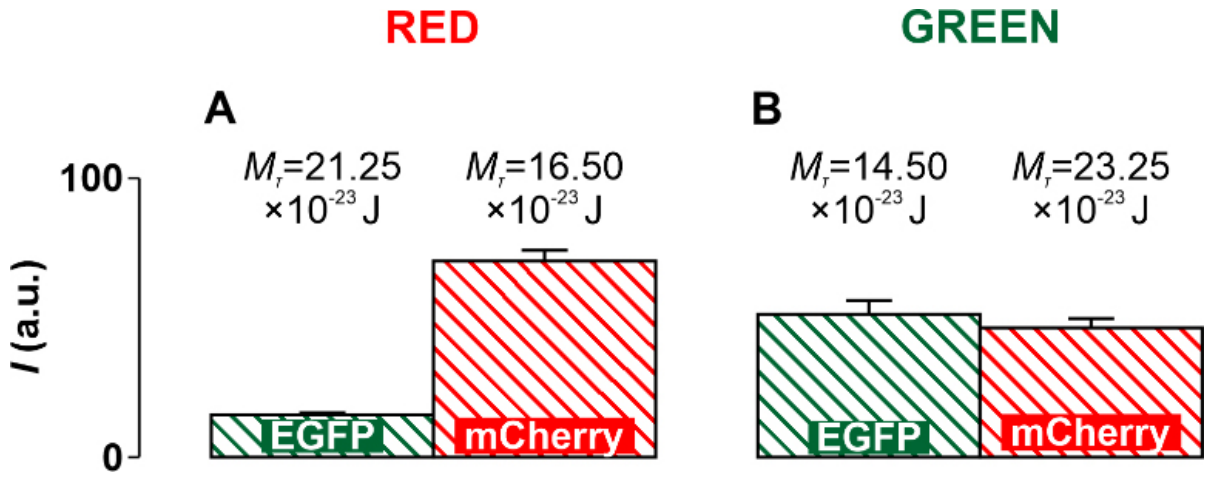

Fig. 8. Diagrams obtained experimentally and supplemented with threshold values of the torsion moment M_τ calculated by mathematical modeling methods. The diagrams are shown for the Red (A) and Green (B) genetic constructions.

4. Discussion

In this work, the relationship between the dynamics of open states (kinks) and promoter functioning in the *app*Y_red and *app*Y_green genetic constructions was investigated by experimental and mathematical modeling methods.

Assessing the functionality of the promoters predicted in [13] using the pPF1 reporter vector revealed TSSs in the studied DNA fragment, indicating the presence of two functionally active divergent promoters, one on each strand. The efficiency of promoters was evaluated by fluorescent protein gene expression intensity in *E. coli* cell colonies transformed with plasmid constructions using fluorescence image and corresponding brightness diagram analysis. It was found that,

in the case of the Red construction, *mCherry* expression is higher than expression of *Egfp*, indicating higher promoter efficiency on the coding strand compared to the promoter on the complementary strand (Fig. 4). On the contrary, in the case of the Green construction, when the *E. coli* fragment is rotated, the expression of *Egfp* is higher than *mCherry*, indicating higher promoter efficiency on the complementary strand than on the coding strand (Fig. 4). This showed that the P1 promoter was stronger than the P2 promoter. Since it has been shown that the promoter strength depends on the orientation of the DNA fragment under study in the plasmid, it can be concluded that the dynamic properties of DNA in general affect the strength of promoters. The presence of the relationship between the strength of promoters and the dynamic properties of DNA, in particular between the promoter strength and DNA supercoiling, was also shown by other authors [11].

From the energetic profiles calculated at the beginning in [17] and continued in this work, we found that the most favorable and preferred site for the formation of kinks in both constructions is the region r2 located between the red and green protein genes. This result is in good agreement with the data [13], according to which the r2 region containing a DNA fragment of *E. coli*, has multiple potential starts of transcription.

From the same energetic profiles, it was found also that in the case of the Red construction, the formation of a kink in the complementary sequence of this r2 region is less expensive. However, the height of the energy barrier that prevents the kink movement towards the red protein gene is less than the height of the energy barrier that prevents the kink movement towards the green protein gene. In the case of Green construction, the situation is exactly the opposite: the formation of a kink in the coding sequence of the region r2 is less expensive, but the energy barrier that the kink needs to overcome in order to move towards the red protein gene turned out to be higher than the energy barrier towards the green protein gene. These results are in good agreement with the above experimental data showing that in the case of the Red construction, *mCherry* expression is higher than *Egfp* expression, and in the case of the Green construction, *Egfp* expression is higher than *mCherry* expression. Indeed, the diagrams presented in Fig. 4 show that in the case of Red construction, the red diagram is higher than the green one (Fig. 4C), and in the case of Green construction, the opposite is true: the green diagram turned out to be higher than the red one (Fig. 4F). Thus, the assumption made in [17] that in the case of the Red construction, more red proteins will be produced, and in the case of the Green construction, on the contrary, more green proteins will be produced, was fully confirmed.

Taking into account the influence of the torsion moment, additional features of the Red and Green internal dynamics are revealed. The existence is shown and estimates of the threshold values of the torsion moment Mτ are made (Tables 1,2). It was demonstrated that when these values are reached, the nature of the kink motion changes significantly: instead of oscillatory motion near one of the boundaries of the intermediate region, the kink passes to translational motion and gets the opportunity to move to neighboring regions, which are genes of fluorescent proteins (Figs. 6,7). In the case of the Red construction, the threshold value of the torsion moment when the kink moves toward the red protein gene is less than the threshold value when the kink moves toward the green protein gene. This suggests that the movement of the kink towards the red protein gene is preferable, and this correlates well with the experimental data (Fig. 8A) and agrees with the results of the energy profile analysis. The opposite result was obtained in the case of the Green construction: the threshold value of the torsion moment when the kink moves towards the green protein gene is less than the similar value when the kink moves towards the red protein gene. From this, it was concluded that it is the movement of the kink towards the green protein gene that will be preferable, which means that more green protein will be synthesized. This also correlates well with the experimental data (Fig. 8B) and the energy analysis results.

5. Conclusions

Thus, two divergent promoters, one on each strand, were identified for the Red and Green genetic constructions, and the dependence of promoter functionality on the orientation of the fragments was shown. A good correlation between the results of experiments and model calculations allows us to state that the mechanism of such dependence is determined by the dynamics of open states in the Red and Green constructions, which depends on the energy profile of the sequence and the magnitude of the torsional moment.

Using the dynamic model of DNA, it was possible to obtain numerical estimates of the minimum values of the torsion moment required for protein synthesis. It has been shown that in the case of the Red construction, the minimum value of the torsion moment required for the synthesis of red proteins lies in the range $\left(16.25 \times 10^{-23};\ 16.50 \times 10^{-23}\right)$ J, and the moment required for the synthesis of green proteins is in the range $(21.00 \times 10^{-23};\ 21.25 \times 10^{-23})$ J. In the case of the Green construction, the minimum value of the torsional moment required for the synthesis of green proteins lies in the interval $(14.25 \times 10^{-23};\ 14.50 \times 10^{-23})$ J, and the minimum value of the torsion moment required for the synthesis of red proteins lies in the interval $(23.00 \times 10^{-23};\ 23.25 \times 10^{-23})$ J. The obtained values indicate that the considered constructions are not equivalent in terms of the sensitivity of their structural-dynamic properties to the level of DNA supercoiling. This, in turn, indicates that the used mathematical model of DNA quite well describes the features of the internal dynamics of the DNA molecule in the presence of an external torsion field.

Availability of Data and Materials

All datasets on which the conclusions of a manuscript depend are available to readers from the corresponding author.

Author Contributions

IM performed experimental part of the research. AG and LY performed mathematical part of the research. All auhors contributed to the design of the research study, data analysis, and writing of the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

We acknowledge the Institute of Cell Biophysics of the Russian Academy of Sciences for administrative and technical support.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/j.fbl2904155.

References

[1] Watson JD, Baker TA, Bell SP, Gann AA, Levine M, Losick RM. Molecular Biology of the Gene. 7th edn. Pearson Education RHLtd: India. 2014.

[2] Alexandrov BS, Gelev V, Yoo SW, Alexandrov LB, Fukuyo Y, Bishop AR, *et al*. DNA dynamics play a role as a basal transcription factor in the positioning and regulation of gene transcription initiation. Nucleic Acids Research. 2010; 38: 1790–1795.

[3] Parker SCJ, Tullius TD. DNA shape, genetic codes, and evolution. Current Opinion in Structural Biology. 2011; 21: 342–347.

[4] Duzdevich D, Redding S, Greene EC. DNA dynamics and single-molecule biology. Chemical Reviews. 2014; 114: 3072–3086.

[5] Yakushevich LV, Krasnobaeva LA. Ideas and methods of nonlinear mathematics and theoretical physics in DNA science: the McLaughlin-Scott equation and its application to study the DNA open state dynamics. Biophysical Reviews. 2021; 13: 315–338.

[6] Duprey A, Groisman EA. The regulation of DNA supercoiling across evolution. Protein Science. 2021; 30: 2042–2056.

[7] Ma J, Wang MD. DNA supercoiling during transcription. Biophysical Reviews. 2016; 8: 75–87.

[8] Ma J, Tan C, Gao X, Fulbright RM, Jr, Roberts JW, Wang MD. Transcription factor regulation of RNA polymerase's torque generation capacity. Proceedings of the National Academy of Sciences of the United States of America. 2019; 116: 2583–2588.

[9] Ma J, Bai L, Wang MD. Transcription under torsion. Science. 2013; 340: 1580–1583.

[10] Li J, Zhang Y. Relationship between promoter sequence and its strength in gene expression. The European Physical Journal. E, Soft Matter. 2014; 37: 44.

[11] Zhi X, Leng F. Dependence of transcription-coupled DNA supercoiling on promoter strength in Escherichia coli topoisomerase I deficient strains. Gene. 2013; 514: 82–90.

[12] Masulis IS, Babaeva ZS, Chernyshov SV, Ozoline ON. Visualizing the activity of Escherichia coli divergent promoters and probing their dependence on superhelical density using dual-colour fluorescent reporter vector. Scientific Reports. 2015; 5: 11449.

[13] Shavkunov KS, Masulis IS, Tutukina MN, Deev AA, Ozoline ON. Gains and unexpected lessons from genome-scale promoter mapping. Nucleic Acids Research. 2009; 37: 4919–4931.

[14] Yakushevich LV. Nonlinear Physics of DNA. Wiley: Weinheim, Germany. 2004.

[15] Grinevich AA, Ryasik AA, Yakushevich LV. Trajectories of DNA bubbles. Chaos, Solitons & Fractals. 2015; 75: 62–75.

[16] Grinevich AA, Yakushevich LV. The influence of the DNA torque on the dynamics of transcription bubbles in plasmid PTTQ18. Journal of Theoretical Biology. 2018; 453: 68–77.

[17] Grinevich AA, Masulis IS, Yakushevich LV. Mathematical Modeling of the Behavior of the Transcription Bubble in the pPF1 Plasmid and Its Modified Versions. A link between the Plasmid Energy Profile and the Direction of Transcription. Biophysics. 2021; 66; 209–217.

[18] Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA. 1989.

[19] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 2012; 9: 671–675.

[20] Brightness of fluorescent proteins. Available at: https://www.fpbase.org/table/ (Accessed: 14 November 2023).

[21] Englander SW, Kallenbach NR, Heeger AJ, Krumhansl JA, Litwin S. Nature of the open state in long polynucleotide double helices: possibility of soliton excitations. Proceedings of the National Academy of Sciences of the United States of America. 1980; 77: 7222–7226.