

## МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПОВЕДЕНИЯ ТРАНСКРИПЦИОННОГО ПУЗЫРЯ В ПЛАЗМИДЕ rPF1 И ЕЕ МОДИФИКАЦИЯХ. СВЯЗЬ МЕЖДУ ЭНЕРГЕТИЧЕСКИМ ПРОФИЛЕМ ПЛАЗМИДЫ И НАПРАВЛЕНИЕМ ТРАНСКРИПЦИИ

© 2021 г. А.А. Гриневич, И.С. Масулис, Л.В. Якушевич

Институт биофизики клетки РАН – обособленное подразделение ФИЦ «Пушкинский научный центр биологических исследований РАН», 142290, Пущино Московской области, Институтская ул., 3

E-mail: grin\_aa@mail.ru

Поступила в редакцию 06.11.2020 г.

После доработки 26.01.2021 г.

Принята к публикации 29.01.2021 г.

С использованием нелинейного уравнения синус-Гордона и его решения – кинка – моделировали и изучали динамику транскрипционного пузыря в плазмиде rPF1. На основе рассчитанного энергетического профиля плазмиды rPF1 и ее модификаций было показано, что минимальная потенциальная энергия формирования кинка или нуклеации транскрипционного пузыря соответствует области между генами белков Egfr и mCherry. Вставка в область между Egfr и mCherry однородных последовательностей показала, что кинк с большей вероятностью активируется в polyT или polyC по сравнению с polyA или polyG, что свидетельствует о зависимости нуклеации транскрипционного пузыря от физической массы оснований. В случае вставок в область между Egfr и mCherry небольших фрагментов природной последовательности *Escherichia coli* модель позволяет выявить нити ДНК с наибольшей вероятностью нуклеации транскрипционного пузыря и соответственно определить направление (в сторону гена Egfr или mCherry) транскрипции, что говорит о связи между направлением транскрипции и энергетическим профилем плазмиды.

*Ключевые слова:* математическое моделирование, нелинейное уравнение синус-Гордона, кинк, солитоны, плаزمиды rPF1, ДНК, транскрипционный пузырь.

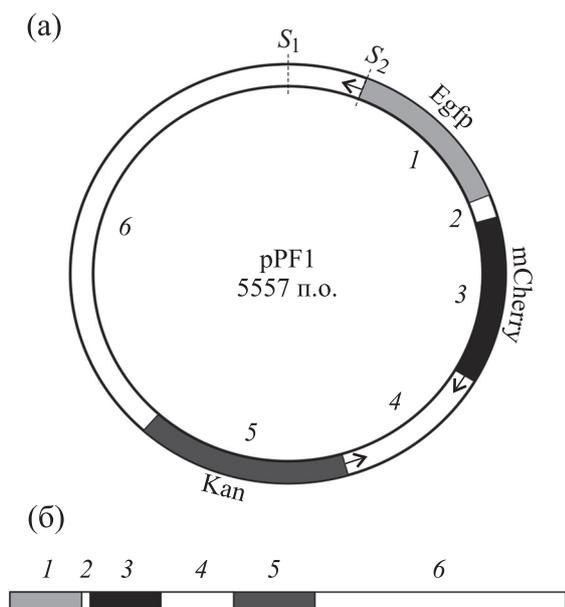
DOI: 10.31857/S0006302921020058

Математическое моделирование с использованием методов нелинейной математики и теоретической физики – один из наиболее простых и эффективных инструментов изучения динамики сложных биологических систем. Постановка и реализация компьютерных экспериментов, заменяющих и дополняющих эксперименты с реальными системами, все более широко и активно осуществляется исследователями из-за целого ряда преимуществ: они дешевле, быстрее и обладают предсказательными возможностями. В настоящей работе методы математического моделирования применяются для изучения особенностей динамики транскрипционного пузыря в плазмиде rPF1 и предсказания с их помощью – в какой области плазмиды с наибольшей вероятностью произойдет нуклеация транскрипционного пузыря и в каком направлении пойдет процесс транскрипции.

*Сокращение:* п.о. – пары оснований.

Плазмиды представляют собой кольцевые молекулы ДНК. Они являются удобными объектами для искусственных манипуляций с генетическим материалом. Векторная плаزمиды rPF1 – типичный представитель этого класса биологических объектов (рис. 1). Она была получена из плазмиды pET-28b [1] и дополнена двумя генами, кодирующими зеленый и красный флуоресцентные белки Egfr и mCherry [2]. Данный вектор, как репортерная конструкция, позволяет одновременно регистрировать транскрипцию в двух направлениях, инициированную на участках последовательности между генами флуоресцентных белков.

Начало процесса транскрипции определяется взаимодействием РНК-полимеразы с промоторными областями ДНК. В результате происходит локальное расплетание двойной спирали ДНК, формирование открытого транскрипционного комплекса (транскрипционного пузыря) протяженностью ~ 13 пар оснований (п.о.) и последующее движение этого пузыря вдоль молекулы ДНК в процессе элонгации транскрипции [3].



**Рис. 1.** Кольцевая (а) и линейная (б) схемы плазмиды рPF1. Ген зеленого белка *Egfr* (участок 1) и ген устойчивости к канамицину *Kan* (участок 5) транскрибируются по комплементарной нити. Ген красного белка *mCherry* (участок 3) транскрибируется по основной нити.  $S_1$  – точка, указывающая начало последовательности;  $S_2$  – точка, указывающая начало первого участка.

Направление синтеза РНК задается конфигурацией транскрипционного пузыря и ориентацией матричной и нематричной нитей в полости РНК-полимеразы (рис. 2), а скорость элонгации зависит от нуклеотидной последовательности транскрибируемого участка [4]. Кинетика формирования пузыря и скорость его движения вдоль молекулы ДНК в процессе транскрипции являются ключевыми факторами, обуславливающими вариацию в уровне экспрессии различных генов в диапазоне 10000 раз [5]. Механо-динамические и энергетические свойства нуклеотидов играют важную роль при формировании транскрипционного пузыря. Известно, что это многостадийный процесс, сопровождающийся образованием промежуточных состояний и вращательной динамикой отдельных оснований [6–9].

Внутренняя (конформационная) лабильность ДНК и ее способность реагировать на точечные возмущения в упорядоченной структуре могут рассматриваться в качестве физической основы для реализации ее основной функции – хранения и экспрессии генетической информации. С этой точки зрения одним из наиболее перспективных направлений является использование математического моделирования конформационной подвижности кольцевых ДНК для предсказания динамического поведения локальных возмущений



**Рис. 2.** Транскрипционный пузырь.

структуры двойной спирали с учетом механохимических свойств.

Одним из инструментов, используемых для математического моделирования конформационной подвижности ДНК, является аппарат нелинейной математики и, в частности, свойства нелинейного уравнения синус-Гордона и его односолитонных решений – кинков [10, 11]. Движение кинков можно рассматривать как движение квазичастиц в потенциальном поле молекулы. Ранее движение транскрипционного пузыря в плазмиде рТТQ18 моделировалось нами как движение кинка в потенциальном поле этой плазмиды [12]. Поведение кинка в такой модели полностью определялось энергетическим профилем потенциального поля плазмиды. Аналогичный математический аппарат применялся для расчета потенциального поля плазмиды рBR322 в работе [13]. В настоящей работе мы используем этот аппарат для расчета энергетического профиля плазмиды рPF1 и различных ее модификаций, полученных путем интегрирования в область, расположенную между двумя генами белков *Egfr* и *mCherry*, однородных последовательностей и небольших фрагментов *Escherichia coli*, взятых из так называемых промоторных островов – областей, высоко насыщенных промотороподобными участками [14]. Цель данной работы – рассчитать энергетические профили плазмиды рPF1 и ее модификаций и использовать полученные результаты для предсказания поведения кинка, что в свою очередь позволит сделать выводы о том, в какой области плазмиды с наибольшей вероятностью произойдет нуклеация транскрипционного пузыря и в каком направлении (в сторону гена *Egfr* или *mCherry*) пойдет процесс транскрипции.



СТРУКТУРА ПОСЛЕДОВАТЕЛЬНОСТИ ПЛАЗМИДЫ pPF1

Полная последовательность нуклеотидов плазмиды pPF1 приведена на рис. 3. Общая длина последовательности равна 5557 п.о. Она содержит три гена: ген зеленого белка Egfp с координатами (333–1049), помеченный на рис. 3 светло-серым маркером, ген красного белка mCherry с координатами (1134–1841), помеченный черным маркером, и ген устойчивости к канамицину Kan с координатами (2569–3381), помеченный темно-серым маркером. Жирным шрифтом белого цвета на темно-сером фоне показан сайт узнавания эндонуклеазы рестрикции BglIII (AGATCT) с координатами (1089–1094).

Кроме генов белков, последовательность плазмиды pPF1 содержит три промежуточные участка. Назовем их вторым (1050–1133), четвертым (1842–2568) и шестым участками (3382–5557) + (1–332).

Основное внимание при генетических манипуляциях уделяется второму участку, так как он расположен между генами белков Egfp и mCherry и имеет удобный для манипуляций нуклеотидный состав. На рис. 4 показана структура этого участка. В центре последовательности находится сайт узнавания эндонуклеазы рестрикции BglIII AGATCT длиной 6 п.о. Слева и справа от него находятся короткие последовательности длиной 39 п.о. каждая, причем короткая последовательность справа от рестриктазного сайта является обратно комплементарной к короткой последовательности слева от рестриктазного сайта.

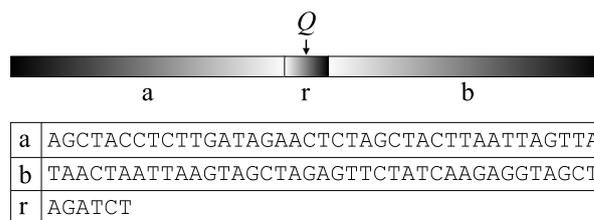


Рис. 4. Схематическое изображение последовательности второго участка. Q – точка, указывающая место, куда будут встроены небольшие модельные последовательности в компьютерных экспериментах.

Как и в экспериментах по искусственным генетическим манипуляциям *in vitro*, в наших компьютерных экспериментах мы будем встраивать во второй участок небольшие модельные последовательности. Сначала для отработки метода в качестве модельных будут использованы однородные последовательности: polyA, polyT, polyG, polyC, а затем – небольшие фрагменты последовательности *Escherichia coli*.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КОНФОРМАЦИОННОЙ ПОДВИЖНОСТИ МОЛЕКУЛЫ ДНК

Для математического описания конформационной подвижности некоторого фрагмента молекулы ДНК, которое включало бы возможность образования открытых состояний, воспользуемся системой из 2N связанных нелинейных дифференциальных уравнений, имитирующих угловые отклонения азотистых оснований в обеих полинуклеотидных цепочках [15]:

$$\begin{aligned}
 & I_{n,1} \frac{d^2 \varphi_{n,1}(t)}{dt^2} - K'_{n,1} [\varphi_{n+1,1}(t) - 2\varphi_{n,1}(t) + \varphi_{n-1,1}(t)] + \\
 & + k_{n,1-2} R_{n,1} (R_{n,1} + R_{n,2}) \sin \varphi_{n,1} - k_{n,1-2} R_{n,1} R_{n,2} \sin(\varphi_{n,1} - \varphi_{n,2}) = \\
 & = -\beta_{n,1} \frac{d\varphi_{n,1}(t)}{dt} + M_0,
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 & I_{n,2} \frac{d^2 \varphi_{n,2}(t)}{dt^2} - K'_{n,2} [\varphi_{n+1,2}(t) - 2\varphi_{n,2}(t) + \varphi_{n-1,2}(t)] + \\
 & + k_{n,1-2} R_{n,2} (R_{n,1} + R_{n,2}) \sin \varphi_{n,2} - k_{n,1-2} R_{n,1} R_{n,2} \sin(\varphi_{n,2} - \varphi_{n,1}) = \\
 & = -\beta_{n,2} \frac{d\varphi_{n,2}(t)}{dt} + M_0.
 \end{aligned} \tag{2}$$

Здесь  $\varphi_{n,i}(t)$  – угловое отклонение n-го азотистого основания i-й цепочки;  $I_{n,i}$  – момент инерции n-го азотистого основания i-й цепочки;  $R_{n,i}$  – расстояние от центра масс n-го азотистого основания i-й цепочки до сахаро-фосфатного остова;  $K'_{n,i} = KR_{n,i}^2$ ; K – константа,

характеризующая жесткость (на растяжение) сахаро-фосфатного остова;  $\beta_{n,i} = \alpha R_{n,i}^2$ ;  $\alpha$  – коэффициент диссипации;  $k_{n,i}$  – константа, характеризующая взаимодействие между основаниями внутри пар;  $M_0$  – постоянный внешний момент;  $i = 1, 2$ ;  $n = 1, 2, \dots, N$ , N – число

пар оснований в рассматриваемом фрагменте. Уравнения (1) и (2) упрощаются, если усреднить коэффициенты  $I_{n,i}$ ,  $R_{n,i}$ ,  $K_{n,i}$ ,  $k_{n,1-2}$  и  $\beta_{n,i}$  методом концентраций [16]:

$$\begin{aligned} I_{n,i} &\rightarrow \bar{I}_i = I_A C_{A,i} + I_T C_{T,i} + I_G C_{G,i} + I_C C_{C,i}, \\ R_{n,i} &\rightarrow \bar{R}_i = R_A C_{A,i} + R_T C_{T,i} + R_G C_{G,i} + R_C C_{C,i}, \\ K_{n,i} &\rightarrow \bar{K}_i = K_A C_{A,i} + K_T C_{T,i} + K_G C_{G,i} + K_C C_{C,i}, \\ k_{n,1-2} &\rightarrow \bar{k}_{1-2} = k_{A-T} (C_{A,1} + C_{T,2}) + k_{G-C} (C_{G,1} + C_{C,2}), \\ \beta_{n,i} &\rightarrow \bar{\beta}_i = \beta_A C_{A,i} + \beta_T C_{T,i} + \beta_G C_{G,i} + \beta_C C_{C,i}, \end{aligned} \quad (3)$$

где  $C_{j,i} = N_{j,i}/N$  – концентрация оснований  $j$ -го типа ( $j = A, T, G, C$ ) в  $i$ -й цепи ( $i = 1, 2$ );  $N_{j,i}$  – количество азотистых оснований  $j$ -го типа в  $i$ -й последовательности.

После проведения процедуры усреднения (3) уравнения (1) и (2) преобразуются к виду:

$$\begin{aligned} \bar{I}_1 \frac{d^2 \varphi_{n,1}(t)}{dt^2} - \bar{K}_1 [\varphi_{n+1,1}(t) - 2\varphi_{n,1}(t) + \varphi_{n-1,1}(t)] + \\ + \bar{k}_{1-2} \bar{R}_1 (\bar{R}_1 + \bar{R}_2) \sin \varphi_{n,1} - \bar{k}_{1-2} \bar{R}_1 \bar{R}_2 \sin(\varphi_{n,1} - \varphi_{n,2}) = \\ = -\bar{\beta}_1 \frac{d\varphi_{n,1}(t)}{dt} + M_0, \end{aligned} \quad (4)$$

$$\begin{aligned} \bar{I}_2 \frac{d^2 \varphi_{n,2}(t)}{dt^2} - \bar{K}_2 [\varphi_{n+1,2}(t) - 2\varphi_{n,2}(t) + \varphi_{n-1,2}(t)] + \\ + \bar{k}_{1-2} \bar{R}_2 (\bar{R}_1 + \bar{R}_2) \sin \varphi_{n,2} - \bar{k}_{1-2} \bar{R}_1 \bar{R}_2 \sin(\varphi_{n,2} - \varphi_{n,1}) = \\ = -\bar{\beta}_2 \frac{d\varphi_{n,2}(t)}{dt} + M_0. \end{aligned} \quad (5)$$

В континуальном приближении эти уравнения приобретают следующий вид:

$$\begin{aligned} \bar{I}_1 \varphi_{1tt} - \bar{K}_1 a^2 \varphi_{1zz} + \bar{k}_{1-2} \bar{R}_1 (\bar{R}_1 + \bar{R}_2) \sin \varphi_1 - \bar{k}_{1-2} \bar{R}_1 \bar{R}_2 \sin(\varphi_1 - \varphi_2) = \\ = -\bar{\beta}_1 \varphi_{1t} + M_0, \end{aligned} \quad (6)$$

$$\begin{aligned} \bar{I}_2 \varphi_{2tt} - \bar{K}_2 a^2 \varphi_{2zz} + \bar{k}_{1-2} \bar{R}_2 (\bar{R}_1 + \bar{R}_2) \sin \varphi_2 - \bar{k}_{1-2} \bar{R}_1 \bar{R}_2 \sin(\varphi_2 - \varphi_1) = \\ = -\bar{\beta}_2 \varphi_{2t} + M_0. \end{aligned} \quad (7)$$

С помощью приближения среднего поля [17] два связанных уравнения (6) и (7) трансформируются в два независимых уравнения:

$$\bar{I}_1 \varphi_{1tt} - \bar{K}_1 a^2 \varphi_{1zz} + \bar{k}_{1-2} \bar{R}_1^2 \sin \varphi_1 = -\bar{\beta}_1 \varphi_{1t} + M_0, \quad (8)$$

$$\bar{I}_2 \varphi_{2tt} - \bar{K}_2 a^2 \varphi_{2zz} + \bar{k}_{1-2} \bar{R}_2^2 \sin \varphi_2 = -\bar{\beta}_2 \varphi_{2t} + M_0. \quad (9)$$

Односолитонные решения этих уравнений – кинки:

$$\begin{aligned} \varphi_{k,1}(z,t) &= 4 \operatorname{arctg} \left\{ \exp \left[ (\gamma_{k,1} / d_{k,1}) (z - v_{k,1} t - z_{0,1}) \right] \right\}, \\ \varphi_{k,2}(z,t) &= 4 \operatorname{arctg} \left\{ \exp \left[ (\gamma_{k,2} / d_{k,2}) (z - v_{k,2} t - z_{0,2}) \right] \right\}. \end{aligned} \quad (10)$$

Кинки используют для имитации открытых состояний [18, 19]. Здесь  $v_{k,i}$  – скорость движения кинка в  $i$ -й цепочке ( $i = 1, 2$ ),  $\bar{C}_i = (\bar{K}_i a^2 / \bar{I}_i)^{1/2}$  – скорость звука в  $i$ -й цепочке,  $d_{k,i} = (\bar{K}_i a^2 / \bar{V}_i)^{1/2}$  – размер кинка в  $i$ -й цепочке,  $\bar{V}_i = \bar{k}_{1-2} \bar{R}_i^2$ ,  $\gamma_{k,i} = (1 - v_{k,i}^2 / \bar{C}_i^2)^{-1/2}$  – фактор Лоренца,

Минимальные значения энергии кинка  $E_{0,i}$ , рассчитанные для шести участков основной и комплементарной последовательностей плазмиды pPF1

Номер участка	Координаты участка, п.о.	$E_{0,1}$ ( $10^{-17}$ Дж)	$E_{0,2}$ ( $10^{-17}$ Дж)
1 (Egfp)	333–1049	0.1639	0.1658
2	1050–1133	0.1566	0.1566
3 (mCherry)	1134–1841	0.1673	0.1673
4	1842–2568	0.1585	0.1585
5 (Kan)	2569–3381	0.1596	0.1596
6	(3382–5557) + (1–332)	0.1631	0.1631

Полная энергия кинка в  $i$ -й цепочке  $E_i$  определяется следующей формулой [20]:

$$E_i = \frac{E_{0,i}}{\sqrt{1 - \frac{v_{k,i}^2}{\bar{C}_i^2}}}, \quad (11)$$

где  $E_{0,i}$  – минимальное значение энергии, необходимое для образования кинка:

$$E_{0,i} = 8\sqrt{\bar{K}_i \bar{V}_i}, \quad (i = 1, 2). \quad (12)$$

В случае малых скоростей ( $v_{k,i} \ll \bar{C}_i$ ) формула (11) приобретает вид:

$$E_i = \frac{E_{0,i}}{\sqrt{1 - \frac{v_{k,i}^2}{\bar{C}_i^2}}} = E_{0,i} + \frac{m_{k,i} v_{k,i}^2}{2} \quad (13)$$

где  $m_{k,i} = \frac{E_{0,i}}{2\bar{C}_i^2}$  – масса кинка в  $i$ -й цепочке.

Формула (13) свидетельствует о том, что кинк можно моделировать как квазичастицу, обладающую массой  $m_{k,i}$ , скоростью  $v_{k,i}$  и обладающую потенциальной энергией  $E_{0,i}$ . В следующем разделе мы воспользуемся квазиоднородным приближением и формулой (12) для расчета потенциальной энергии кинка внутри каждого из участков.

### ЭНЕРГЕТИЧЕСКИЙ ПРОФИЛЬ ПЛАЗМИДЫ pPF1

При построении энергетического профиля плазмиды pPF1 использовались значения коэффициентов уравнений (1) и (2) из работы [12] и

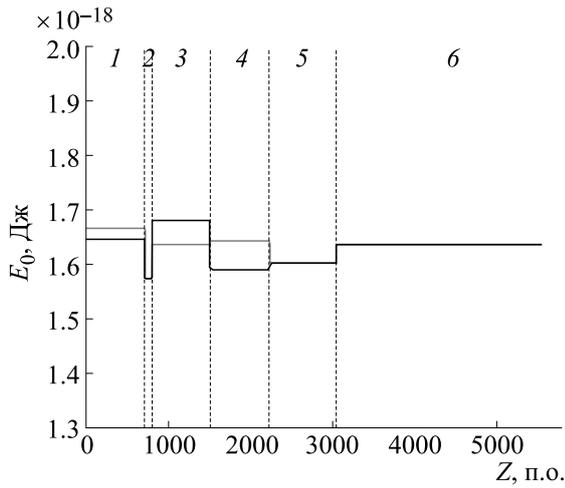
формула (12) для расчета минимальных значений энергии  $E_{0,i}$ , необходимых для образования кинка в каждом из шести участков  $i$ -й цепочки ДНК ( $i = 1, 2$ ): в трех участках, отвечающих генам (Egfp, mCherry, Kan) и трех участках, отвечающих промежуточным областям (второй, четвертый и шестой участки). Результаты расчета приведены в таблице. Построенные энергетические профили для основной и комплементарной последовательностей плазмиды pPF1 приведены на рис. 5.

Из рис. 5 видно, что самая глубокая потенциальная яма наблюдается во втором участке и основной (черная кривая), и комплементарной (серая кривая) последовательностей. Это говорит о том, что вероятность образования кинка во втором участке выше, чем в других. Видно также, что во втором участке черная и серая кривые совпадают. Это означает, что энергии активации кинка в основной и в комплементарной последовательностях одинаковые. Таким образом, с энергетической точки зрения кинку все равно, в какой из двух этих последовательностей активироваться. Какого-либо предпочтения нет.

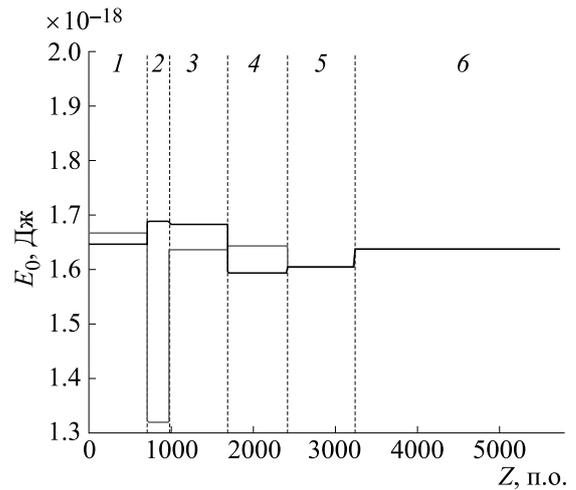
### ЭНЕРГЕТИЧЕСКИЕ ПРОФИЛИ МОДИФИЦИРОВАННЫХ ПЛАЗМИД

**Компьютерные эксперименты с однородными последовательностями, интегрированными в плазмиду pPF1.** В первом компьютерном эксперименте плазида pPF1 модифицировалась вставкой во второй участок однородной модельной последовательности polyA длиной 183 п.о. Схематическое изображение последовательности второго участка со вставкой представлено на рис. 6.

Результаты расчета энергетических профилей основной и комплементарной последовательности



**Рис. 5.** Энергетический профиль основной (черная кривая) и комплементарной (серая кривая) последовательностей плазмиды pPF1. Цифрами 1–6 обозначены участки последовательностей. Вертикальные пунктирные линии обозначают границы между смежными участками.



**Рис. 7.** Энергетический профиль основной (черная кривая) и комплементарной (серая кривая) последовательностей модифицированной плазмиды pPF1 с однородной последовательностью polyA, интегрированной во второй участок. Цифрами 1–6 обозначены участки. Вертикальные пунктирные линии обозначают границы между смежными участками.

стей модифицированной плазмиды представлены на рис. 7.

Из рис. 7 видно, что во втором участке комплементарной последовательности polyT (серая кривая) наблюдается самая глубокая потенциальная яма, а в том же участке основной последовательности polyA (черная кривая), напротив, наблюдается барьер. Это говорит о том, что вероятность активации кинка во втором участке комплементарной последовательности выше, чем в том же участке основной последовательности.

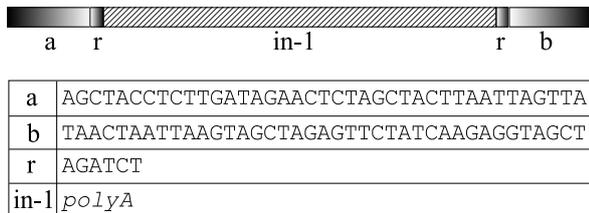
Во втором компьютерном эксперименте плазмиды pPF1 модифицировалась вставкой во второй участок другой однородной последовательности (polyG) той же длины (рис. 8).

Результаты расчета энергетических профилей основной и комплементарной последовательностей модифицированной плазмиды pPF1 с инте-

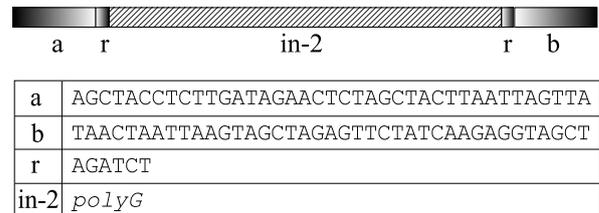
грированной в нее последовательностью polyG представлены на рис. 9.

Из рис. 9 видно, что самая глубокая потенциальная яма наблюдается во втором участке комплементарной последовательности polyC (серая кривая). В этом же участке основной последовательности polyG (черная кривая), напротив, наблюдается барьер. Таким образом, наибольшая вероятность локального расплетания ДНК будет наблюдаться в комплементарной последовательности второго участка.

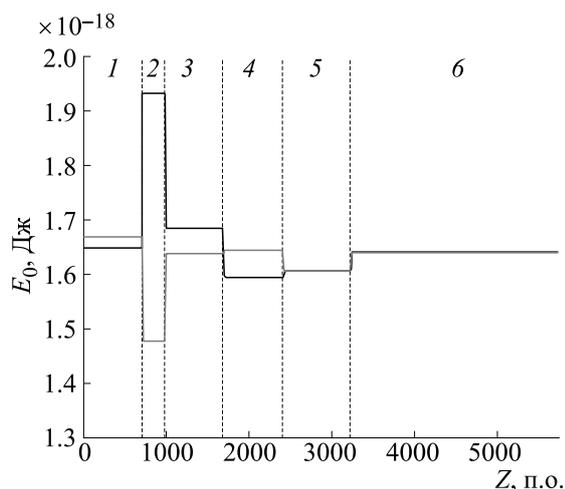
**Компьютерные эксперименты с неоднородными последовательностями, интегрированными в плазмиду pPF1.** В следующем компьютерном эксперименте плазмиды pPF1 модифицировалась вставкой во второй участок небольшого фрагмента реальной последовательности *Escherichia coli*, взятого из так называемых промоторных островков – областей высоко насыщенных промотор-



**Рис. 6.** Схематическое изображение последовательности второго участка с интегрированной в него последовательностью polyA.



**Рис. 8.** Схематическое изображение последовательности второго участка плазмиды pPF1 с интегрированной в него модельной однородной последовательностью polyG.



**Рис. 9.** Энергетический профиль основной (черная кривая) и комплементарной (серая кривая) последовательностей модифицированной плазмиды pPF1, с однородной последовательностью *polyG*, интегрированной во второй участок. Цифрами 1–6 обозначены участки. Вертикальные пунктирные линии обозначают границы между смежными участками.

подобными участками [14]. На рис. 10 представлена схема второго участка после модификации.

Результаты расчета энергетического профиля основной и комплементарной последовательностей модифицированной вышеуказанным способом плазмиды pPF1 представлены на рис. 11.

Из рис. 11 видно, что самая глубокая потенциальная яма приходится на второй участок основной последовательности (черная кривая). Это говорит о том, что наибольшая вероятность активации кинка будет наблюдаться во втором участке основной последовательности, соответствующей нематричной нити ДНК. В данном случае это можно трактовать следующим образом: именно второй участок ответственен за разрушение водородных связей и нуклеацию транскрипционного пузыря. Развитие нуклеации на основной последовательности указывает на направление движе-

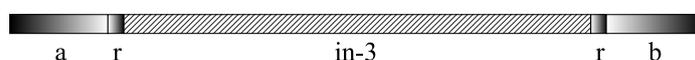
ния пузыря: от 5'- к 3'-концу. Такое движение соответствует движению в направлении гена красного белка. Можно ожидать, что в данном случае преимущественно будет синтезироваться белок mCherry.

В следующем компьютерном эксперименте во второй участок плазмиды pPF1 был встроено фрагмент последовательности обратного комплементарной к последовательности фрагмента, который встраивался во второй участок в предыдущем эксперименте. Последовательность второго участка, полученная в результате такой модификации, показана на рис. 12, а на рис. 13 представлены результаты построения энергетического профиля модифицированной плазмиды.

Из рис. 13 видно, что в отличие от предыдущего случая самая глубокая потенциальная яма приходится на второй участок комплементарной последовательности (серая кривая). Таким образом, с более высокой вероятностью кинк будет активироваться не в основной, а в комплементарной последовательности второго участка. С учетом того, что движение пузыря всегда происходит от 5'- к 3'-концу, можно ожидать, что с большей вероятностью пузырь будет двигаться в направлении гена зеленого белка, для которого в данной конструкции нематричной нитью является комплементарная последовательность. Таким образом, вероятность синтеза белка Egfp будет выше по отношению к mCherry.

## ОБСУЖДЕНИЕ И ВЫВОДЫ

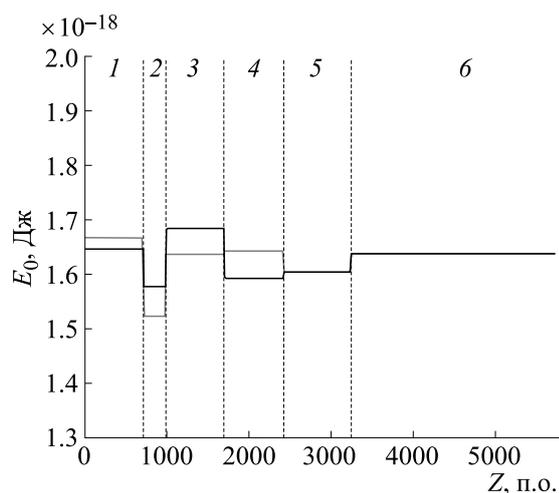
В данной работе с помощью методов нелинейной математики были рассчитаны энергетические профили потенциальных полей плазмиды pPF1 и ее модификаций. Модификации плазмиды получали вставкой во второй участок, расположенный между генами зеленого и красного белков, небольших фрагментов однородных (синтетических) и неоднородных (природных) последовательностей.



a	AGCTACCTCTTGATAGAACTCTAGCTACTTAATTAGTTA
b	TAACATAATTAAGTAGCTAGAGTTCTATCAAGAGGTAGCT
r	AGATCT
in-3	TAATCAGGATGATGTGCATTTATATTTTTATACACAAAATTATAGTTTGCA AATTTTAATAAATTTTCATTTAAGATTTAAATATATATGATGTTGTTTTTA TTCTAASCTATTTTCAAAGTTACATTTTTCAACGCTTACTATGCTTTTTAT AACATAAACTCACTACAACGCACCTGAA

**Рис. 10.** Схематическое изображение последовательности второго участка плазмиды pPF1 с интегрированным в него первым фрагментом реальной последовательности *Escherichia coli* (in-3).





**Рис. 13.** Энергетический профиль основной (черная кривая) и комплементарной (серая кривая) последовательностей модифицированной плазмиды pPF1, где во второй участок интегрирован небольшой фрагмент последовательности *Escherichia coli* (in-4). Цифрами 1–6 обозначены участки последовательностей. Вертикальные пунктирные линии обозначают границы между смежными участками.

жения и не учитываются продольные, поперечные и радиальные динамические степени свободы нуклеотидов. Используется приближение среднего поля, при котором явная динамика оснований одной нити описывается в среднем поле, формируемом комплементарной нитью. Не учитывается спиральность молекулы ДНК, и не рассматривается сам процесс инициации и образования транскрипционного пузыря, который формируется при взаимодействии РНК-полимеразы с ДНК. Несмотря на эти ограничения модели, она позволяет на основе расчета энергетических профилей рассматриваемых молекул ДНК сделать качественную оценку вероятности нуклеации пузыря и направления его движения, результаты которой согласуются с экспериментальными данными.

Результаты моделирования показали, что эффективность транскрипции в заданном направлении синтеза РНК зависит от энергии активации кинка в нематричной нити. Полученные данные свидетельствуют о целесообразности применения расчетного метода на основе аппарата нелинейной динамики для прогнозирования биологических характеристик природных последовательностей ДНК.

#### КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

#### СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания исследований с использованием людей и животных в качестве объектов.

#### СПИСОК ЛИТЕРАТУРЫ

1. Map of plasmid pET-28b [https://www.snapgene.com/resources/plasmid-files/?set=pet\\_and\\_duet\\_vectors\\_\(novagen\)&plasmid=pET-28b\(%2B\)](https://www.snapgene.com/resources/plasmid-files/?set=pet_and_duet_vectors_(novagen)&plasmid=pET-28b(%2B)) (дата обращения: 01.11.2020).
2. I. S. Masulis, Z. Sh. Babaeva, S. V. Chernyshov, et al., *Sci. Rep.* **5**, 11449 (2015).
3. S. Borukhov and E. Nudler, *Trends Microbiol.* **16** (3), 126 (2008).
4. J. Y. Kang, T. V. Mishanina, R. Landick, et al., *J. Mol. Biol.* **431** (20), 4007 (2019).
5. M. T. Jr. Record, W. S. Reznikoff, M. L. Craig, et al., in *Escherichia coli and Salmonella Cellular and Molecular Biology*, Ed. by F.C. Neidhardt (ASM Press, Washington, DC, USA, 1996), pp. 792–821.
6. H. Boyaci, J. Chen, R. Jansen, et al., *Nature* **565** (7739), 382 (2019).
7. J. Chen, C. Chiu, S. Gopalkrishnan, et al., *Mol Cell.* **78** (2), 275 (2020).
8. E. F. Ruff, M. T. Jr. Record, and I. Artsimovitch. *Biomolecules* **5** (2), 1035 (2015).
9. J. T. Winkelman, I. O. Vvedenskaya, Y. Zhang, et al., *Science* **351** (6277), 1090 (2016).
10. P. J. Caudrey, J. C. Eilbeck, and J. D. Gibbon, *Nuovo Cimento B* **25** (2), 497 (1975).
11. V. G. Ivancevic and T. T. Ivancevic, *J. Geometry and Symmetry in Physics* **31**, 1 (2013).
12. A. A. Grinevich, A. A. Ryasik, and L. V. Yakushevich, *Chaos, Solitons & Fractals* **75**, 62 (2015).
13. L. V. Yakushevich and L.A. Krasnobaeva, *Math. Biol. Bioinform.* **14** (1), 327 (2019).
14. K. S. Shavkunov, I. S. Masulis, M. N. Tutukina, et al., *Nucl. Acids Res.* **37** (15), 4919 (2009).
15. L. A. Krasnobaeva and L. V. Yakushevich, *J. Bioinform. Comput. Biol.* **13** (1), 1540002 (2015).
16. L. V. Yakushevich and L. A. Krasnobaeva, *Int. J. Non-linear Mech.* **43**, 1074 (2008).
17. L. V. Yakushevich, *J. Biol. Phys.* **43**, 113 (2017).
18. W. Englander, N. R. Kallenbach, A. J. Heeger, et al., *Proc. Natl. Acad. Sci. USA* **77**, 7222 (1980).
19. Л. В. Якушевич, Л. А. Краснобаева, А. В. Шаповалов и др., *Биофизика* **50**, 450 (2005).
20. Л. В. Якушевич и А. А. Рясик, *Компьютерные исследования и моделирование* **4** (1), 209 (2012).
21. Y. Zhang, Y. Feng, S. Chatterjee, et al., *Science* **338** (6110), 1076 (2012).
22. E. Heyduk, K. Kuznedelov, K. Severinov, et al. *J. Biol. Chem.* **281** (18), 12362 (2006).
23. M. E. Karpen and P. L. deHaseth, *Biomolecules* **5** (2), 668 (2015).
24. N. S. Roy, S. Debnath, A. Chakraborty, et al., *Phys. Chem. Chem. Phys.* **20** (14), 9449 (2018).
25. F. Colizzi, C. Perez-Gonzalez, R. Fritzen, et al., *Proc. Natl. Acad. Sci. USA* **116** (45), 22471 (2019).

**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**

**A.A. Grinevich, I.S. Masulis, and L.V. Yakushevich**

*Institute of Cell Biophysics, Russian Academy of Sciences,  
Institutskaya ul. 3, Pushchino, Moscow Region, 142290 Russia*

In this study, we use the kink solutions of the nonlinear sine-Gordon equation to apply the methods of mathematical modeling to investigate the dynamics of the transcription bubble in the pPF1 plasmid. Based on the calculated energy profile for the pPF1 plasmid and its modified versions, it was shown that the minimum potential energy of the kink formation or transcription bubble nucleation corresponds to the region between the genes of the Egfp and mCherry proteins. The insertion of homogeneous sequences into the region between Egfp and mCherry showed that the kink is more likely to be activated in polyT or polyC as compared to polyA or polyG, which indicates the dependence of nucleation of the transcription bubble on the molecular weight of base pairs. In the case of insertions into the region between Egfp and mCherry of small fragments of the native sequence of *Escherichia coli*, the model identifies DNA strands with the highest probability of nucleation of the transcription bubble, and, accordingly, determines the direction (towards the Egfp or mCherry gene) of transcription, indicating a link between the direction of transcription and the energy profile of the plasmid.

*Keywords: mathematical modeling, nonlinear sine-Gordon equation, kink, solitons, pPF1 plasmid, DNA, transcription bubble*