

УДК:577.0; 57.036

Случайные блуждания по псевдоспецифическим состояниям как модель облегченной диффузии ДНК-связывающего белка

©2015 Сибатов Р.Т., Саенко Ю.В.

Ульяновский государственный университет

Аннотация. Рассматривается облегченная диффузия регуляторного белка в процессе поиска специфического сайта-мишени на молекуле ДНК. На основе гипотезы о баллистическом движении между рассеяниями на псевдоспецифических участках предложена феноменологическая модель одномерного движения белка. Сочетание этого подхода с моделью дискретных случайных блужданий позволила получить зависимости среднего времени первого достижения от энергии связи, среднего пробега между псевдоспецифическими состояниями, скорости белка и температуры. Выявлены температурный и энергетический оптимумы, при которых достигаются минимальные значения среднего времени поиска.

Ключевые слова: облегченная диффузия, фактор транскрипции, ДНК, случайное блуждание, время первого достижения.

1. Введение

Экспрессия генов в живых клетках управляется регуляторными белками, которые связываются со специфичными участками молекулы ДНК (сайтами), стимулируя или подавляя, тем самым, транскрипцию гена [1, 2, 3, 4]. Регуляторный белок в этом случае называют фактором транскрипции (ФТ). Один из важнейших вопросов молекулярной биологии состоит в том, каким образом ДНК-связывающие белки осуществляют поиск сайта на ДНК. Как показывают наблюдения, времена поиска много меньше предсказываемых теорией диффузии, предполагающей трехмерное блуждание белка в окружающем молекулу ДНК пространстве. Многочисленные экспериментальные и теоретические исследования указывают на то, что движение белка происходит в режиме комбинированной диффузии: трехмерное блуждание белка чередуется с квазиодномерным движением вдоль молекулы ДНК. Такое поведение впервые предложено Бергом и вон Хиппелем [5] существенно уменьшает время нахождения и связывания белка с мишенью на ДНК. Этот процесс был назван *облегченной диффузией*. Впоследствии вводились дополнительные концепции, в частности, обсуждались межсегментные переходы белков на гибкой ДНК молекуле, предложена модель «притягивающей антенны» сайта, рассмотрены неупорядоченные профили потенциала взаимодействия между участками ДНК и белком [6, 7, 8, 9]. В ряде работ [10, 11, 12, 13, 14, 15] отмечается необходимость учета трехмерной структуры ДНК, субдиффузии белка, нелинейной зависимости темпа поиска от концентрации ФТ вблизи ДНК, недиффузионного характера продольного движения.

¹ren_sib@bk.ru

²saenkoyv@yandex.ru

В работах [16, 8, 17] обсуждается влияние неоднородности нуклеотидной последовательности, которая приводит к неупорядоченному профилю свободной энергии для движения белка вдоль ДНК. В модели из [18] постулируется возможность двух состояний для белка, связанного с ДНК: неспецифическая мода с постоянной энергией связи ϵ_{ns} и нуклеотидно-зависимая специфическая мода с энергией связи $\epsilon > \epsilon_{ns}$. Профиль свободной энергии состоит преимущественно из участков неспецифического взаимодействия с энергией ϵ_{ns} с небольшим количеством ловушек с $\epsilon > \epsilon_{ns}$.

В работе [15] исследуется обобщенная модель облегченной диффузии, в которой чередуются два конформационных состояния в процессе движения вдоль ДНК: поиск и распознавание. ФТ, находясь в состоянии поиска, совершает одномерное блуждание с относительно высоким коэффициентом диффузии. Стохастически ФТ переключается в состояние распознавания, которое характеризуется значительно меньшим коэффициентом диффузии. Именно в этом состоянии ФТ может распознать и связаться с сайтом-мишенью. Расчеты [15] показывают, что в новой модели для типичных параметров, скорость ассоциации с целью меньше по сравнению с предыдущими моделями. Кажется правдоподобным, что более эффективной стратегия поиска будет не при случайном переключении в режим распознавания, а при реализации механизма, при котором распознавание происходит на псевдоспецифических участках (близких по нуклеотидной последовательности к сайту мишени) и не затрачивается время на диффузию на интервалах с малой энергией связи. Феноменологическая модель такого механизма и будет рассмотрена в данной работе.

2. Рассеяние на псевдоспецифических участках

Прямое наблюдение случайного блуждания белка, помеченного флуоресцентной меткой, вдоль ДНК даёт ценную информацию о процессе поиска. Например, на рис. 1 приведены четыре экспериментальные траектории белка р53 вдоль молекулы ДНК, оцифрованные из работ [8, 19]. С помощью этих траекторий построены гистограммы продольных пробегов между точками поворота, квантиль-квантильная диаграмма для смещений между точками поворота, а также гистограмма времени между точками поворота и скорости. Средний пробег и средняя скорость равны $l \approx 0.16$ мкм и $\langle v \rangle \approx 3.07$ мкм/с, соответственно. Квантиль-квантильная диаграмма свидетельствует о том, что смещения между точками поворота не распределены по гауссову закону, присутствует избыточная доля достаточно больших пробегов, т.е. процесс продольного блуждания в строгом смысле нельзя моделировать броуновским движением. Относительно большие значения среднего пробега, негауссов характер смещений между точками поворота указывают на то, что участки, характеризующиеся большой энергией связи с белком (*псевдоспецифические состояния*, ПС), могут служить точками поворота скользящего по инерции белка. Поскольку гистограмма (б) близка к экспоненциальной функции (это следует из полулогарифмического масштаба), то можно считать что эти состояния представляют собой одномерный пуассоновский ансамбль. Тот факт, что только ПС служат точками поворота позволяет не затрачивать время на процесс распознавания на неспецифических участках.

Рассмотрим феноменологическую модель механизма рассеяния на псевдоспецифических участках. Пусть энергия связи белок-ДНК распределена с некоторой плотностью $f(\epsilon)$. Предполагаем, что существует некоторый транспортный уровень ϵ_m , превышая который белок приобретает поступательную кинетическую энергию продольного движения $\epsilon_{kin} = \epsilon - \epsilon_m$, где ϵ — энергия белка. Будучи локализованным белок имеет энергию ϵ_t . Делокализация предполагается тепловой, так

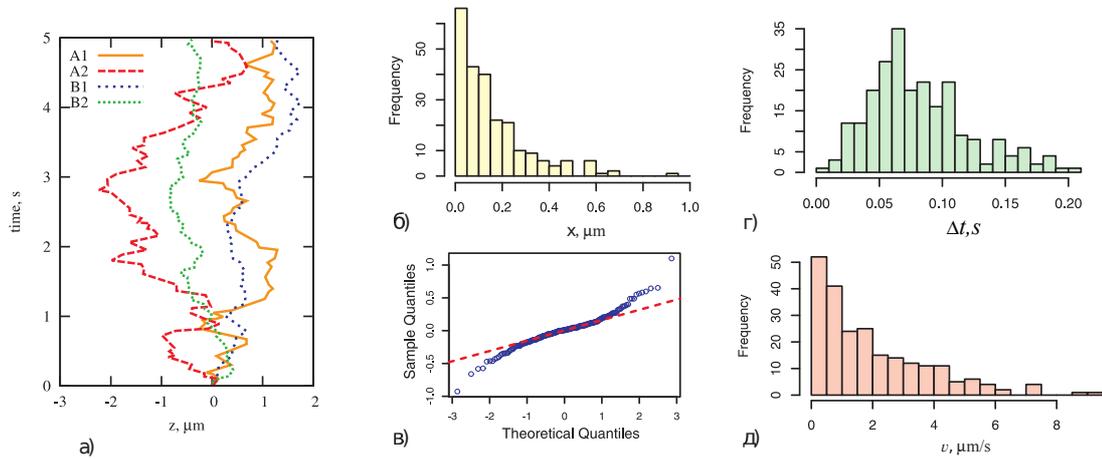


Рис. 1. а) Пространственно-временные траектории блуждания белка р53 вдоль молекулы ДНК (данные оцифрованы из работ [8, 19]). б) Гистограмма свободных пробегов вдоль ДНК между точками поворота, средний пробег $l \approx 0.16$ мкм. в) Квантиль-квантильная диаграмма для смещений белка в сравнении с нормальным распределением. Хорошо описывается нормальным распределением в диапазоне $(-0.5$ мкм, 0.5 мкм).

что приращение энергии имеет вид

$$P\{\Delta E > \epsilon\} = \exp(-\epsilon/\gamma kT),$$

где γ – некоторый безразмерный множитель, при этом темп делокализации

$$\nu = \nu_0 \exp(-\epsilon_t/kT)$$

определяет среднее время нахождения белка в локализованном состоянии.

Распределение кинетической энергии тоже имеет экспоненциальный вид,

$$P\{E_{\text{kin}} > \epsilon_{\text{kin}}\} = P\{\Delta E > \epsilon_{\text{kin}} + \epsilon_t | \Delta E > \epsilon_t\} = \exp(-\epsilon_{\text{kin}}/\gamma kT).$$

Предполагаем, что уменьшение импульса белка происходит за счет взаимодействия белок-ДНК,

$$\frac{dmv}{dt} = -f(\epsilon_t).$$

Тогда диссипация кинетической энергии при прохождении расстояния $dx = vdt$ определяется соотношением

$$\frac{d(mv^2/2)}{dx} = -f(\epsilon_t).$$

Таким образом, функцию f можно интерпретировать как потери кинетической энергии направленного движения белка на единицу длины ДНК.

Плотность распределения «силы трения» f связана с плотностью распределения энергии связности ДНК-белок соотношением

$$p(f) = D(\epsilon_t(f)) \frac{d\epsilon_t(f)}{df},$$

где $D(\epsilon)$ – плотность состояний. Выберем ее в виде экспоненты $D(\epsilon) = \epsilon_0^{-1} \exp(-\epsilon/\epsilon_0)$, а зависимость «силы трения» f от энергии связи в виде возрастающей функции с характерным параметром ϵ_1 .

Средний пробег определяется средней кинетической энергией и средней диссипацией:

$$l = \frac{\langle E_{\text{kin}} \rangle}{\langle f \rangle} = \langle f \rangle^{-1} \left[\gamma k T - (\gamma k T + \epsilon_{\text{off}}) \exp\left(-\frac{\epsilon_{\text{off}}}{\gamma k T}\right) \right]. \quad (1)$$

Здесь ϵ_{off} – энергия диссоциации, тогда соответствующая вероятность диссоциации

$$\eta = P(E_{\text{kin}} > \epsilon_{\text{off}}) = \exp(-\epsilon_{\text{off}}/\gamma k T). \quad (2)$$

Средняя кинетическая энергия определяется с помощью интеграла

$$\langle E_{\text{kin}} \rangle = \int_0^{\epsilon_{\text{off}}} \frac{\epsilon \exp(-\epsilon/\gamma k T)}{k T [1 - \exp(-\epsilon_{\text{off}}/\gamma k T)]} d\epsilon.$$

Для нахождения $\langle f \rangle$ необходимо знать зависимость потерь энергии от ϵ_t . Рассмотрим две модельных функции

$$f(\epsilon_t) = f_0 [\exp(\epsilon_t/\epsilon_1) - 1]. \quad (3)$$

$$f(\epsilon_t) = f_0 [1 - \exp(-\epsilon_t^2/\epsilon_1^2)]. \quad (4)$$

В первом случае приходим к плотности распределения силы трения в виде

$$p(f) = \frac{\epsilon_1}{f_0 \epsilon_0} \left(1 + \frac{f}{f_0} \right)^{-1 - \frac{\epsilon_1}{\epsilon_0}},$$

и при $d = \epsilon_1/\epsilon_0 > 1$:

$$\langle f \rangle = \frac{f_0}{d - 1}, \quad d = \frac{\epsilon_1}{\epsilon_0}. \quad (5)$$

Во втором случае плотность распределения диссипации энергии

$$p(f) = \frac{\epsilon_1}{2\epsilon_0(f_0 - f) \ln\left(\frac{f_0}{f_0 - f}\right)} \exp\left(-\frac{\epsilon_1}{\epsilon_0} \sqrt{\ln\frac{f_0}{f_0 - f}}\right), \quad 0 \leq f \leq f_0,$$

и среднее значение диссипации энергии на единицу длины

$$\langle f \rangle = f_0 \left[1 - \frac{\sqrt{\pi} \epsilon_1}{2\epsilon_0} \exp\left(\frac{\epsilon_1^2}{4\epsilon_0^2}\right) \operatorname{erfc}\left(\frac{\epsilon_1}{2\epsilon_0}\right) \right]. \quad (6)$$

Из формулы (1) следует, что с ростом температуры средний пробег между точками поворота увеличивается, но при этом ограничен сверху значением

$$l_{\text{max}} = \lim_{T \rightarrow \infty} l(T) = \frac{\epsilon_{\text{off}}}{2\langle f \rangle}.$$

Связано это с тем, что при большой температуре вероятность диссоциации η близка к 1, а E_{kin} является равномерно распределенной величиной в интервале $(0, \epsilon_{\text{off}})$.

3. Время первого достижения

С помощью предложенной феноменологической модели оценим зависимость среднего времени первого достижения θ от физических параметров. Для расчета θ расширим применимость результатов дискретной модели [14, 21] облегченной диффузии белка. Авторы [14, 21] рассматривают одиночный белок, который блуждает вдоль одномерной молекулы ДНК или окружающем растворе. Молекула ДНК содержит N узлов ($i = 1, 2, \dots, N$). Один из узлов с номером $i = m$ является мишенью. Белок симметрично блуждает вдоль ДНК с темпом u . Белок может отделяться от ДНК с темпом k_{off} . Интересное предположение авторов [14, 21], неочевидное, но значительно упрощающее расчет заключается в том, что белок из состояния 3D-диффузии может быть захвачен любым узлом ДНК с равной вероятностью с темпом k_{on} . Оно основано на представлении о сложной трехмерной архитектуре ДНК и достаточно быстрой диффузии в растворе.

Для лучшего понимания модели полезно записать уравнение для эволюции плотности распределения $w_n(t)$ времени первого достижения [21]:

$$\frac{dw_n(t)}{dt} = u[w_{n+1}(t) - w_n(t)] + u[w_{n-1}(t) - w_n(t)] + k_{\text{off}}[w_{3d}(t) - w_n(t)].$$

Для граничных узлов оно принимает вид:

$$\frac{dw_1(t)}{dt} = u[w_2(t) - w_1(t)] + k_{\text{off}}[w_{3d}(t) - w_1(t)]$$

$$\frac{dw_N(t)}{dt} = u[w_{N-1}(t) - w_N(t)] + k_{\text{off}}[w_{3d}(t) - w_N(t)].$$

Переход в состояние трехмерной диффузии и адсорбция молекулой ДНК учитывается соотношением

$$\frac{dw_{3d}(t)}{dt} = \frac{k_{\text{on}}}{N} \sum_{n=1}^N [w_n(t) - w_{3d}(t)].$$

С помощью преобразования Лапласа авторы [21] находят решение для трансформанты $\tilde{w}_n(s) = \int_0^\infty w_n(t)e^{-st}dt$ и приходят к выражению для среднего времени поиска мишени, расположенной в узле m , с помощью формулы $\theta_m = -N^{-1} \sum_{n=1}^N \tilde{w}'_n(s)$. Перепишем θ_m в виде (подробнее см. [21]):

$$\theta_m = \left(\frac{1}{k_{\text{off}}} + \tau_{3d} \right) (\xi_m(y) - 1),$$

где

$$\xi_m(y) = N \frac{(1-y)(1+y^m)(1+y^{N+1-m})}{(1+y)(1-y^{N+1})}, \quad y = 1 + \frac{k_{\text{off}}}{2u} - \sqrt{\left(1 + \frac{k_{\text{off}}}{2u}\right)^2 - 1}.$$

Усредняя по возможным положениям мишени $\theta = N^{-1} \sum_{m=1}^N \theta_m$. При достаточно большом N ($N > 100$): $\theta \approx \theta_{N/2}$. Недостатком модели [14, 21] является фиксированное распределение пробегов (экспоненциальное) между поворотами с неизменным средним значением пробега и отсутствие скорости продольного движения (u — темп перехода между узлами, измеряемый в c^{-1}).

Включить в модель среднюю длину пробега, скорость движения между узлами и, в конечном счете, зависимость от температуры можно с привлечением модели,

предложенной в параграфе 2, если считать, что рассматриваемые узлы представляют собой псевдоспецифические участки. Тогда

$$N = \frac{L}{l}, \quad u = \frac{v}{l}, \quad \eta = k_{\text{off}} \frac{l}{v}. \quad (7)$$

Среднее время поиска,

$$\theta(T) = \left(\frac{1}{k_{\text{off}}(T)} + \tau_t(T) + \tau_{3d}(T) \right) (\xi(y, T) - 1), \quad (8)$$

$$\xi(y, T) = \frac{L(1-y) [1 + y^{(1+L/l)/2}]^2}{l(1+y)(1-y^{1+L/l})}, \quad y = 1 + \frac{\eta(T)}{2} - \sqrt{\left(1 + \frac{\eta(T)}{2}\right)^2 - 1}. \quad (9)$$

Здесь $\tau_t(T)$ – среднее время локализации на псевдоспецифическом участке.

Средний пробег и среднеквадратическая скорость

$$l(T) = \langle E_{\text{kin}} \rangle \langle f \rangle^{-1}, \quad v(T) = \sqrt{2 \langle E_{\text{kin}} \rangle / m} \quad (10)$$

определяются средней кинетической энергией

$$\langle E_{\text{kin}} \rangle = \gamma kT - \frac{\epsilon_{\text{off}}}{\exp\left(\frac{\epsilon_{\text{off}}}{\gamma kT}\right) - 1}. \quad (11)$$

Остальные параметры

$$\eta(T) = \exp(-\epsilon_{\text{off}}/\gamma kT), \quad k_{\text{off}}(T) = \frac{v(T)\eta(T)}{l(T)},$$

$$\tau_t = \nu_0^{-1} \exp\left(\frac{\epsilon_0}{kT}\right), \quad \tau_{3d}(T) = a \sqrt{\frac{m}{\gamma kT}}. \quad (12)$$

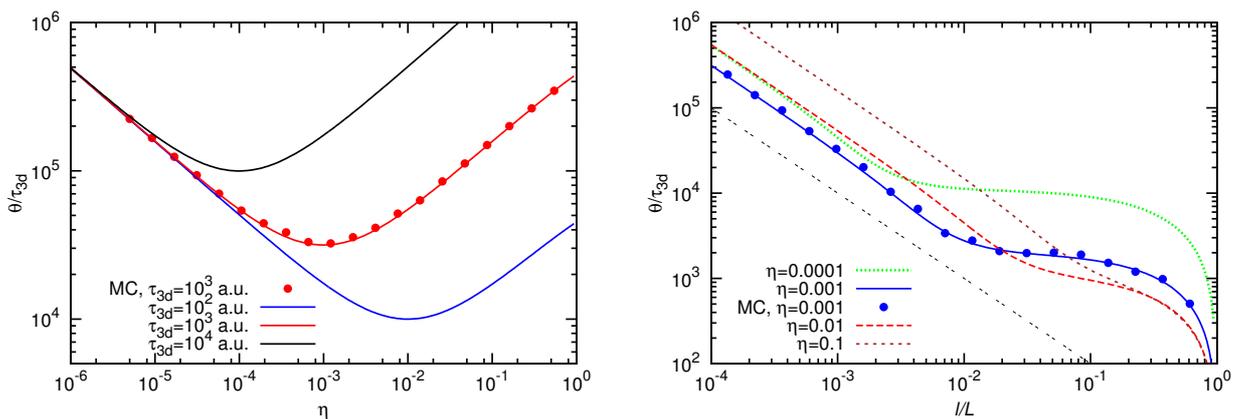


Рис. 2. Зависимость среднего времени первого достижения от вероятности диссоциации η (слева) и средней длины пробега l между псевдоспецифическими состояниями (справа) ($L = 10^6$ п.о., $\gamma = 1$, $l = 10^{-4}L$, $\tau_t = 10^{-4}\tau_{3d}$).

На рис. 2–4 представлены зависимости среднего времени первого достижения от вероятности диссоциации η , средней длины пробега l , температуры и энергии диссоциации ϵ_{off} . На основе этих зависимостей могут быть выявлены температурный

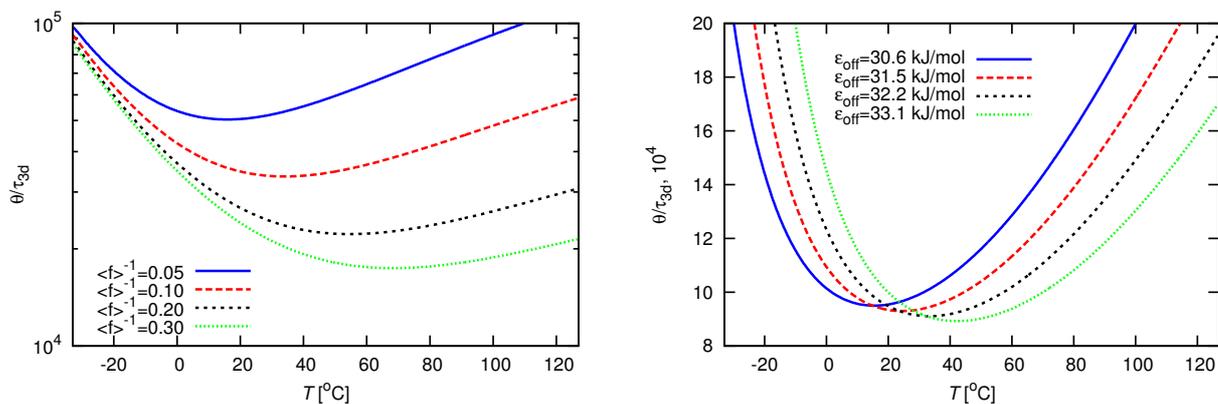


Рис. 3. Зависимость среднего времени первого достижения от температуры для различных значений $\langle f \rangle^{-1}$ (слева) и ϵ_{off} (справа) ($L = 10^8$ п.о., $\gamma = 1$).

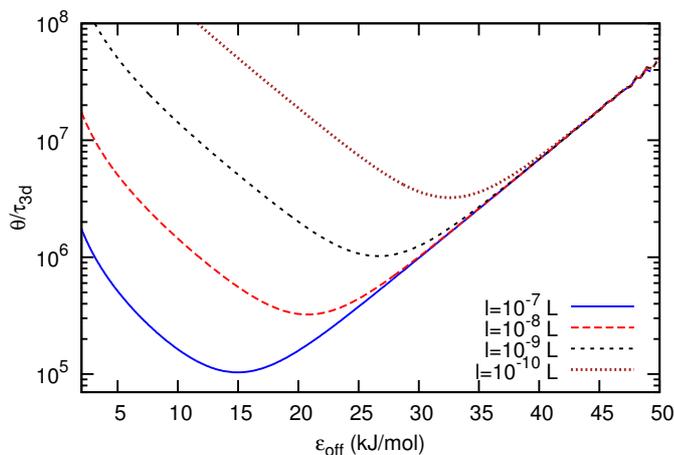


Рис. 4. Зависимость среднего времени первого достижения от энергии диссоциации ϵ_{off} для различных значений среднего пробега при температуре 37 °С. Параметры: $L = 10^9$ п.о., $\gamma = 1$, $T = 310$ К, $v = 0.6$ мкм/с.

и энергетический оптимумы, при которых достигаются минимальные значения среднего времени поиска. Температурный оптимум объясняется тем, что с ростом T одновременно увеличиваются и вероятность диссоциации белок-ДНК η и средняя длина пробега l между псевдоспецифическими участками, при этом $\theta(l)$ монотонно убывает (рис. 2, справа), а зависимость $\theta(\eta)$ имеет ярко выраженный минимум (рис. 2, слева). Этот минимум проявляется и в зависимости времени поиска θ от энергии диссоциации ϵ_{off} (рис. 4).

4. Заключение

Рассмотрена облегченная диффузия регуляторного белка в процессе поиска специфического сайта-мишени на молекуле ДНК в рамках модели случайных блужданий. Характер экспериментальных траекторий [19] продольного движения белка вдоль молекулы ДНК позволил выдвинуть гипотезу о баллистическом движении между рассеяниями на псевдоспецифических участках. Траектории свидетельствуют, что эти участки не служат ловушками при наблюдаемых температурах, тепловой энергии достаточно, чтобы практически мгновенно привести белок в

продольное движение, а ее избыток расходуется на кинетическую энергию белка. Тот факт, что только псевдоспецифические состояния служат точками поворота на пространственно-временных траекториях позволяет не затрачивать время на процесс распознавания на неспецифических участках. На основе гипотезы предложена феноменологическая модель одномерного движения, в которой механизм диссипации кинетической энергии приводит к преимущественному рассеянию на участках с наибольшей энергией связи. Сочетание механизма скольжения белка с моделью дискретных случайных блужданий позволило получить зависимости среднего времени первого достижения от энергии химического взаимодействия между белком и ДНК, среднего пробега между псевдоспецифическими состояниями, скорости белка и температуры. На этих зависимостях наблюдаются температурный и энергетический оптимумы, при которых достигаются минимальные значения среднего времени поиска.

Работа выполнена при поддержке Министерства образования и науки РФ в рамках государственного задания (2014/296; 6.1617.2014/К).

СПИСОК ЛИТЕРАТУРЫ

1. Schena M., Shalon D., Davis R.W., Brown P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995. V. 270. No. 5235. P. 467–470.
2. DeRisi J.L., Iyer V.R., Brown P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*. 1997. V. 278. No. 5338. P. 680–686.
3. McAdams H.H., Arkin A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*. 1997. V. 94. No. 3. P. 814–819.
4. Elf J., Li G.-W., Xie X.S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*. 2007. V. 316. No. 5828. P. 1191–1194.
5. von Hippel P.H., Berg O. Facilitated target location in biological systems. *Journal of Biological Chemistry*. 1989. V. 264. No. 2. P. 675–678.
6. Slutsky M., Mirny L.A. Kinetics of protein-dna interaction: facilitated target location in sequence-dependent potential. *Biophysical Journal*. 2004. V. 87. No. 6. P. 4021–4035.
7. Hu T., Shklovskii B. How does a protein search for the specific site on DNA: the role of disorder. *Physical Review E*. 2006. V. 74. No. 2. P. 021903.
8. Mirny L., Slutsky M., Wunderlich Z., Tafvizi A., Leith J., Kosmrlj A. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*. 2009. V. 42. No. 43. P. 434013.
9. Loverdo C., Benichou O., Voituriez R., Biebricher A., Bonnet I., Desbiolles P. Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. *Physical Review Letters*. 2009. V. 102. No. 18. P. 188101.
10. Lomholt M.A., van den Broek B., Kalisch S.-M.J., Wuite G.J., Metzler R. Facilitated diffusion with DNA coiling. *Proceedings of the National Academy of Sciences*. 2009. V. 106. No. 20. P. 8204–8208.
11. Li G.-W., Berg O.G., Elf J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nature Physics*. 2009. V. 5. No. 4. P. 294–297.
12. Zaid I.M., Lomholt M.A., Metzler R. How subdiffusion changes the kinetics of binding to a surface. *Biophysical Journal*. 2009. V. 97. No. 3. P. 710–721.
13. Kolomeisky A.B. Physics of protein–DNA interactions: mechanisms of facilitated target search. *Physical Chemistry Chemical Physics*. 2011. V. 13. No. 6. P. 2088–2095.
14. Kolomeisky A.B., Veksler A. How to accelerate protein search on DNA: Location and dissociation. *The Journal of Chemical Physics*. 2012. V. 136. No. 12. P. 125101.
15. Bauer M., Metzler R. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophysical Journal*. 2012. V. 102. No. 10. P. 2321–2330.

16. Brackley C., Cates M., Marenduzzo D. Facilitated diffusion on mobile DNA: configurational traps and sequence heterogeneity. *Physical Review Letters*. 2012. V. 109. No. 16. P. 168103.
17. Barbi M., Place C., Popkov V., Salerno M. A model of sequence-dependent protein diffusion along DNA. *Journal of Biological Physics*. 2004. V. 30. No. 3. P. 203–226.
18. Gerland U., Moroz J. D., Hwa T. Physical constraints and functional characteristics of transcription factor–DNA interaction. *Proceedings of the National Academy of Sciences*. 2002. V. 99. No. 19. P. 12015–12020.
19. Tafvizi A., Mirny L.A., van Oijen A.M. Dancing on DNA: kinetic aspects of search processes on DNA. *Chemphyschem*. 2011. V. 12. No. 8. P. 1481–1489.
20. Park K., Kim T., Kim H. Facilitated protein-DNA binding: Theory and Monte Carlo simulation. *Bull. Korean Chem. Soc.* 2012. V. 33. No. 3. P. 971.
21. Veksler A., Kolomeisky A.B. Speed-selectivity paradox in the protein search for targets on DNA: is it real or not? *The Journal of Physical Chemistry B*. 2013. V. 117. P. 12695.

Материал поступил в редакцию 11.09.2015, опубликован 30.11.2015.