

## ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ СТРУКТУРЫ $\sigma^{70}$ -СУБЪЕДИНИЦЫ РНК-ПОЛИМЕРАЗЫ *Escherichia coli* ОТ ИОННОЙ СИЛЫ МЕТОДОМ МОЛЕКУЛЯРНОЙ ДИНАМИКИ

© 2015 г. А.П. Толстова, Е.В. Дубровин, О.Н. Королева\*

Физический факультет Московского государственного университета им. М.В. Ломоносова,  
119991, Москва, ГСП-1, Ленинские горы;

\*Химический факультет Московского государственного университета им. М.В. Ломоносова,  
119991, Москва, ГСП-1, Ленинские горы, 1/3

E-mail: [tolstova@physics.msu.ru](mailto:tolstova@physics.msu.ru)

Поступила в редакцию 03.09.15 г.

Проведено моделирование белка  $\sigma^{70}$ -субъединицы РНК-полимеразы *E. coli* (небольшого белка, входящего в состав холофермента РНК-полимеразы и отвечающего за специфичность инициации транскрипции конститутивных генов) при разных концентрациях солей. Обнаружены два варианта расположения С-концевого домена 4 белка. В одном из них, при низкой концентрации солей в растворе, домен 4 взаимодействует с обогащенной отрицательно заряженными аминокислотными остатками областью 190–210 АК домена NCR. Во втором варианте высокое содержание солей приводит к экранированию заряженной области NCR и домен 4 становится свободен, что, предположительно, приводит к росту скорости полимеризации. Результаты моделирования не подтверждают существующие на данный момент гипотезы о расположении N- и С-концов белка при самоингибировании.

Ключевые слова:  $\sigma^{70}$ -субъединица РНК-полимеразы, молекулярная динамика.

$\sigma^{70}$ -Субъединица РНК-полимеразы *E. coli* – небольшой белок, который входит в состав холофермента РНК-полимеразы и отвечает за специфичность инициации транскрипции генов [1]. Трехмерная структура полноразмерной  $\sigma^{70}$ -субъединицы РНК-полимеразы в свободном состоянии в настоящее время остается неизвестной. В то же время исследователям удалось расшифровать структуру отдельных фрагментов  $\sigma^{70}$ -субъединицы РНК-полимеразы *E. coli* [2]. Было обнаружено, что структура индивидуальных доменов белка в свободном состоянии очень сходна с их структурой в составе холофермента РНК-полимеразы [3]. Известно, что  $\sigma^{70}$ -субъединица в свободном виде не взаимодействует с промоторами. Было предложено несколько механизмов, объясняющих ингибирование связывания ДНК свободной  $\sigma^{70}$ -субъединицей [4–7], однако до настоящего времени не получено данных, которые бы однозначно свидетельствовали в пользу какого-либо из них.

В литературе имеются данные о способности  $\sigma^{70}$ -субъединицы формировать линейные агрегаты в условиях, близких к физиологическим [8–10]. Исследование морфологии и физико-хи-

мических свойств белка может пролить свет на механизм фибриллообразования.

### МАТЕРИАЛЫ И МЕТОДЫ

Белок  $\sigma^{70}$ -субъединица РНК-полимеразы *E. coli* представлен в банке структурных данных RCSB PDB 23 вариантами. Из них 20 содержат практически всю структуру за исключением небольших пропусков. Во всех вариантах  $\sigma^{70}$ -субъединица находилась в составе холофермента РНК-полимеразы. Это примечательный факт, поскольку в таком состоянии белок демонстрирует конформацию, наиболее удобную для выполнения своей функции, а не кристаллическую форму, порой весьма далекую от нативной.

Названия и некоторые особенности структур РНК-полимеразы *E. coli* перечислены в табл. 1.

Для того чтобы оценить, насколько разные получаются структуры  $\sigma^{70}$ -субъединицы в кристалле и как они зависят от лигандов, было проведено попарное выравнивание с подсчетом среднеквадратичного отклонения (*RMSD*) в

**Таблица 1.** Структуры РНК-полимеразы *E. coli* из банка данных RCSB PDB

Название структуры в PDB банке	Лиганд	Автор	Наличие контакта между $\sigma^{70}$ -субъединицей и лигандом	Разрешение, Å
4YLO	RNA DNA	Zuo	+	6,0
4YLN	RNA DNA	Zuo	+	5,50
4YLP	RNA DNA	Zuo	+	5,50
4YFK	Squaramide 8	Molodtsov	–	3,57
4LJZ	–	Bae	–	3,59
4YFN	Squaramide 14	Molodtsov	–	3,82
4LK0	T7 Gp2	Bae	–	3,91
4MEX	Salinamide A	Feng	–	3,90
4MEY	–	Feng	–	3,95
4LK1	–	Bae	–	3,84
4LLG	Gp2	Bae	+	3,79
4YG2	–	Murakami	–	3,70
4JKR	ppGpp	Zuo	–	4,20
1SIG	–	Malhotra	–	2,60
4YFX	Мухорыронин B	Molodtsov	–	3,84
4KMU	Rifampin	Murakami	–	3,85
4KN4	Benzoxazinofiramycin-2b	Murakami	–	3,96
4KN7	Benzoxazinofiramycin-2c	Murakami	–	3,69
4JK2	pppGpp	Murakami	–	4,20
4JK1	ppGpp	Murakami	–	3,90

**Таблица 2.** Содержащиеся в структурах  $\sigma^{70}$ -субъединицы аминокислотные остатки

Название структуры	Номера содержащихся в структуре аминокислот
YLO, YLP, YLN	79–171, 210–613
4YG2, 4YFK, 4LJZ, 4LK0	94–167, 213–236, 243–613
4YFX, 4KMU, 4KN4, 4KN7, 4JK1, 4JK2	95–154, 212–609
4MEX, 4MEY	95–107, 114–165, 210–273, 242–612
4LK1, 4LLG	7–56, 94–167, 213–236, 243–613
4JKR	95–171, 210–613
1SIG	113–167, 173–191

программе MaxCluster [11] по следующей формуле:

$$RMSD = \sqrt{\sum_i^N \frac{d_i^2}{N}},$$

где  $d_i$  – расстояние между выбранной парой атомов двух сравниваемых белков,  $N$  – число пар для сравнения. Среднеквадратичное отклонение подсчитывали с использованием суперпозиции белков таким образом, чтобы достичь минимального значения. Выравнивание белков

проводили с использованием матрицы вращения Кабсча [12,13].

Для моделирования целого белка предлагалось использовать программу предсказания структуры его недостающих частей [14,15].

Для того чтобы выбрать наиболее полную структуру для достраивания, была составлена табл. 2, в которой структуры разбиты на группы по количеству отсутствующих в них аминокислот.

Из табл. 2 видно, что лучше всего для достраивания подходят структуры 4LK1 и 4LLG, содержащие пропуски размером в 6, 38,

46 и 7 аминокислотных остатков. Самый большой «пустой» участок в 46 аминокислот между остатками 167 и 213 частично перекрывается аминокислотными остатками из структуры 1SIG (рис. 1). Среднеквадратичное отклонение между структурами 4LK1 и 1SIG составляло 1,24 Å, тогда как между 4LLG и 1SIG оно равнялось 1,6 Å. Поэтому в качестве структуры для достраивания была выбрана структура 4LK1. Было проведено выравнивание структур 4LK1 и 1SIG, после чего недостающий участок из 1SIG был добавлен в 4LK1, а затем проведено достраивание недостающих участков белка с помощью программы FALC-Loop [16].

Далее полученная полная структура  $\sigma^{70}$ -субъединицы без пропусков подвергалась моделированию в течение 150 нс в растворах солей с концентрацией солей 0 мМ; 20 мМ NaCl и 5 мМ MgCl<sub>2</sub>; 40 мМ NaCl и 5 мМ MgCl<sub>2</sub>; 40 мМ NaCl. Моделирование проводили в среде Gromacs 5.0.4 [17] и силовом поле AMBER99sb-ILDN [18].

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

По результатам моделирования мономеров  $\sigma^{70}$ -субъединицы с полной структурой было обнаружено, что наибольшей подвижностью обладают области N- и C-концов белка (рис. 2). И если подвижность N-конца подтверждается наличием пропусков в структуре, полученной методом рентгеноструктурного анализа, то подвижность C-конца может объясняться функциональной чувствительностью к ионной силе раствора, в котором содержится белок. В таком случае при отсутствии ионов в кристалле будет наблюдаться одна и та же структура C-конца белка, тогда как в растворе могут наблюдаться различные варианты конформаций.

В построенной модели наиболее структурно-стабильным участком оказалась область, соответствующая домену NCR (non conservative region, 127–370 АК). Структурное выравнивание, в целом по белку дающее различие в *RMSD* от 16 до 20 Å, в случае этого домена давало 1,83 Å для структур на рис. 2г и 2д, 1,80 Å для структур на рис. 2г и 2в, 1,87 Å для структур на рис. 2г и 2б, а также 1,97 Å для структур на рис. 2г и 2а. Длина стабильной области составляла 250–300 АК.

В работе [19] было показано, что C-конец белка, а именно домен 4.2 (566–613 АК), отвечает за распознавание промотера ДНК, а N-концевой домен 1.1 (1–100 АК) ингибирует этот участок в случае, когда  $\sigma^{70}$ -субъединица находится не в составе РНК-полимеразы. Это оз-

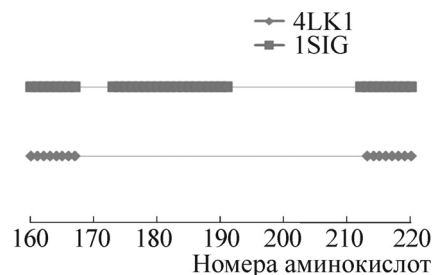


Рис. 1. Аминокислотные остатки, присутствующие в структурах 4LK1 и 1SIG.

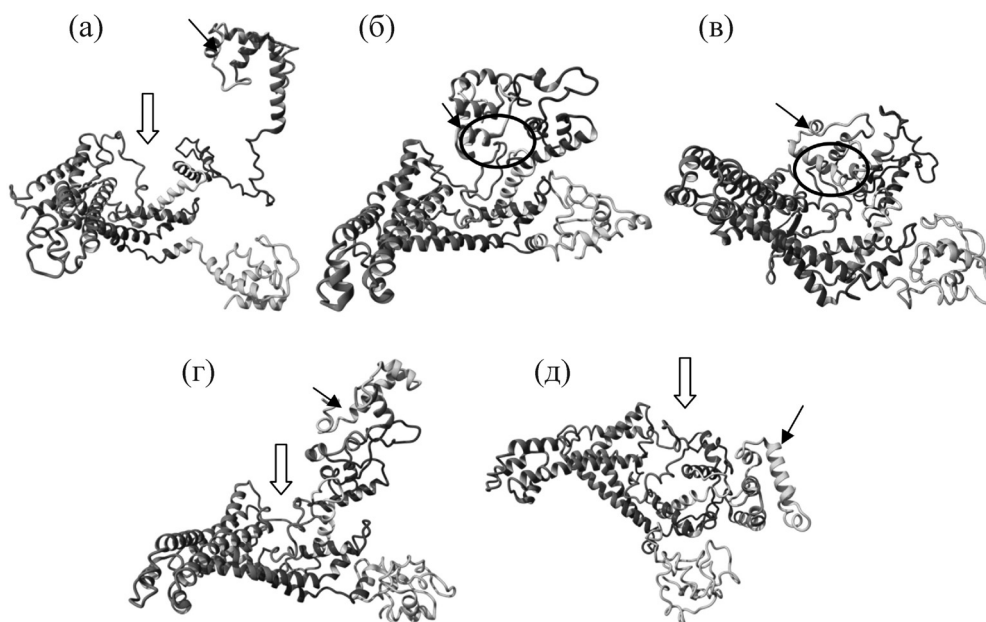
начает, что в свободной  $\sigma^{70}$ -субъединице N- и C-концы белка должны быть расположены в непосредственной близости друг от друга. Однако в модели не было выявлено контакта между доменами 1.1. и 4.2.

Домен 1.1, по данным статей [20,21] предположительно гидрофобно взаимодействующий с доменами 2.3 и 2.4, во всех случаях был расположен на удалении от них.

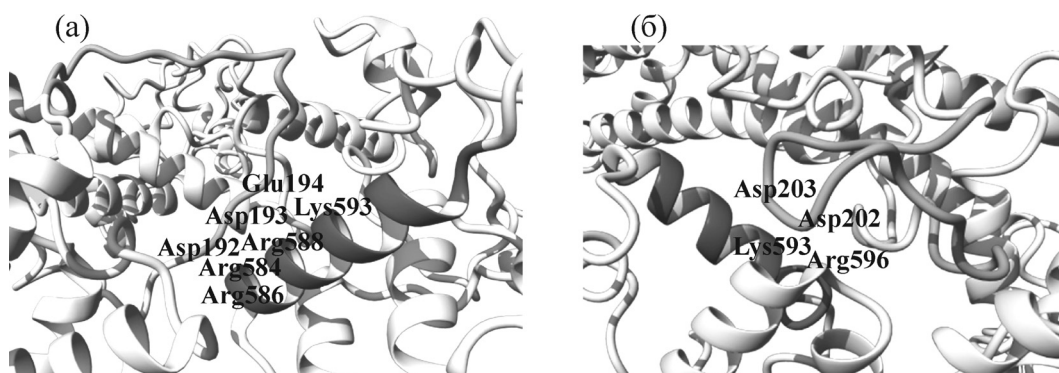
Рис. 2, тем не менее, демонстрирует некоторые общие тенденции, характерные для поведения мономера  $\sigma^{70}$ -субъединицы, находящегося не в составе РНК-полимеразы. Все структуры на рис. 2б–д, по сравнению с начальной структурой на рис. 2а из кристалла, показывают более компактное расположение домена 4, близкое к остальным частям молекулы. В случае низких концентраций солей наблюдалось взаимодействие домена 4.2 с областью 190–210 АК из домена NCR (рис. 2б,в), тогда как при высоких концентрациях солей эти домены не взаимодействовали.

На итоговых структурах был проведен анализ расположения заряженных областей. В целом, хотя белок имеет высокий отрицательный заряд ( $-44 e^-$ ), заряды распределены по белку довольно равномерно. Имеется только одна крупная область, обогащенная отрицательно заряженными остатками: 191–209 АК,  $-16 e^-$ .

При низких концентрациях соли наблюдалось электростатическое взаимодействие с этим районом небольшого участка домена 4.2, соответствующего 584–599 АК и имеющего суммарный положительный заряд  $+5 e^-$ . А именно, аминокислоты Glu198 и Glu199 взаимодействовали с Arg596; Asp202 и Asp204 – с Glu585 и Lys593 при концентрации солей 20 мМ NaCl и 5 мМ MgCl<sub>2</sub>, Arg584 – с Asp191–Asp193, Glu194 и Asp195 при концентрации солей 0 мМ (рис. 3).



**Рис. 2.** Результаты моделирования полной  $\sigma^{70}$ -субъединицы РНК-полимеразы *E. coli* при разных концентрациях солей: (а) – начальная структура после добавления недостающих участков пептидной цепи, (б) –  $\sigma^{70}$ -субъединица после 150 нс моделирования с нулевой концентрацией ионов, (в) – белок после 150 нс моделирования в 20 мМ NaCl и 5 мМ MgCl<sub>2</sub>, (г) – белок после 150 нс моделирования в 40 мМ NaCl, (д) – белок после 150 нс моделирования в 40 мМ NaCl и 5 мМ MgCl<sub>2</sub>. Тонкой стрелкой выделен домен 4.2 (566–613 АК), овалом выделена область контакта между доменом 4.2 и областью отрицательного заряда в домене NCR (190–210 АК), широкой стрелкой выделена область отрицательного заряда в домене NCR.



**Рис. 3.** Взаимодействие домена 4.2 с областью отрицательного заряда домена NCR при концентрациях соли 0 мМ (а), 20 мМ NaCl и 5 мМ MgCl<sub>2</sub> (б).

Ранее мы проводили исследование полимеризации  $\sigma^{70}$ -субъединицы РНК-полимеразы *E. coli* при различных концентрациях солей [22,23]. Известно, что  $\sigma^{70}$ -субъединица формирует амилоидные фибриллы в условиях, близких к физиологическим, однако механизм этого процесса не выяснен. Методом атомно-силовой микроскопии было обнаружено, что самая высокая скорость фибриллообразования возникала при высоких концентрациях солей, а именно при 40 мМ NaCl. В экспериментах по модели-

рованию полимеризации  $\sigma^{70}$ -субъединицы было обнаружено, что одна из областей контакта между белками в фибрилле – это домен 4.2, взаимодействующий с доменом 4.2 из соседней молекулы (данные не опубликованы). Возможно, что рост скорости фибриллообразования как раз и объясняется тем, что домен 4.2 при высокой ионной силе перестает взаимодействовать с областью отрицательного заряда в домене NCR и становится доступен для контакта с соседней молекулой.

## ВЫВОДЫ

В результате моделирования полной структуры  $\sigma^{70}$ -субъединицы РНК-полимеразы *E. coli* было обнаружено два варианта расположения С-концевого домена 4 белка. В одном из них при низкой концентрации солей в растворе домен 4 взаимодействует с сильно отрицательно заряженной областью 190–210 АК домена NCR. В другом варианте высокое содержание солей приводит к экранированию заряженной области NCR и домен 4 становится свободен, что, предположительно, приводит к росту скорости полимеризации.

Авторы выражают благодарность В.А. Твердислову и В.Л. Друзе за обсуждение результатов.

Работа выполнена с использованием ресурсов суперкомпьютерного комплекса МГУ им. М.В. Ломоносова при финансовой поддержке Российского фонда фундаментальных исследований (гранты №№ 15-32-20629 и 13-04-01504).

## СПИСОК ЛИТЕРАТУРЫ

1. T. M. Gruber and C. A. Gross, *Annu. Rev. Microbiol.* **57**, 441 (2003).
2. A. Malhotra, E. Severinova, and S. A. Darst, *Cell* **87**, 127 (1996).
3. E. A. Campbell, O. Muzzin, M. Chlenov, et al., *Mol. Cell* **9**, 527 (2002).
4. E. C. Schwartz, A. Shekhtman, K. Dutta, et al., *Chem. Biol.* **15**, 1091 (2008).
5. A. J. Dombroski, W. A. Walter, and C. A. Gross, *Genes Dev.* **7**, 2446 (1993).
6. S. Callaci, E. Heyduk, and T. Heyduk, *Mol. Cell* **3**, 229 (1999).
7. S. Callaci and T. Heyduk, *Biochemistry* **37**, 3312 (1998).
8. P. A. Lowe, U. Aebil, C. Gross, and R. R. Burgess, *J. Biol. Chem.* **256**, 2010 (1981).
9. A. L. Ferguson, A. D. Hughes, U. Tufail, et al., *FEBS Lett.* **481**, 281 (2000).
10. S. Callaci, E. Heyduk, and T. Heyduk, *J. Biol. Chem.* **273**, 32995 (1998).
11. <http://www.sbg.bio.ic.ac.uk/~maxcluster/index.html>.
12. W. Kabsch, *Acta Crystallographica Section A* **32**, 922 (1976).
13. W. Kabsch, *Acta Crystallographica Section A* **34**, 827 (1978).
14. W. Pirovano and J. Heringa, *Methods Mol. Biol.* **609**, 327 (2010).
15. P. Y. Chou and G. D. Fasman, *Biochemistry* **13** (2), 222 (1974).
16. J. Ko, D. Lee, H. Park, et al., *Nucl. Acids Res.* **39**, W210 (2011).
17. D. Van der Spoel, R. Van Drunen, and H. J. C. Berendsen, *GRONingen MACHine for Chemical Simulation* (BIOSON Research Institute, Nijenborgh 4NL-9717 AG, Groningen; 1994).
18. V. Hornak, R. Abel, A. Okur, et al., *Proteins* **65**, 712 (2006).
19. E. C. Schwartz, A. Shekhtman, K. Dutta, et al., *Chemistry & Biology* **15**, 1091 (2008).
20. A. J. Dombroski, W. A. Walter, and C. A. Gross, *Genes & Development* **7**, 2446 (1993).
21. J. A. Camarero, A. Shekhtman, E. A. Campbell, et al., *Proc. Natl. Acad. Sci. USA* **99** (13), 8536 (2002).
22. E. V. Dubrovin, O. N. Koroleva, Y. A. Khodak, et al., *Nanomedicine: Nanotechnology, Biology, and Medicine* **8**, 54 (2012).
23. O. N. Koroleva, E. V. Dubrovin, Y. A. Khodak, et al., *Cell Biochem. Biophys.* **66** (3), 623 (2013).

## Investigation of the Dependence of *Escherichia coli* RNA Polymerase $\sigma^{70}$ -Subunit Structure on Ionic Strength by Molecular Dynamics Simulation Method

A.P. Tolstova\*, E.V. Dubrovin\*, and O.N. Koroleva\*\*

\*Department of Physics, Lomonosov Moscow State University, Leninskiye Gory, Moscow, 119991 Russia

\*\*Department of Chemistry, Lomonosov Moscow State University, Leninskiye Gory 1/3, Moscow, 119991 Russia

The  $\sigma^{70}$ -subunit of *E. coli* RNA polymerase (a small protein, being a part of RNA holoenzyme, and responsible for initiation of transcription of constitutive genes) is modeled at different ionic strengths. Two variants of the location of C-end domain 4 are obtained. At low ionic strength domain 4 interacts with the region of high negative charge 190–210 AK within NCR domain. At high ionic strength this region was screened and domain 4 was free and set away from domain NCR. We suppose that this leads to the increase in polymerization rate. Simulation data do not confirm any hypothesis about a self-inhibition mechanism.

*Key words:*  $\sigma^{70}$ -subunit of RNA polymerase, molecular dynamics simulation