

Ю.А. Мазур^{1,2}, Н.Ю. Опарина²

¹ Московский физико-технический институт (государственный университет)

² Институт молекулярной биологии им. В.А. Энгельгардта РАН

Молекулярное моделирование распознавания стоп-кодона фактором терминации трансляции 1-го типа eRF1

Для протекания терминации трансляции у эукариот необходимо распознавание трёх стоп-кодонов: UAA, UAG и UGA, которое осуществляется фактором eRF1. Известно, что N-концевой домен этого фактора отвечает за декодирование, однако вопрос о непосредственном сайте связывания и механизме распознавания оставался нерешённым. В нашей работе рассматривается взаимодействие 2-го и 3-го пуриновых нуклеотидов стоп-кодонов с N-концевым доменом белка eRF1. Применены методы молекулярного моделирования и получены модели комплексов N-концевого домена eRF1 с динуклеотидами AA, AG и GA, отвечающие требованиям специфического распознавания. Показано, что для «неспецифического» динуклеотида GG возможно связывание с N-концевым доменом eRF1 с энергиями, близкими к таковым для «специфических» динуклеотидов AA, AG и GA, однако при этом задействованы разные сайты связывания. Показано непосредственное контактирование с лигандом в полученных моделях аминокислотных остатков eRF1 Thr122, Ser123 и Tyr125, для которых ранее экспериментально показана их функциональная важность при декодировании стоп-кодонов. Виртуальный мутагенез Thr122, Ser123 → Gln122, Ph123 показал изменение энергий образования комплекса, согласующееся с известными экспериментальными данными.

Ключевые слова: молекулярное моделирование, декодирование стоп-кодонов, фактор eRF1, динуклеотиды, виртуальный мутагенез.

I. Введение

Терминация трансляции — важнейший завершающий этап синтеза белка, при котором обеспечивается высокая точность распознавания «бессмысленных» стоп-кодонов и новосинтезированные полипептиды диссоциируют от рибосом. В отличие от прокариот, у которых декодирование трёх стоп-кодонов UAA, UAG и UGA осуществляется с помощью двух паралогичных факторов терминации RF1 и RF2, отличающихся специфичностью по отношению к стоп-кодонам, у эукариот и архебактерий все три стоп-кодона с одинаковой эффективностью декодируются одним белковым фактором — eRF1/aRF1 [1–4]. Собственно распознавание стоп-кодонов происходит после связывания фактора eRF1/aRF1 с A-сайтом рибосомы, однако есть ряд свидетельств, что основную роль в специфическом декодировании играет не рибосома, а eRF1/aRF1, точнее его N-концевой домен [5–10]. Применение методов случайного [11] и направленного мутагенеза [8, 9] позволили

выявить в N-концевом домене eRF1 ряд аминокислотных остатков, замены в которых меняют специфичность декодирования стоп-кодонов. Однако несмотря на расположение в одном домене и сближенность некоторых из выявленных участков, в целом они расположены в различных областях декодирующего домена eRF1 и не образуют четких пространственных кластеров. Одним из многообещающих подходов для выявления механизма распознавания стоп-кодона служит исследование филогенетических групп, в которых наблюдаются отклонения от стандартного генетического кода. В частности, для инфузорий *Stylonychia*, *Tetrahymena* и *Paramecium* стоп-кодом служит только UGA, а триплеты UAA и UAG кодируют глутамин, а для *Euplotes* показано использование в качестве стоп-кодонов только UAA и UAG, но не UGA, который кодирует цистеин [7, 12–15]. В перечисленных работах были предприняты попытки выявления области eRF1, отвечающей за специфическое декодирование стоп-кодонов, с

помощью сравнительного анализа выравненных аминокислотных последовательностей eRF1, различающихся набором распознаваемых стоп-кодонов. Однако и в этом случае не выявили четкого кластера аминокислотных замен, который можно было бы считать потенциальным сайтом декодирования. На основании сравнения eRF1 из групп с одинаковым изменением специфичности распознавания стоп-кодонов экспериментально показано, что, по-видимому, один и тот же тип изменения специфичности связывания N-концевого домена eRF1 со стоп-кодонами можно создать рядом независимых способов с использованием пространственно разнесенных аминокислотных замен [10, 14].

Следует отметить, что во всех перечисленных случаях имеющиеся экспериментальные данные позволяют сопоставлять способность eRF1 распознавать только 2-й и 3-й пуриновые нуклеотиды стоп-кодона (AA, AG или GA) и отличать их от 2-го и 3-го гуаниновых нуклеотидов триптофанового кодона UGG, сведения же о распознавании первого инвариантного урацила стоп-кодонов незначительны.

В нашей работе проведено молекулярное моделирование образования комплексов N-концевого домена eRF1 человека (PDB 1dt9) и «специфических» рибонуклеотидов AA, AG и GA в сравнении с «неспецифическим» рибонуклеотидом GG.

II. Материалы и методы

II.1. Исходные данные

Для анализа взаимодействия N-концевого домена белка eRF1 человека с пуриновыми динуклеотидами стоп-кодонов и родственным «неспецифическим» динуклеотидом GG использована структура 1DT9 из банка данных RCSB (www.rcsb.org). Из структуры полноразмерного белка eRF1 была вырезана с использованием пакета программ Sybyl (www.tripos.com) структура N-концевого домена, состоящая из аминокислотных остатков 5-140. С помощью программы Biopolymer из пакета программ Sybyl были получены модели рибонуклеотидов PO4-5'-AA-3'-ОН, PO4-5'-AG-3'-ОН, PO4-5'-GA-3'-ОН и PO4-5'-GG-3'-ОН в виде одноцепочечных фрагментов A-формы

РНК. Необходимые точечные мутации в комплексы N-домен+dNTP вносились с помощью программ SwissPDBViever и пакета программ Sybyl.

II.2. Молекулярный докинг

Молекулярный докинг пуриновых рибонуклеотидов на N-концевой домен eRF1 человека выполнен с помощью программы Autodock 3.0 в отсутствие начальных ограничений. Для оптимизации скоринг-функции применяли параметры для докинга РНК (Moitessier et. al., 2006) [16]. Использован генетический алгоритм для автоматического докинга. Для каждого рибонуклеотида проводили 100 независимых раундов молекулярного докинга.

II.3. Анализ результатов молекулярного докинга и кластеризация потенциальных моделей комплексов N-концевого домена eRF1 с рибонуклеотидами

Среди вариантов расположения лигандов на первом этапе были выбраны только те, для которых энергия взаимодействия согласно скоринг-функции программы Autodock была минимальна.

Далее с помощью in-house программы фильтрации результатов молекулярного докинга были рассчитаны межмолекулярные взаимодействия для РНК-белковых комплексов (включая водородные связи и ван-дер-ваальсовы взаимодействия) и применены следующие фильтры: отсутствие плотного (ближе 4Å) контакта с белком для 5'/3' концов рибонуклеотида (данный параметр определяется тем, что изучаемые рибонуклеотиды на деле представляют собой фрагменты мРНК); основная часть (более 80%) межмолекулярных взаимодействий для рибонуклеотида должна приходиться на азотистые основания, а не на сахарофосфатный остов; наличие среди контактов не менее 3 связей с длиной до 3,5Å; соотношение контактов между первым и вторым пуринами не должно превышать 3:1 (этот параметр введен, чтобы избежать отбора стабильных комплексов, для которых большая часть РНК-белковых взаимодействий приходится только на один из нуклеотидов, что не позволяет считать такие взаимодействия специфичными).

На втором этапе все модели комплексов, прошедшие фильтр, были классифицированы согласно их пространственному расположению на N-концевом домене eRF1 (согласно совпадению аминокислотных остатков, взаимодействующих с 1-м и 2-м пуринами рибонуклеотида соответственно). По результатам работы программы были построены кластеры возможных специфических РНК-белковых взаимодействий с использованием пакета программ CLUTO (glaros.dtc.umn.edu/gkhome/views/cluto) и in-house разработанного программного продукта. Для каждого из кластеров были посчитаны средние энергии, и если эти средние энергии оказывались отрицательными, то из таких кластеров отбиралась лучшая модель и далее использовалась для молекулярной динамики.

II.4. Молекулярная динамика

Молекулярная динамика отобранных в докинге моделей была выполнена с помощью пакета программ Amber. Были использованы условия непрямого указания растворителя. Последовательность действий при молекулярной динамике была следующей: этап минимизации структуры, медленный нагрев от 0 до 300 К и проведения динамики при 300 К до достижения равновесного состояния. Поскольку параметры диэлектрической проницаемости рибосомы в области A-сайта неизвестны, была выбрана условная единица 20, основанная на большей гидрофильности РНК, нежели внутренняя часть белковой глобулы (где диэлектрическая проницаемость равна 4). Аналогичная процедура проводилась для свободного N-домена eRF1 и для свободных динуклеотидов AA, AG, GA и GG. Это позволило определить потенциальную энергию в равновесном состоянии и разницу этой энергии для комплекса и для пары «свободный белок + свободный лиганд».

Кроме того, для комплексов AG/eRF1 и GA/eRF1 получены «мутантные» структуры, в которых искусственно аденин был заменен на гуанин (программа SwissPDBViewer), при этом расположение нуклеотидов не меняли. Молекулярная динамика этих комплексов была проведена при тех же условиях.

III. Результаты

При проведении «слепого» молекулярного докинга подавляющее большинство моделей соответствовало расположению лиганда в области, близкой к функционально важному Y-C-F району [1, 8–9] для каждого из динуклеотидов AA, AG, GA и GG. Однако отобранные модели не позволяли с помощью фильтрации по энергии взаимодействия, определённой с помощью скоринг-функции Autodock 3.0, выявить более локальный участок взаимодействия. Поэтому нами была применена система фильтрации моделей комплексов, необходимая для отбора потенциально специфических комплексов динуклеотид — eRF1 (см. материалы и методы). При этом были отобраны кластеры моделей расположения пуринового динуклеотида в комплексе с eRF1 (рис. 1). Однако на основании одного только сравнения энергий взаимодействия также не удаётся определить сайт декодирования стоп-кодона в eRF1. Нами было замечено, что среди выявленных ориентаций динуклеотидов AA, AG и GA присутствует чрезвычайно близкое направление (отмеченное черной стрелкой), в то же время комплексов с аналогичным расположением динуклеотидного лиганда для моделей связывания GG и eRF1 не обнаружено (существует лишь близкое направление). Это позволило нам предположить, что в отличие от РНК-белковых комплексов, образующихся в растворе, специфический комплекс eRF1–стоп-кодон не способен образовываться в отсутствие рибосомы, которая обеспечивает правильное взаимное расположение домена 1 eRF1 и стоп-кодона в мРНК. Таким образом, модели комплексов динуклеотидов AA, AG и GA, соответствующие общему кластеру расположения лиганда, отсутствующему среди комплексов eRF1 с GG, были отобраны для дальнейшего исследования с помощью молекулярной динамики. Как указывалось выше (см. материалы и методы), для дальнейшего анализа также были использованы представители других кластеров.

Для более точного анализа взаимодействий в комплексе РНК–белок, а также для оценки его стабильности была проведена молекулярная динамика отобранных моделей. На основании анализа энер-

гий взаимодействия домена 1 eRF1 со специфическими динуклеотидами AA, AG и GA и с неспецифическим GG были отобраны три модели (модели AA_82, AG_66, GA_96), по 1 модели для каждого специфического динуклеотида, отвечающие следующим требованиям: энергии взаимодействия eRF1 со специфическими лигандами должны быть близки, но в то же время быть значительно ниже, чем энергии связывания eRF1 с неспецифическим лигандом GG (табл. 1).

Проведённый анализ контактных аминокислотных остатков показал, что аминокислоты T122, S123, L124 и Y125 близко контактируют со всеми специфическими динуклеотидами (табл. 2). Значительная часть контактных аминокислотных остатков консервативна или инвариантна для eRF1 с неизменной специфичностью распознавания стоп-кодона.

Мы провели анализ расположения вероятных водородных связей и ван-дер-ваальсовых взаимодействий для отобранных моделей комплексов. Как видно из таблицы 3, часть определённых контактных остатков участвует во взаимодействии

одновременно и с первым, и со вторым нуклеотидами лиганда (соответствуют второму и третьему нуклеотидам стоп-кодона). Однако в целом можно говорить о предпочтительном участии аминокислотных остатков 122, 123 в распознавании первого, а 125 — второго нуклеотидов лиганда.

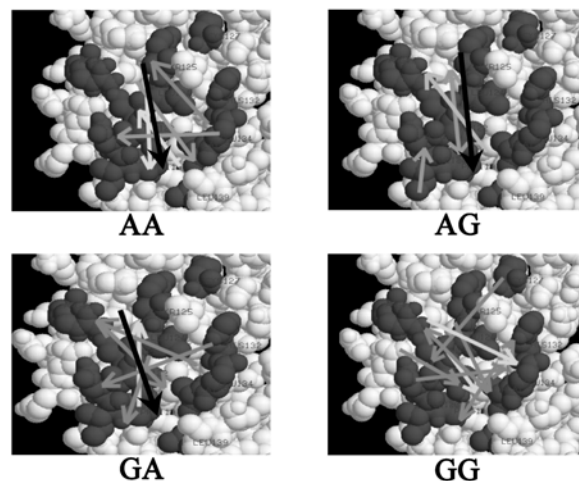


Рис. 1. Расположение кластеров динуклеотидных лигандов в комплексе с доменом 1 eRF1. Стрелками схематически показаны направления, характерные для кластеров. Направление стрелок указывает направление от 5' конца мРНК к 3' концу

Т а б л и ц а 1

Энергии взаимодействия специфических и неспецифического динуклеотидов с доменом 1 eRF1

	Потенциальная энергия	Разница в энергии комплекса и пары (свободный белок+свободный лиганд)	
AA			
AA_26	-3048,1199	-13,397	
AA_82	-3066,478	-31,7551	
AA_69	-3028,9956	5,7273	
AG			Сравнение с «мутантом» AG-GG
AG_41	-3085,8551	-16,5292	14,7129
AG_66	-3094,7455	-25,4196	-2,1839
GA			Сравнение с «мутантом» GA-GG
GA_29	-3049,7892	21,4498	-26,8293
GA_34	-3099,3908	-28,1518	-28,8103
GA_96	-3103,9689	-32,7299	-6,3395

Т а б л и ц а 2

**Контактирующие с динуклеотидным лигандом
аминокислотные остатки в домене 1 eRF1**

Intersections			Variation among eRF1s
AA	AG	GA	
I38			Variable
	P40	P40	Conserved
	R47	R47	Conserved
		V48	Variable
		M51	Variable
D43			Conserved
P119			Invariant
	N121	N121	Invariant
T122	T122	T122	Conserved
S123	S123	S123	Conserved
L124	L124	L124	Conserved
Y125	Y125	Y125	Invariant
L126			Conserved
	H132		Conserved
	E134		Conserved
A135	A135		Conserved
L136			Invariant
A138			Variable
L139			Invariant

Показаны близкие ($< 3\text{Å}$) контакты для моделей комплексов с AA, AG и GA. Аминокислоты и их нумерация соответствуют eRF1 человека. Отмечена консервативность соответствующих остатков для eRF1, распознающих все три стоп-кодона. Выделены аминокислоты, контактирующие со всеми вариантами динуклеотидных лигандов.

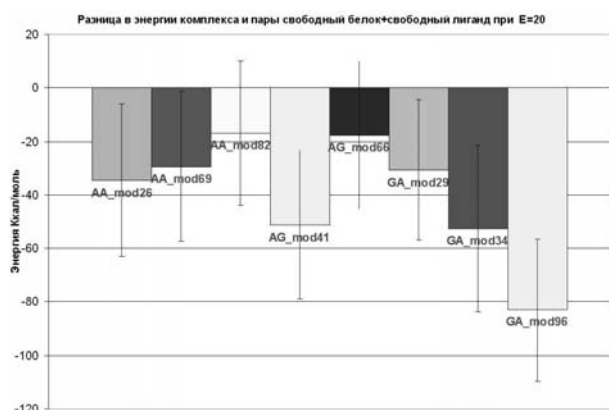


Рис. 2. Данные приведены для комплекса с мутантным белком (Thr122-Ser123 → Gln122-Phe123)

В конце мы провели виртуальный мутагенез, при котором заменили две аминокислоты Thr122-Ser123 на Gln122-Phe123. После проведённой 15 пс молекулярной динамики были определены энергии взаимодействия (рис. 2). Как легко заметить,

наиболее стабильным среди проанализированных вариантов оказался комплекс с GA, соответствующий стоп-кодону UGA, но не комплексы с AA и AG, представляющими фрагменты стоп-кодонов UAA и UAG. Таким образом, полученные данные полностью соответствуют известным результатам эксперимента, что подтверждает высокую специфичность взаимодействия полученных нами молекулярных моделей [10]. Необходимо отметить, что наилучшее соответствие с экспериментом показала тройка моделей AA_mod82, AG_mod66, GA_mod96, которые соответствуют немутированным моделям AA_82, AG_66, GA_96. Поскольку отбор этих моделей был независимым в двух виртуальных экспериментах, можно говорить о данных моделях как о потенциальных кандидатах в качестве специфических комплек-

сов посадки мРНК на N-концевой домен eRF1 человека.

Т а б л и ц а 3

Аминокислотные остатки домена 1 eRF1, образующие вероятные водородные связи и/или ван-дер-ваальсовы взаимодействия с первым и вторым нуклеотидами специфического динуклеотидного лиганда

Amino acid	AA	AG	GA
ILE38	A2		
ARG47		A1	G1
VAL48			G1
MET51			G1
ASN121		A1	A2
THR122	A1, A2	A1	G1
SER123	A1	A1	G1
LEU124	A2	G2	A2
TYR125	A1, A2	G2	A2, G1
HIS132		G2	
GLU134		G2	
ALA135	A2		

IV. Выводы

Впервые применены методы молекулярного моделирования для определения декодирующего участка на N-концевом домене eRF1 человека. Показано, что как для специфических пуринов стоп-кодонов (AA, AG и GA), так и для неспецифического динуклеотида GG удаётся выявить сайты связывания, близкие по пространственному расположению и энергии. Тем не менее, благодаря использованным в настоящей работе параметрам фильтрации моделей РНК-белковых комплексов удаётся выявить такое расположение пуриновых динуклеотидов на N-концевом домене eRF1, которое отвечает требованиям высокоспецифического связывания. Дополнительно проведённый виртуальный мутагенез части контактных аминокислот показал совпадение результатов молекулярного моделирования с известными ранее экспериментальными данными.

СПИСОК ЛИТЕРАТУРЫ

1. Kisselev L., Ehrenberg M., Frolova L. Termination of translation: interplay of mRNA, rRNAs and release factors? //

EMBO J. — 2003. — V. 22, N. 2. — P. 175–182.

2. Nakamura Y., Ito K. Making sense of mimic in translation termination // Trends Biochem. Sci. — 2003. — V. 28, N. 2. — P. 99–105.

3. Inge-Vechtomov S., Zhouravleva G., Philippe M. Eukaryotic release factors (eRFs) history // Biol. Cell. — 2003. — V. 95, N. 3–4. — P. 195–209.

4. Poole E.S., Askarian-Amiri M.E., Major L.L., McCaughan K.K., Scarlett D.J., Wilson D.N., Tate W.P. Molecular mimicry in the decoding of translational stop signals // Prog. Nucleic Acid Res. Mol. Biol. — 2003. — V. 74. — P. 83–121.

5. Bertram G., Bell H.A., Ritchie D.W., Fullerton G., Stansfield I. Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition // RNA. — 2000. — V. 6, N. 9. — P. 1236–1247.

6. Frolova L., Seit-Nebi A., Kisselev L. Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1 // RNA. — 2002. — V. 8, N. 2. — P. 129–136.

7. Kervestin S., Frolova L., Kisselev L., Jean-Jean O. Stop codon recognition in ciliates: Euplotes release factor does not respond to reassigned UGA codon // EMBO Rep. — 2001. — V. 2, N. 8. — P. 680–684.

8. Seit-Nebi A., Frolova L., Kisselev L. Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1 // EMBO Rep. — 2002. — V. 3, N. 9. — P. 881–886.

9. Kolosov P., Frolova L., Seit-Nebi A., Dubovaya V., Kononenko A., Oparina N., Justesen J., Efimov A., Kisselev L. Invariant amino acids essential for decoding function of polypeptide release factor eRF1 // Nucleic Acids Res. — 2005. — V. 33, N. 19. — P. 6418–6425.

10. Lekomtsev S., Kolosov P., Bidou L., Frolova L., Rousset JP., Kisselev L. Different modes of stop codon restriction by the Stylonychia and Paramecium eRF1 translation termination factors // Proc. Natl. Acad. Sci. USA. — 2007. — V. 104, N. 26. — P. 10824–10829.

11. Bertram G., Bell H.A., Ritchie D.W., Fullerton G., Stansfield I. Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon

recognition // RNA. — 2000. — V. 6, N. 9. — P. 1236–1247.

12. *Inagaki Y., Doolittle W.F.* Class I release factors in ciliates with variant genetic codes // Nucleic Acids Res. — 2001. — V. 29, N. 4. — P. 921–927.

13. *Kim O.T., Yura K., Go N., Harumoto T.* Newly sequenced eRF1s from ciliates: the diversity of stop codon usage and the molecular surfaces that are important for stop codon interactions // Gene. — 2005. — V. 346. — P. 277–286.

14. *Salas-Marco J., Fan-Minogue H., Kallmeyer A.K., Klobutcher L.A., Farabaugh P.J., Bedwell D.M.* Distinct

paths to stop codon reassignment by the variant-code organisms *Tetrahymena* and *Euplotes* // Mol. Cell. Biol. — 2006. — V. 26, N. 2. — P. 438–447.

15. *Liang H., Landweber L.F., Fresco JR.* Are stop codons recognized by base triplets in the large ribosomal RNA subunit? // RNA. — 2005. — V. 11, N. 10. — P. 1478–1484.

16. *Moitessier N., Westhof E., Hanessian S.* Docking of aminoglycosides to hydrated and flexible RNA // J. Med. Chem. — 2006. — V. 49, N. 3. — P. 1023–1033.

Поступила в редакцию 26.01.2008.