

УДК 577.112.4

## БЕЛКОВЫЙ СПЛАЙСИНГ

© 2007 г. П. Л. Старокадомский\*

Институт молекулярной биологии и генетики Национальной академии наук Украины, Киев, 03143, Украина

Поступила в редакцию 01.06.2006 г.

Принята к печати 07.07.2006 г.

Белковым сплайсингом называется процесс, в результате которого внутренняя часть белка (интеин) посттрансляционно вырезается, а фланкирующие ее последовательности (экстеины) лигируются. Для протекания белкового сплайсинга не требуется присутствие каких-либо кофакторов или ферментов, что отличает этот процесс от других вариантов процессинга белков. Белковый сплайсинг катализируется внутренним доменом самого интеина – так называемым Hint-доменом. В обзоре описаны основные закономерности и молекулярные механизмы этого процесса. Рассмотрены также функции Hint-доменов других семейств белков (Hh-белки, VIL-домены бактерий и т.п.). Исследование белкового сплайсинга имеет большое прикладное и научное значение. Так, анализ работ по сравнению интеинов, обнаруженных в различных ветвях живого, иллюстрирует роль горизонтального переноса в процессе распространения интеинов. В результате обобщения последних данных автором показано возможное участие белкового сплайсинга в регуляции ряда клеточных процессов.

*Ключевые слова:* интеин, белковый сплайсинг, Hint-домен, экстеин.

PROTEIN SPLICING, by P. L. Starokadomskii\* (Institute of Molecular Biology and Genetics, National Academy of Sciences, Kiev, 03143 Ukraine; \*e-mail: p.l.starokadomski@imbg.org.ua). Protein splicing is a post-translational autocatalytic excision of internal protein sequence (intein) with the subsequent ligation of the flanking polypeptides (exteins). This process doesn't require any cofactors or enzymes, which distinguish it from other variants of protein processing. Protein splicing is catalyzed by Hint-domain – internal intein's domain. In review the main molecular mechanisms of this process are described. Function of analogous Hint-domains of other proteins families (Hh-proteins, VIL-domains etc) are considered. The analysis of inteins characteristic to different branches of life illustrates the role of horizontal transfer in intein's distribution and evolution. A possible role of inteins in regulation of different cell processes is discussed.

*Key words:* intein, protein splicing, Hint-domain, extein.

Сплайсинг (от англ. to splice – сшивать, соединять) – это форма процессинга, в результате которого из молекулы-предшественника удаляется внутренняя последовательность с последующим лигированием (сплайсингом) фланкирующих последовательностей. Сплайсинг является одним из мощнейших способов быстрой и случайной перетасовки наследственной информации, что в некоторых случаях помогает живым системам быстро приспособиться к изменившимся условиям. Классическим примером служат клетки иммунной системы, универсальность которых обеспечивается альтернативным сплайсингом на уровне ДНК и РНК. До недавнего времени сплайсинг приписывали только нуклеиновым кислотам. Однако в 1990 г. группой Т. Стивенса было установлено, что ряд белков также могут приобретать способность к посттрансляционной модификации путем специфического процесса, аналогичного сплай-

сингу ДНК и РНК [1, 2]. Так в литературе появился термин “белковый сплайсинг”.

Белковый сплайсинг был открыт при исследовании дрожжевого гена *VMA1*, кодирующего субъединицу Vma1 вакуолярной АТФазы. Оказалось, что у других организмов высококонсервативный ген *VMA1* кодирует белок с молекулярной массой около 70 кДа, тогда как у дрожжей этот ген кодирует значительно больший белок. При этом для концевых последовательностей дрожжевого гена *VMA1* характерна высокая степень гомологии с последовательностями других микроорганизмов, тогда как в центральной части гомология нарушается. Далее было показано, что мРНК гена *VMA1* транслируется полностью, а центральная часть полипептида-предшественника удаляется уже посттрансляционно [1]. В результате созревания центральная часть белка удаляется, а фланкирующие ее последовательности лигируются. Эти и ряд других результатов позволили утверждать: после трансляции в неко-

\* Эл. почта: p.l.starokadomski@imbg.org.ua

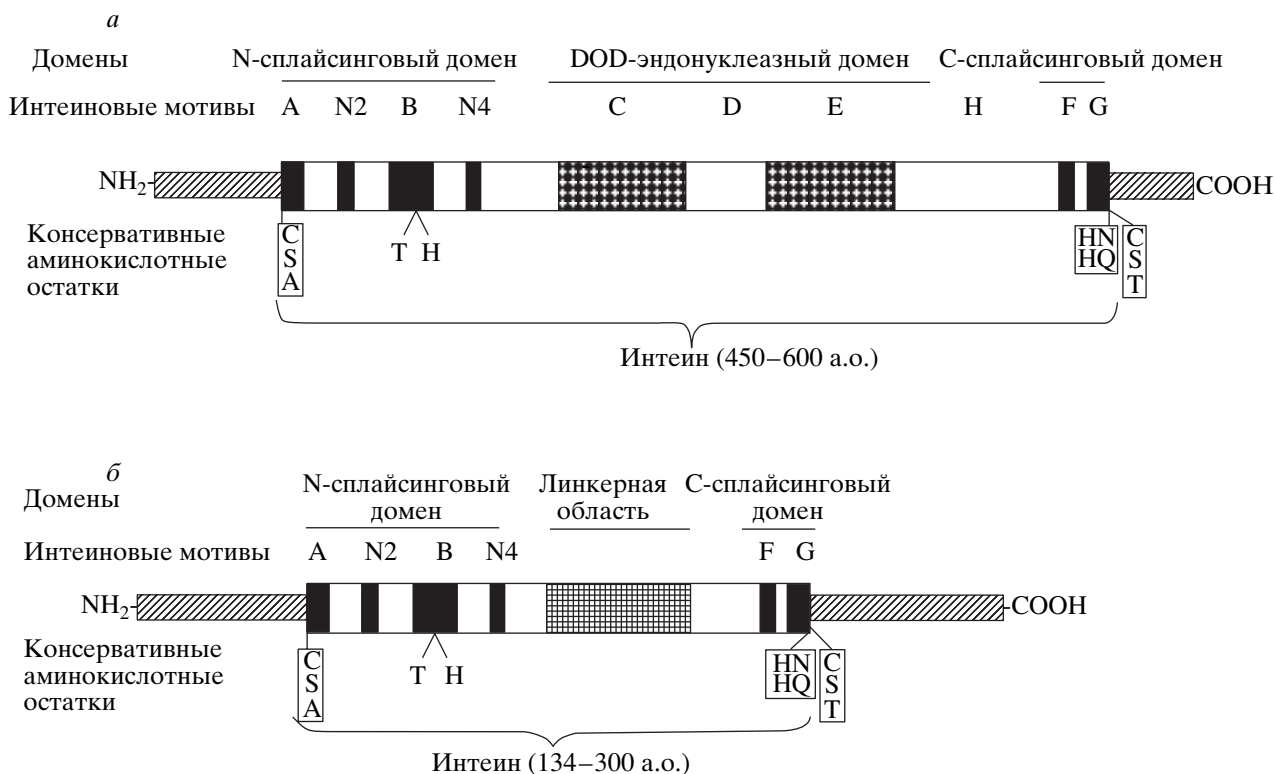


Рис. 1. Схематическое строение основных типов интееинов: а – классический DOD-содержащий интееин, б – мини-интееин.

торых случаях может происходить процесс, аналогичный сплайсингу пре-РНК. Центральную часть белка, которая удаляется, назвали интееином (от **internal protein**), фланкирующие последовательности – соответственно N- и C-экстеинами (**external protein**), а сам процесс – белковым сплайсингом [2, 3]. Характерно, что процесс является автотокаталитическим и не требует присутствия каких-либо кофакторов или ферментов, что отличает его от посттрансляционного процессинга, а лигирование экстеинов отличает белковый сплайсинг от других форм автопротеолиза, который характерен для созревания ряда белков (гликозиласпарагиназы, белки семейства *hedgehog* и т.п.) [4].

По принятой классификации наименование интееина происходит от сокращенного названия вида, у которого он обнаружен, и названия гена, кодирующего интееин-содержащий белок [3]. Например, упомянутый выше интееин субъединицы вакуолярной АТФазы Vma1 дрожжей *Saccharomyces cerevisiae* получил название SceVMA. Если в молекуле белка интееинов несколько, римскими цифрами обозначают его порядковый номер (TerRIR I, TerRIR II и т.д.). Если интееин обладает эндонуклеазной активностью (см. ниже), то его обозначают по аналогии с эндонуклеазами, кодируемыми интронами 1-го типа: например, PI-PfuI и PI-PfuII – интееины 1 и 2 рибонуклеотидредуктазы *Pyrococcus furiosus*, где “PI” обозначает Protein

Insert, I и II-порядковый номер интееина (если их несколько).

### ОБЩАЯ СХЕМА ОРГАНИЗАЦИИ ИНТЕИНА

Длина всех известных на сегодня интееинов варьирует в среднем от 134 до 650 аминокислотных остатков (а.о.) [5, 6], хотя обнаружены интееины длиной 1308 и 1650 а.о. [6, 7]. Условно интееины делятся на две большие группы – классические интееины и мини-интееины (рис. 1). Классический интееин состоит из двух доменов: Hint-домена, обеспечивающего катализ белкового сплайсинга, и центрального эндонуклеазного домена. У мини-интееинов центральный эндонуклеазный домен заменен на линкерную последовательность, не обладающую каталитической активностью.

Анализ большинства интееинов показал, что в последовательности усредненного интееина можно выделить 10 консервативных аминокислотных мотивов, обозначаемых латинскими буквами A, N2, B, N4, C, D, E, H, F, G (рис. 1а) [8–11]. У мини-интееинов центральные мотивы C, D, E и H отсутствуют (рис. 1б).

Коротко рассмотрим основные особенности каждого мотива (таблица).

**А-мотив.** Это короткая последовательность на N-конце интееина, которая состоит из 13 аминокислотных остатков (а.о.), два из которых (первая

Обобщенная формула аминокислотных последовательностей различных интеиновых мотивов

Домены	Мотивы	Аминокислотная последовательность	Обозначения
N-концевой сплайсинговый домен	A N2 B N4	Ch..Dp.hhh..G ...GD.. G..h.hT..H.hhh .....GD.....	h – гидрофобные остатки (G, V, L, I, A, M) a – кислые остатки (D, E)
Эндонуклеазный домен	C D E H	LhG..hhaG .K.IP..h .L.GhFahDG p.S..hh..h..LL..hGI	г – ароматические остатки (F, Y, W) р – полярные остатки (S, T, C) . – неконсервативный остаток p <sup>+1</sup> – +1 остаток С-экстеина (S, T, C) ▲ – место разрыва на С-конце
С-концевой сплайсинговый домен	F G	rVYDLpV..a..HNFh NGhhhHN p <sup>+1</sup> ▲	

и последняя) высококонсервативны, что свидетельствует об их исключительной важности для инициации и протекания сплайсинга. Практически всегда на N-конце интеина (позиция 1) находится Cys. Очень редко в позиции 1 могут находиться Ala, Gln или Ser. В 13-ой позиции, кроме Gly, иногда находятся Ala, Lys, Thr, Arg, Tyr, Asn.

**Мотив N2** состоит из 7 а.о., из которых один – Asp-5 или Glu-5 – является консервативным. Чаще всего этому остатку предшествует остаток Gly.

**В-мотив** состоит из 14 а.о. Среди них необходимо выделить His-10, который присутствует в данной позиции во всех идентифицированных интеинах. Чаще всего рядом с ним, в позиции 7, находится аминокислота Thr [9]. Эти две консервативные аминокислоты активно участвуют в обеспечении инициации сплайсинга.

**Мотив N4** состоит из 16 а.о., из которых в позиции 11 находится высококонсервативный Asp или Glu. Аналогично N2-мотиву, в большинстве случаев перед ними находится Gly-10. Однако N4-мотив обнаружен не во всех интеинах (отсутствует у SceVMA, CtrVMA, CeuClpP и некоторых других) [11].

Мотивы A, N2, B и N4 формируют N-концевой сплайсинговый домен, функция которого заключается в облегчении разрыва пептидной связи на N-конце интеина [11]. В среднем общая длина N-концевого домена составляет около 150–200 а.о. [11]. Большинство из указанных консервативных аминокислот абсолютно необходимы для N-концевого расщепления – аминокислотные замены в указанных выше позициях часто полностью блокируют инициацию расщепления и сплайсинга белков-предшественников.

**С-мотив и Е-мотив** формируют основу DOD-эндонуклеазного домена [13]. Аналогично известным DOD-эндонуклеазам они содержат последо-

вательности длиной 9 и 10 а.о., которые формируют центры узнавания двухцепочечной ДНК и разделены линкером, состоящим приблизительно из 90–130 а.о. В активном центре выделяют консервативные остатки Gly в позициях 3, 9 и 4, 10 мотивов С и Е соответственно. Каждый мотив содержит также каталитически активные остатки Asn и Lys (см. ниже).

**Д-мотив** (8 а.о.) является линкерной последовательностью между С- и Е-мотивами. Замена Lys во второй позиции этого мотива может полностью блокировать эндонуклеазную активность DOD-домена [9]. Из этого вытекает, что наряду с С- и Е-мотивами Д-мотив участвует в формировании активного эндонуклеазного домена.

**Н-мотив** состоит из 19 а.о., в которых довольно консервативна последовательность Leu-13–Leu-14. Вероятно, они принимают участие во взаимодействии молекулы интеина с ДНК.

Мотивы С, D, E, H формируют DOD-эндонуклеазный домен, который присутствует во многих, но не во всех интеинах. Необходимо подчеркнуть – все экспериментальные данные свидетельствуют о том, что DOD-домен не требуется для протекания белкового сплайсинга [8, 9, 11, 12, 14], но обеспечивает хоуминг интеина, на чем мы остановимся ниже.

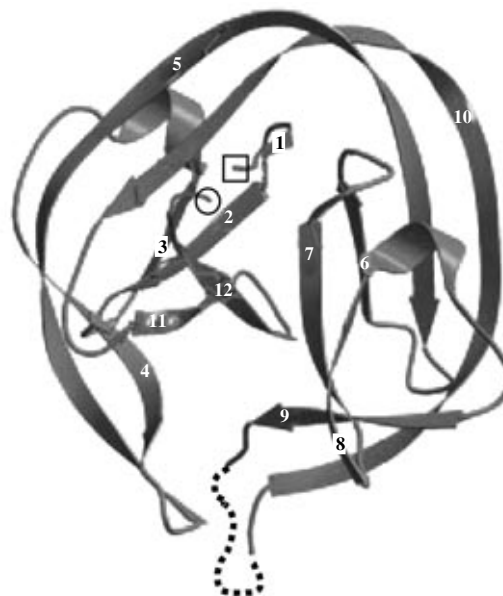
С-концевой сплайсинговый домен, длина которого составляет около 25–40 а.о., образуется мотивами F и G [11]. **Ф-мотив** состоит из 16 а.о., из которых половина высококонсервативна (таблица). **Г-мотив** – это короткая С-концевая последовательность, состоящая из 8 а.о., из которых 7 аминокислот принадлежат интеину и одна – N-концевая аминокислота С-экстеина. F- и G-мотивы разделены очень небольшим линкером, обычно от 2 до 5 а.о. Последней аминокислотой в большинстве случаев является Asn (очень редко – Gln, Asp), а предпоследней – His. Эти две послед-

ние аминокислоты играют важную роль при гидролизе пептидной связи на С-конце интеина [9, 11]; в то же время, для лигирования экстеинов критически важна природа N-концевой аминокислоты С-экстеина (в дальнейшем мы будем ее обозначать термином “+1-аминокислота”). В позиции +1 большинства известных С-экстеинов находится одна из трех аминокислот – Ser, Thr или Cys.

Эндонуклеазный и сплайсинговые регионы образуют в молекуле интеина два пространственно разделенных домена. N- и С-сплайсинговые домены формируют подковообразную структуру, названную Hint-домен [11, 15]. Оба сайта расщепления находятся в центре Hint-домена, на концах двух антипараллельных  $\beta$ -цепей (рис. 2). N- и С-концы интеина сближены до расстояний, соотносимых с длиной ковалентных связей (порядка 3.4 Å). Пространственное расположение фиксировано многочисленными водородными связями, электростатическими и гидрофобными взаимодействиями. Именно пространственное сближение боковых радикалов ключевых аминокислотных остатков играет важную роль в процессе автокаталитического отщепления экстеинов и их последующего лигирования. Отмечено, что структура каталитического центра Hint-домена аналогична структуре активного центра сериновых протеаз [5]. Известно, что сериновые протеазы катализируют разрыв пептидных связей субстрата с помощью высококонсервативных остатков Ser, His и Asp [5]. Автосплайсинг интеина катализируется сходным образом – при участии высококонсервативных остатков Ser, Thr или Cys на N-конце интеина и His-Asn-Ser или His-Asn-Cys на С-конце. Например, у интеина MxeGyrA каталитический центр формируется при участии Ser-1, His-197, Asn-198, а также Thr-72, Asn-74 и His-75. Вся эта структура сближена с N-концевым сплайсинговым сайтом [4]. То есть и в случае сериновых протеаз, и в случае интеинов боковая группа His активирует гидроксильную группу Ser, которая непосредственно координирует гидролиз пептидной связи при участии молекул воды [5, 11]. На рис. 3 представлено сравнение пространственного расположения ряда активных групп каталитического центра мини-интеина MxeGyrA и химотрипсина. Интересно, что пространственное положение двух аминокислот идентично, несмотря на то, что сами сравниваемые белки достаточно гетерогенны.

#### СОВРЕМЕННЫЕ ПРЕДСТАВЛЕНИЯ О МЕХАНИЗМЕ БЕЛКОВОГО СПЛАЙСИНГА

Многими исследователями показано, что белковый сплайсинг является аутокаталитическим процессом и для своего осуществления не требует присутствия других кофакторов [2, 5, 9, 16]. Однако определить точный механизм сплайсинга белков долгое время не удавалось – процесс протекает

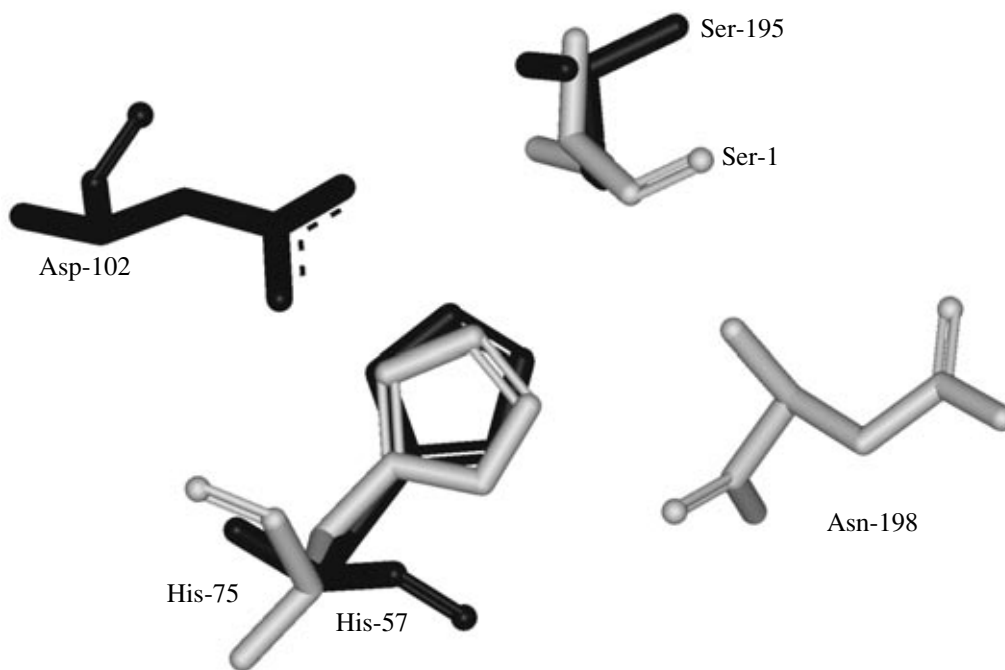


**Рис. 2.** Ленточная модель N- и С-сплайсинговых доменов интеина SspDna B.  $\beta$ -складки обозначены стрелками, квадратом – N-концевой сайт расщепления, кругом – С-концевой. Центральная часть интеина (условно обозначена пунктиром) не показана.

очень быстро, и обычными методами обнаружить промежуточные соединения не представлялось возможным. Решить эту задачу удалось группе Ф. Перлер. Они показали, что замены ряда консервативных аминокислот, в первую очередь последовательности His-Asn(Gln) и +1-аминокислоты экстеина, блокируют белковый сплайсинг на разных этапах, что приводит к накоплению промежуточных продуктов реакции [5, 11]. Исследование последних из упомянутых авторов позволило предложить механизм белкового сплайсинга.

Согласно принятой теории, белковый сплайсинг состоит из серии последовательных нуклеофильных перестановок [5, 11, 16, 17]. Событием, запускающим белковый сплайсинг, является аутокаталитический N–O- или N–S-сдвиг (первая аминокислота на N-конце интеина Ser или Cys соответственно) на N-концевом сайте сплайсинга (рис. 4, шаг 1). В результате образуется высокореакционная эфирная или тиоэфирная связь. Возможность такой перестановки неоднократно подтверждалась экспериментально, поскольку у интеин-содержащих белков обнаруживались эфирные связи, не характерные для белков. В то же время, образование таких связей не наблюдается при нарушении третичной структуры интеина детергентами, что свидетельствует о взаимосвязи между появлением эфирных связей и протеканием белкового сплайсинга [18].

С точки зрения химии, N–O/N–S-сдвиг не является энергетически выгодным процессом, по-



**Рис. 3.** Сравнение пространственного расположения боковых радикалов некоторых консервативных аминокислот в активных центрах сериновой протеазы (химотрипсин, черный, His-57; Asp-102; Ser-195) и трипсина (MxeGytA, серый, Ser-1; His-75; Asn-198).

скольку в результате реакции происходит разрыв амидной (пептидной) связи и образуется высокоэнергетическая эфирная/тиоэфирная связь. Поэтому такой процесс должен протекать в присутствии катализатора. Действительно, реакция разрыва пептидной связи на N-конце обусловлена, как минимум, двумя факторами. Во-первых, процесс катализируется Hint-доменом интеина. Во-вторых, на эффективность N-концевого расщепления определенное влияние оказывают экстеины. В случае интеина MxeGytA пептидная связь, связывающая N-экстеин и первую аминокислоту интеина, находится в редкой и не характерной для белков *цис*-конформации. Поскольку такая связь энергетически невыгодна, то ее наличие облегчает протекание N–O- или N–S-сдвигов, т.е. инициацию сплайсинга [4]. Замена природных экстеинов рекомбинантными белками часто уменьшает или даже блокирует N-концевое расщепление [19].

На втором этапе белкового сплайсинга происходит нуклеофильная атака образовавшейся эфирной связи OH- или SH-группой +1-аминокислоты C-экстеина. В результате происходит реакция трансэтерификации, т.е. перенос N-концевого экстеина на боковую группу +1-аминокислоты C-экстеина (рис. 4, шаг 2) и образуется разветвленное промежуточное соединение – белок, у которого имеется два N-конца. Такая перестановка приводит к смещению зарядов, что в свою очередь индуцирует циклизацию боковой группы Asn на C-конце интеина с образованием сукцинимида

(рис. 4, шаг 3). Действительно, возникновение сукцинимида было показано с помощью хроматографии C-концевых пептидов свободного интеина, отщепленных в результате обработки бромцианом, а также с помощью масс-спектрологии [20]. Катализ этого процесса также обеспечивается Hint-доменом интеина. Наиболее важную роль в образовании разрыва пептидной связи на C-конце интеина играет высококонсервативный предпоследний остаток His [5, 20]. Кроме того, на эффективность расщепления сильно влияет природа C-экстеина: например, замена природного экстеина биотехнологическими аналогами может уменьшить или заблокировать C-концевое расщепление [19].

Циклизация боковой группы Asn приводит к разрыву пептидной связи между интеином и C-экстеином – разветвленная структура распадается на свободный интеин и лигированные экстеины, связанные друг с другом эфирной связью (рис. 4, шаг 3). Последний шаг сплайсинга белка происходит спонтанно (рис. 4, шаг 4). Как указано выше, эфирная/тиоэфирная связь энергетически невыгодна, поэтому происходит обратный N–O- или N–S-сдвиг с образованием энергетически более выгодной пептидной связи, что и завершает процесс лигирования экстеинов (рис. 4, шаг 4б). В свою очередь на C-конце интеина сукцинимид может гидролизироваться с образованием аспарагина или изоаспарагина (рис. 4, шаг 4а) [20].

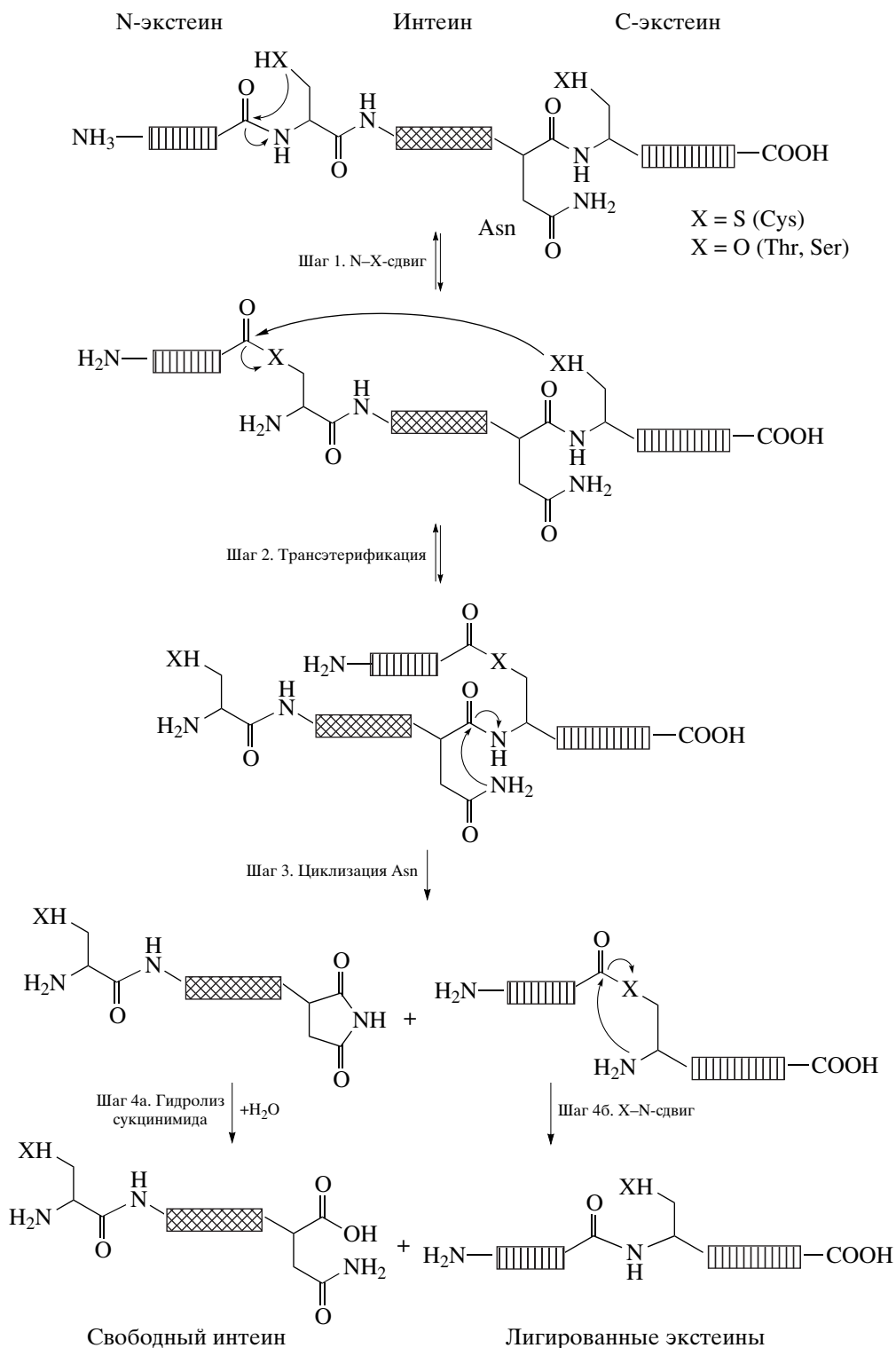


Рис. 4. Схема классического белкового сплайсинга [9].

Более детальное описание сложного, с точки зрения химии, процесса, в этом обзоре мы не приводим. Подробное описание реакции можно найти в работе Динга и соавт. [11].

#### ОСНОВНЫЕ ПРИЗНАКИ ИНТЕИНОВ

Как видно, интеины по своей структуре довольно гетерогенны, однако всем им присущи некоторые общие свойства. В настоящее время вы-

деляют четыре основных признака интеиновой последовательности.

1. В гене-носителе интеина имеется последовательность, которой нет в аналогичных генах других организмов.

2. Размеры зрелого белка отличаются от размеров, предсказанных в результате анализа его кодирующей последовательности.

3. Присутствие специфических мотивов A, B, F, G (наличие или отсутствие DOD-эндонуклеазного домена характерным признаком интеина не является).

4. Наличие четырех консервативных аминокислотных остатков: на N-конце предполагаемого интеина Ser, Thr или Cys; на C-конце – His-Asn или His-Gln; на N-конце C-экстеина – Ser, Thr или Cys [9].

Существуют и некоторые исключения из этих правил. Коротко рассмотрим их.

#### *Интеины без предпоследнего His*

В ряде интеинов предпоследняя аминокислота (His) заменена на другую [21]. И хотя наличие у большинства интеинов предпоследнего His критически важно для обеспечения циклизации Asp, почти все интеины, содержащие такую замену, сохраняют ту или иную степень активности (за исключением неактивного интеина ClpPIS2, вместо His содержащего Gly [21, 22]). Генно-инженерная “обратная” замена предпоследней аминокислоты на His не всегда дает однозначный результат: в двух случаях она снижала уровень сплайсинга, в трех случаях – увеличивала. В случае интеина MjaK1bA такая замена усиливает C-концевое отщепление, но не сплайсинг экстеинов [19], в случае Mje RNAApo1A замена предпоследнего Ser на His снижает выход сплайсированного активного белка на 27%, хотя эффективность N-концевого отщепления не изменяется, в случае неактивного интеина ClpPIS2 замена полностью восстанавливает сплайсинговую активность [21, 22]. Все эти данные могут свидетельствовать, во-первых, о том, что все интеины изначально содержали в качестве предпоследней аминокислоты His, которая в результате точечной мутации могла быть утеряна; во-вторых, что интеины с измененными предпоследними аминокислотными остатками могут вытеснить свой His-содержащий гомолог за счет приобретения каких-либо, пока неизвестных, свойств; в-третьих, что такие интеины могут являться представителями различных этапов развития интеинов, поскольку аналогов этих интеинов, содержащих His, не обнаружено [19]. Кроме того, логично предположить, что в таких мутантах роль предпоследней аминокислоты мог бы взять на себя другой His (возможно, даже “в теле” экстеина), про-

странственно сближенный с C-концевым сайтом расщепления.

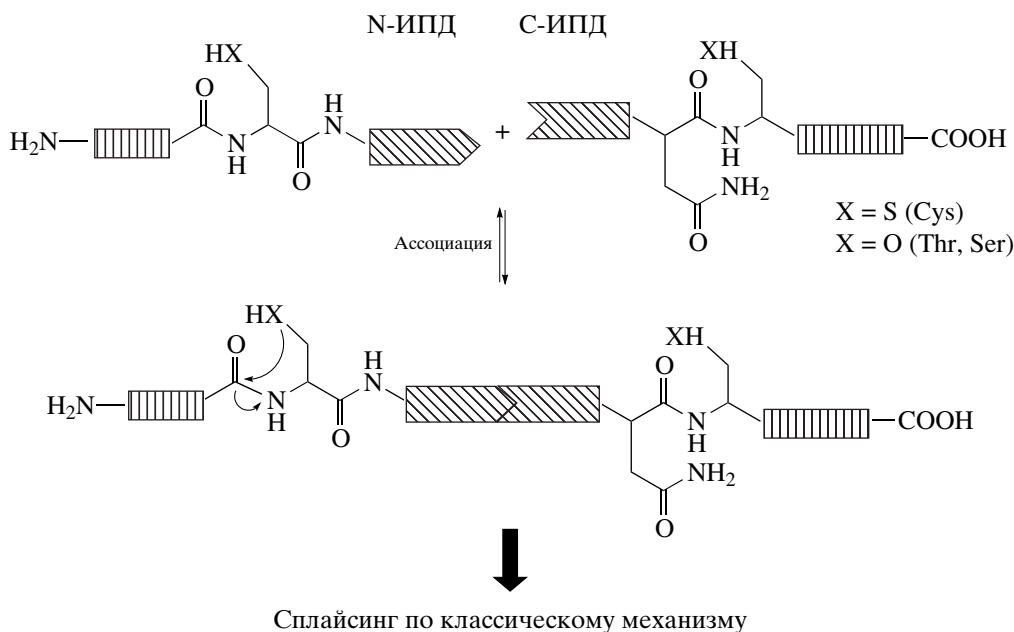
#### *Интеины с последней аминокислотой Gln*

По крайней мере, три функционально активных интеина – CIVRNR, PabPolIII и ChuRNR – имеют на C-конце аминокислотные остатки, отличные от Asp: Gln у CIVRNR и PabPolIII; Asp у ChuRNR. Как и в случае с His, в опытах по обратной замене Gln на Asp не получено однозначного результата: при модификации ChuRNR количество сплайсированного белка снижается в 7 раз, а в случае PabPolIII – повышается в 4 раза. Вероятно, C-концевой разрыв пептидной связи после Gln возникает вследствие его циклизации (с образованием аминоклутаримида). В случае же ChuRNR замена Asp на Asp полностью блокирует сплайсинг, однако не препятствует C-концевому расщеплению. Интересно, что разрыв пептидной связи на C-конце ChuRNR не блокируется даже при замене Asp на Ala, для которого циклизация боковой группы в принципе невозможна [19]. Объяснить этот факт пока что не удается.

#### *Интеины с первой аминокислотой Ala*

У трех интеинов – DraSnf2, MjaK1bA и MavDnaB – первой аминокислотой является Ala. На примере интеина MjaK1bA показано, что сплайсинг в этом случае обеспечивается прямой нуклеофильной атакой пептидной связи на N-конце интеина радикалом +1-аминокислоты C-концевого экстеина. Сплайсинг у таких белков начинается прямо со второго шага механизма, описанного выше (рис. 4). Вероятно, в этом случае N-концевой разрыв происходит легче благодаря особенностям трехмерной структуры данных интеинов, что компенсирует отсутствие первого шага – разрыва пептидной связи на N-конце интеина путем N=O/N-S-сдвига. В пользу такого предположения свидетельствует то, что в большинстве других интеинов генно-инженерная замена первой аминокислоты на Ala практически полностью блокирует N-концевое расщепление [19].

Перечисленные выше исключения наводят на мысль, что могут существовать другие варианты механизма сплайсинга, которые пока не изучены. В пользу этого свидетельствует то, что существуют интеины, сильно отличающиеся от классических, например, MjaK1bA. Этот мини-интеин (168 а.о., NCBI # Q58191) белка K1bA из архебактерии *Methanocaldococcus jannaschii* не содержит эндонуклеазного домена, в качестве первой аминокислоты вместо Ser/Cys у него выступает Ala, а предпоследний His заменен на Ser. Аналогичное строение имеют интеины гомологичных белков у *Pyrococcus abyssi* (PabK1bA), *P. furiosus* (PfuK1bA), *P. horikoshii* OT3 (PhoK1bA), *Thermococcus kodaka-*



**Рис. 5.** Схема механизма белкового *транс*-сплайсинга: N-ИПД – N-концевой интеин-подобный домен; С-ИПД – С-концевой интеин-подобный домен.

*raensis* KOD1 (ТкоK1bA), хотя у трех последних эндонуклеазный домен уже присутствует [9]. Точный механизм, по которому они обеспечивают свои функции, пока неизвестен.

Кроме *цис*-сплайсинга, т.е. автокаталитического удаления интеина из белка-предшественника, у многих организмов обнаружено явление *транс*-сплайсинга [15, 23, 24], в результате которого происходит лигирование двух белков, кодируемых двумя различными генами. Этот процесс опосредуется так называемыми интеин-подобными доменами (ИПД), которые находятся на N- или С-конце обоих белков. Обладая высоким средством друг к другу, они могут объединяться, образуя полноценный интеин (рис. 5). Далее сплайсинг проходит по механизму *цис*-сплайсинга, описанному выше, в результате чего происходит лигирование полипептидов, прилегающих к интеин-подобным доменам. По такому механизму, например, происходит образование белка DnaE у *Synechocystis* sp. PCC6803 [15].

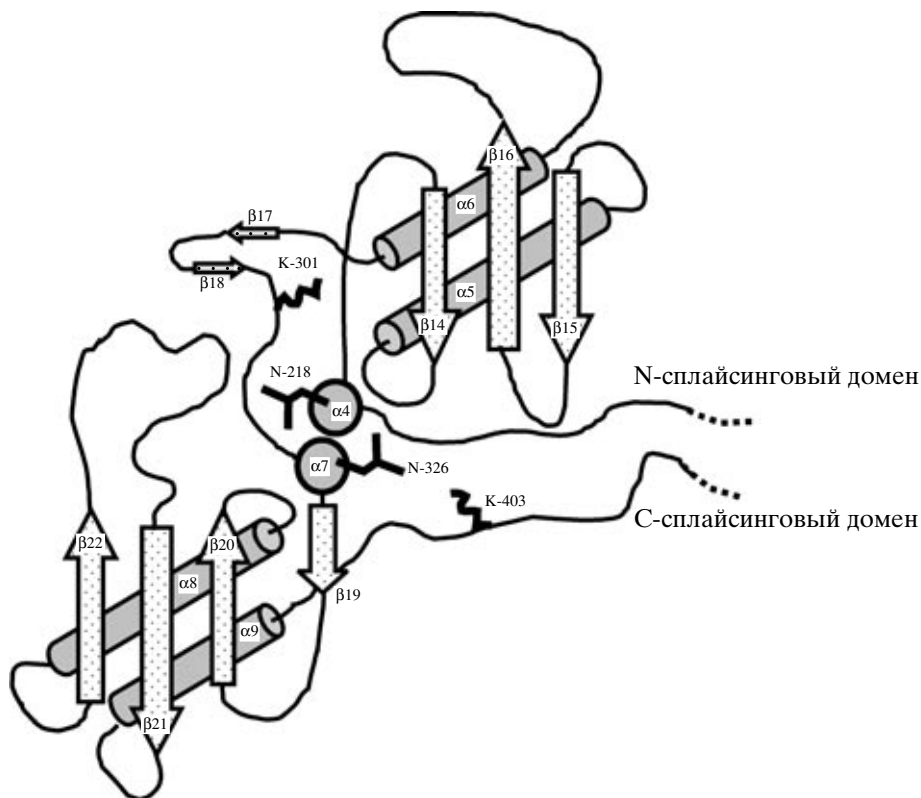
### ЭНДОНУКЛЕАЗНАЯ АКТИВНОСТЬ ИНТЕИНОВ

Значительная часть интеинов обладает эндонуклеазной активностью, которая обеспечивает направленный перенос последовательности интеина из гена, ее содержащего (аллель (+)), в аллельный ген, ее не содержащий (аллель (-)) [5, 7, 18]. Этот феномен называется хоумингом интеинов, по аналогии с хоумингом интронов мРНК группы I [24, 25].

Хоуминг-эндонуклеазы – это большой класс сайт-специфических ДНКаз, которые кодируются мобильными генетическими элементами (интроны группы I, интеины) [13]. Все хоуминг-эндонуклеазы разделяют на 4 семейства в соответствии со строением их активного центра: LAGLIDADG-, или DOD-семейство, GIY-YIG-семейство, His-Cys-семейство, HNH-семейство [18]. Большинство интеинов – DOD-эндонуклеазы, в активном центре которых находится высококонсервативная последовательность LAGLIDADG. Они содержат два DOD-мотива, которые составляют основу DOD-эндонуклеазного домена (рис. 1). По принципу действия DOD-эндонуклеазы являются ДНК- и  $Mg^{2+}$ -зависимыми АТРазами [14], причем присутствие  $Mg^{2+}$  необходимо для обеспечения эндонуклеазной активности, а не для связывания с ДНК [26, 27]. Описано также несколько интеинов (*SspGyrB*, *TerGyrB*), которые содержат HNH-эндонуклеазные домены. Кроме того, целый ряд интеинов несет укороченный набор мотивов в эндонуклеазном домене – С, D, Н (*Seu*, *ClpP*); С, Е, Н (*CivRIRI*, *Mch RecA*, *Mfa RecA*); D, Е, Н (*MjaHyp-1*, *MjaRpolA*) и т.д. [9]. Отдельно стоят мини-интеины, которые вообще не содержат эндонуклеазных доменов (*MxeCyrA*, *Mja KlbA* и др.).

Структура большинства DOD-доменов сходна [13, 14, 28]. Рассмотрим строение DOD-домена на примере *PI-SceI* – эндонуклеазы интеина *SceVMA* – одного из первых интеиновых эндонуклеазных доменов, чья структура определена с помощью рентгеноструктурного анализа (рис. 6). В середине эндонуклеазного домена две антипараллельные





**Рис. 6.** Структура эндонуклеазного домена PI-SceI (интеин SceVMA) [19]. Консервативные аминокислотные остатки обозначены однобуквенными символами.

$\alpha$ -спирали ( $\alpha$ -4 и  $\alpha$ -7) формируют активный центр, содержащий два LAGLIDADG-мотива – по одному на каждой  $\alpha$ -спирали. Непосредственно с ДНК связываются по 6 аминокислотных остатков каждого DOD-мотива [18]. На С-концах этих спиралей находятся два высококонсервативных аспарагина – Asn-218 и Asn-326. Они абсолютно необходимы для обеспечения эндонуклеазной активности. Третья необходимая аминокислота – Lys-301. Asn-218, Asn-326 и Lys-301 формируют каталитическую триаду активного центра эндонуклеазы [29]. Кроме того, в ряде работ подчеркивается важная роль симметричной Lys-301 аминокислоты Lys-403 (рис. 6) [26].

Эндонуклеазные домены интеинов, как правило, узнают непалиндромные ДНК-последовательности длиной около 14–40 п.н. [12, 28]. Кроме DOD-домена, в обеспечении эндонуклеазной активности принимает участие N-концевой сплайсинговый домен, содержащий ДНК-узнающий регион (DRR–DNA recognition region) [30]. Значение N-конца интеина для обеспечения эндонуклеазной активности показано с помощью сайт-направленного мутагенеза. Замены определенных аминокислот в N-концевом сплайсинговом домене резко снижают специфичность и активность DOD-домена [28, 30]. Оказалось, что интеины

узнают два специфических участка ДНК. Первым является сам сайт рестрикации, с которым связывается DOD-домен. Второй участок – дистальная область целевой последовательности, узнается DRR-регионом интеина [30, 31]. Этот участок включает в себя около 15–17 н. и существен для специфического связывания и узнавания точки рестрикации [26].

Механизм хоуминга показан на рис. 7. Интеин катализирует двухцепочный разрыв в аллель(–) в точке интеиновой вставки. При этом образуется 3'-выступающий конец из четырех нуклеотидов (“липкий” конец) [28–30]. Далее клеточными ферментами проводится репарация двухцепочечного разрыва с использованием в качестве матрицы аллельной ДНК, содержащей последовательность интеина. Это обеспечивает его перенос в аллель(–) [5, 9]. Поскольку исходной аллелью может быть и плазмидная, и вирусная ДНК, то таким образом может обеспечиваться и горизонтальный перенос, что приводит к распространению интеинов. Однако на самом деле процесс хоуминга происходит крайне редко в силу ряда причин [32].

Около 15% известных на сегодня интеинов (мини-интеины) не содержат центрального эндонуклеазного домена [9, 11]. Наиболее изучен ми-

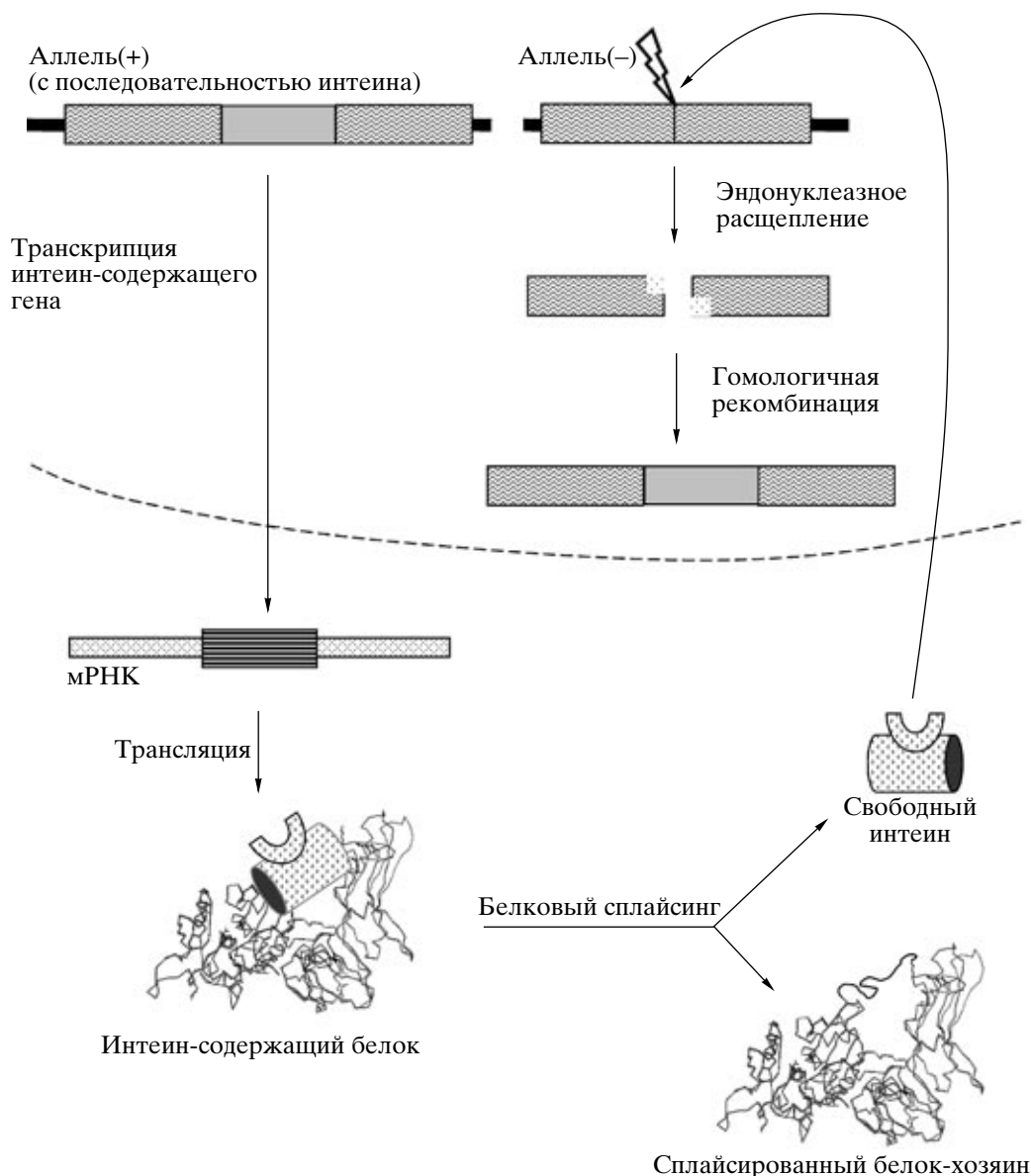


Рис. 7. Механизм хоуминга интеинов.

ни-интеин белка GyrA у *Mycobacterium xenopi* (MxeGyrA) [4, 11]. При анализе локуса *gyrA* у различных микобактерий показано, что интеины имеются в той же позиции у *M. leprae*, *M. kansasii*, *M. flavescens* и *M. goodii*. Однако только у Mxe GyrA отсутствует центральный эндонуклеазный домен (отсутствие вставки между 108 и 142 а.о.) [4].

Сравнительный анализ аминокислотных последовательностей мини-интеинов и классических интеинов свидетельствует о том, что DOD-домен был приобретен эволюционно позже. Поэтому мини-интеины считаются эволюционно более древними формами. Показано, что Hint-домен интеинов состоит из двух высокогомологичных субдоменов, которые, вероятно, являются продуктами дупли-

кации [4, 9, 33, 34]. Эндонуклеазный домен у классических интеинов находится посередине второго субдомена в виде вставки, нарушая его целостность [9]. У мини-интеинов такая вставка отсутствует. В то же время, N- и C-сплайсинговые домены, не затронутые вставкой, у всех интеинов высокогомологичны. Кроме того, Hint-домены, гомологичные мини-интеинам, обнаружены в ряде других белков, не обладающих эндонуклеазной активностью (Hh-белки, VII-домены, см. ниже) [32–34]. Это свидетельствует о том, что DOD-домен был приобретен уже функционировавшими мини-интеинами.

Приобретение специфической эндонуклеазной активности значительно облегчило распростра-

нение классических интеинов по сравнению с мини-интеинами. Возможно, поэтому сегодня мини-интеинов обнаруживается значительно меньше. Дискуссионным является вопрос о способности к хоумингу мини-интеинов. Теоретически, их хоуминг может опосредоваться эндонуклеазой, кодируемой другим, связанным или независимым, геном [4], однако это утверждение нуждается в доказательствах.

### РАСПРОСТРАНЕНИЕ ИНТЕИНОВ

Очевидно, интеины возникли на ранних этапах эволюции [5, 32]. Об этом косвенно свидетельствует то, что интеины имеются у некоторых представителей всех трех ветвей жизни – эукариот, эубактерий и архебактерий [9, 35, 36]. Однако распространение их весьма неоднородно. Например, у *Methanococcus jannaschii* интеины обнаружены в 19 генах. С другой стороны, названо более 30 видов (прокариот и эукариот), в геномах которых интеинов нет (*Arabidopsis thaliana*, *Agrobacterium tumefaciens* C58 Cereon, *Escherichia coli* штамм K-12, *Lactococcus lactis* IL1403, *Staphylococcus aureus* N315 *Vibrio cholerae* El Tor N16961, *Drosophila melanogaster*, *Homo sapiens* и др.) [9].

Подавляющее большинство интеинов обнаружено у прокариот, в основном архебактерий и эубактерий. Чаще всего они встречаются в генах ферментов, взаимодействующих с нуклеиновыми кислотами (ДНК- и РНК-полимеразы, RecA, GyrA, DnaB, хеликазы, гиразы, лигазы и т.д.) или с производными нуклеотидов (VMA, анаэробная рибонуклеозидтрифосфатредуктаза, UDP-глюкозодегидрогеназа и т.д.) [5, 9]. Однако интеины обнаружены и в других группах ферментов (например, фосфоенолпируватсинтазе). Как правило, в генах находят один, реже – два интеина. Однако есть и исключения. Одним из них является ген рибонуклеотидредуктазы океанской N<sub>2</sub>-фиксирующей цианобактерии *Trichodesmium erythraeum*. Ген (13650 п.н.) содержит 4 интеина (Ter RIR1 – Ter RIR4) и 3 интрона типа II, а сам белок кодируется фрагментом, составляющим только 17% от общей последовательности гена. При этом каждый интеин функционален и обеспечивает сплайсинг экстеинов в системе экспрессии в *E. coli*. Стоит также упомянуть два гигантских интеина, обнаруженных у цианобактерий *Trichodesmium erythraeum* IMS101. Это первые интеины генов хеликазы *dnaE-d* (TerDnaE-1 1308 а.о.) и  $\alpha$ -субъединицы ДНК-полимеразы *dnaB-b* (Ter DnaB-1 1650 а.о.). Отметим, что в каждом гене имеются дополнительно еще несколько участков, кодирующих интеины, но они значительно короче (TerDnaE-2 – 428 а.о., TerDnaE-3с – 36 а.о., TerDnaE-3п – 102 а.о. и TerDnaB-2 – 177 а.о.) [7, 9].

Что касается эукариот, интеины найдены примерно в 50 таксонах, но, в основном, в геномах

простейших эукариот. Например, у *S. cerevisiae* (интеин белка VMA SceVMA) и *Debaryomyces hansenii* (Dha VMA), *Candida guilliermondii* и *Podospira anserine* (интеин глутаматсинтазы Cgu GLT1 и PanGLT1), *Rhodothermus marinus*, *Podospira anserine* (интеин хитинсинтазы Pan CHS2), *Cryptococcus neoformans* (мини-интеин белка сплайсеосом PRP8 Cne PRP8) и его DOD-содержащий гомолог у *C. laurentii* и у некоторых аскомицетов (Ani PRP8, Nfi PRP8, Afu PRP8, Hca PRP8, Cla PRP8) [9, 32, 36–40]. Кроме того, интеины обнаружены в пластидах ряда одноклеточных. Например, в хлоропластах *Chlamydomonas eugametos*, в гене протеазы *clpP*, обнаружена последовательность (вставка IS2), обладающая большинством признаков интеина, за исключением того, что предпоследней аминокислотой вместо His является Gly-455, что, как было упомянуто выше, блокирует сплайсинг белка ClpP при его наработке в *E. coli* [21]. Интеиновые последовательности обнаружены также в ДНК хлоропластов водорослей *Porphyra purpurea*, *Euglena gracilis* и *Chlorella vulgaris* и других [9, 32, 39, 40].

Интеины обнаружены в геномах ряда прокариотических и эукариотических вирусов [41–44]. Например, интеин APMV Pol (351 а.о.) кодируется геном *polB* недавно описанного ДНК-содержащего гигантского Мимивируса [43, 44]. Присутствие последовательностей интеинов в геноме вирусов является лучшей иллюстрацией механизма, обеспечивающего горизонтальный перенос интеинов. Действительно, появление в клетке интеин-кодирующей последовательности может создавать дополнительную аллель (при условии, что последовательность интеина фланкирована геномной ДНК клетки-хозяина). Точно так же в геном вируса, плазмиду и другую эписомную ДНК может переноситься последовательность интеина из ДНК клетки-хозяина [5]. Действительно, горизонтальный перенос информации – один из важных факторов, обуславливающих распространение интеиновых последовательностей. Об этом свидетельствует тот факт, что интеины гомологичных белков у разных видов часто имеют много общего. О возможности горизонтального переноса интеинов между прокариотами и эукариотами свидетельствует следующий пример: интеины гена *dnaB* у *Rhodothermus marinus* и *Synechocystis* sp. PCC6803 имеют гораздо более высокую гомологию, чем экстеины [39, 45]. Другой пример – эукариотический интеин Pan CHS2 хитинсинтазы, обнаруженный только у *P. anserine*; его аналогов в генах хитинсинтаз других видов пока что не описано. Более того, у *P. anserine* +1-аминокислота С-концевого экстеина необходима для сплайсинга Cys, тогда как у остальных гомологичных хитинсинтаз в этом месте обычно располагается неактивный Val. Кроме того, неожиданным оказался тот факт, что последователь-

ность *Ran* CHS2 высокогомологична интеину глутаматсинтазы *Ran* GLT1 [32]. Это позволяет выдвинуть гипотезу о возможности переноса кодирующей последовательности интеина между неаллельными генами, возможно, за счет эндонуклеазного разрыва сходного, но неидентичного сайта [32]. В пользу этой гипотезы свидетельствует то, что в ДНК *M. leprae* и *M. tuberculosis* имеются последовательности трех гомологичных интеинов в генах *ppls1*, *recA* и *dnaB*, однако каждый из них вставлен в негомологичных сайтах [9]. Чтобы обеспечить такой перенос, после неспецифического эндонуклеазного разрыва должна произойти неспецифическая репарация, где матрицей будет служить некомплементарная цепь. Возможно ли это – авторами не обсуждается [32].

Однако ряд данных свидетельствует против того, что интеины произошли от одного белка-предшественника и распространились путем простого горизонтального и вертикального переноса между видами [35]. Хотя все известные интеины имеют сходное строение концевых каталитических доменов, в общем строении для них характерна крайне высокая гетерогенность [45, 46]. Так, например, гомология кодирующих последовательностей генов ДНК-гираз у шести видов *Mycobacteria* составляет около 96%, в то время как последовательности интеинов в этих генах имеют гораздо более низкую гомологию – около 65%. Это можно объяснить тем, что после горизонтального переноса в процессе эволюции каждый интеин приспособляется к конкретному хозяину, постепенно утрачивая гомологию с интеинами в других организмах [47].

## АНАЛОГИ ИНТЕИНОВ

Hint-домены обнаружены в белках других семейств, автопроцессинг которых базируется на превращениях, аналогичных белковому сплайсингу. Среди них можно назвать бактериальные интеин-подобные домены А- и В-типов (VIL-домены) и белки семейства *Hedgehog* (Hh) у многоклеточных. Кроме того, обнаружено несколько семейств ферментов, автопроцессинг которых инициируется N–O-сдвигом аналогично тому, как это происходит на первом шаге белкового сплайсинга [39, 45–48].

### *Интеин-подобный процессинг эмбриональных сигнальных белков семейства Hh (Hedgehog Protein)*

Сигнальные белки семейства Hh обнаружены у человека, мыши, курицы, лягушки, рыбы, морского ежа, пиявки и насекомых и у других многоклеточных. Во многих работах показано исключительно важное влияние белков семейства Hh на процессы формирования эмбриона [5, 45, 49, 50]. Для рассмотрения в данном обзоре важен автока-

талитический процесс созревания этих белков. Hh-белки экспрессируются в виде неактивных предшественников, состоящих из двух доменов – сигнального N-концевого домена (Hh-N) и каталитического C-домена (Hh-C). Hh-C образует типичный Hint-домен, в составе которого обнаружены консервативные мотивы, аналогичные мотивам интеинов А и В. Сходство в строении Hint-доменов интеинов и Hh-белков свидетельствует об их возможном общем происхождении. Но организмы-хозяева у них различные: интеины обнаружены, в основном, у прокариот, одноклеточных эукариот, в пластидах и вирусах. Hh-белки обнаружены пока только в многоклеточных организмах [51]. В отличие от интеинов, расщепление Hh-белка идет по одному сайту, разделяя белок на две, а не на три части. Кроме того, для протекания реакции необходим экзогенный индуктор – молекула холестерина.

Механизм созревания Hh-белков следующий. В точке расщепления находится аминокислотный остаток Cys. Расщепление инициируется в результате N/S-сдвига, катализируемого Hint-доменом, что аналогично первому шагу стандартного механизма белкового сплайсинга (рис. 8б). Следующий шаг – нуклеофильная атака гидроксильной группы холестерина карбонильной группы пептидной связи на C-конце сигнального домена Hh-N. Это вызывает разрыв пептидной связи между доменами Hh-N и Hh-C и приводит к ковалентной сшивке холестерина и Hh-N. Модификация холестерина сигнального Hh-N-домена обеспечивает его миграцию и встраивание в плазматическую мембрану в качестве клеточного рецептора или выход белка из клетки, что позволяет ему выполнять свои функции [45, 50].

### *VIL-домены*

VIL-домены состоят из 130–155 а.о. и имеют высокую гомологию с Hint-доменами интеинов и Hh-белков. Однако есть ряд отличий, характерных именно для VIL-доменов. В первую очередь, во всех них имеется один или два дополнительных мотива длиной 10–14 а.о. В VIL-доменах, в отличие от интеинов, отсутствуют эндонуклеазные домены, а в отличие от Hh-белков они фланкированы полипептидными последовательностями с двух боков. Кроме того, VIL-домены обнаружены только у одноклеточных [51]: более чем у 50 видов бактерий и у некоторых видов реснитчатых.

Различают два типа VIL-доменов – А и В. VIL-домены А-типа (VIL-A) имеют аналогичный интеинам Hint-домен, состоящий из сходных консервативных мотивов, а также еще одного дополнительного (14 а.о.). У VIL-A два C-концевых остатка полностью соответствуют таковым у интеинов – His-Asp, перед ними почти всегда находится Val. Функции VIL-A заключаются в катализе N- и C-концевого отщепления, а также сплайсинга прилегающих полипептидов.

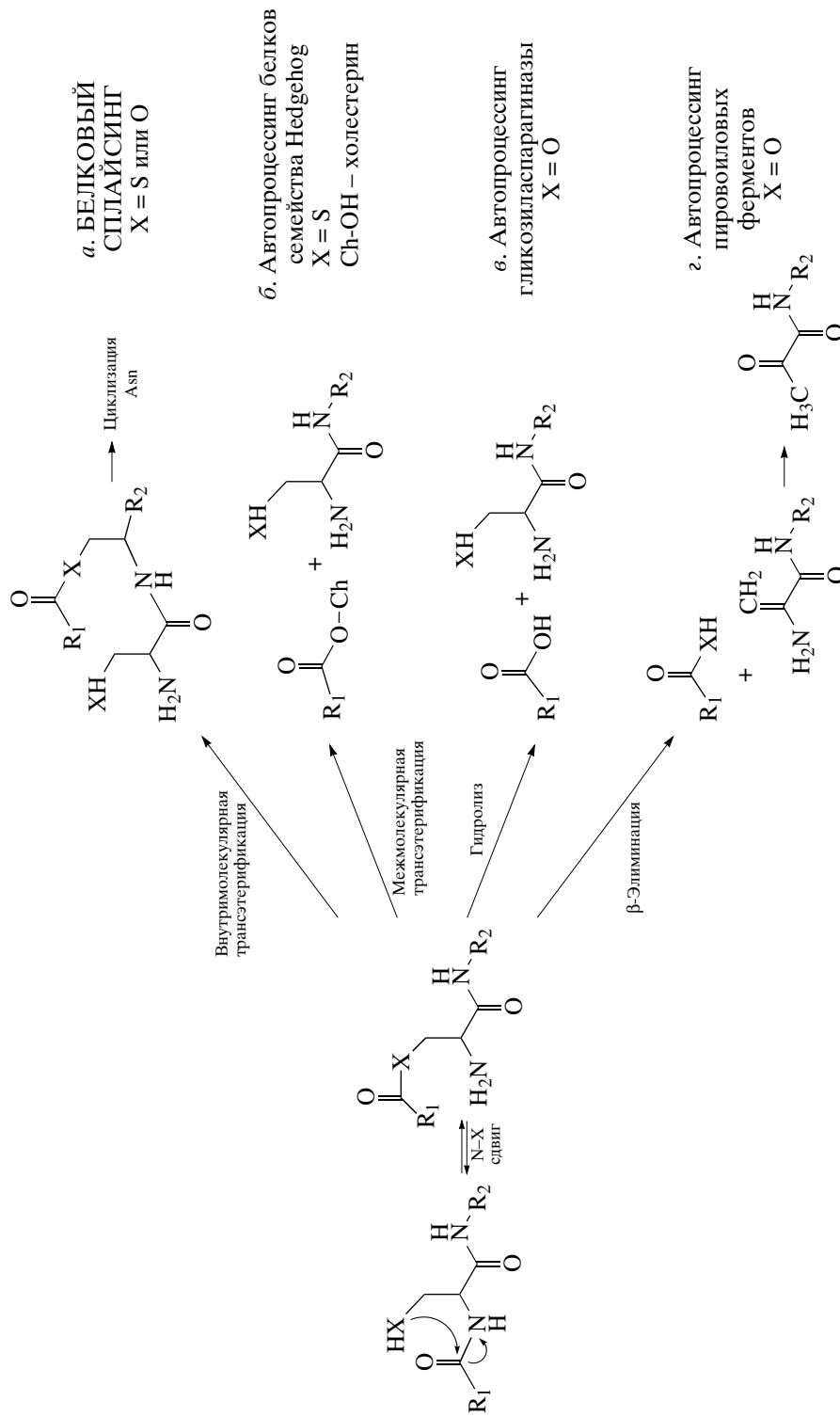


Рис. 8. Роль N–O и N–S ацильных перестановок в обеспечении автопроцессинга различных белков [48].

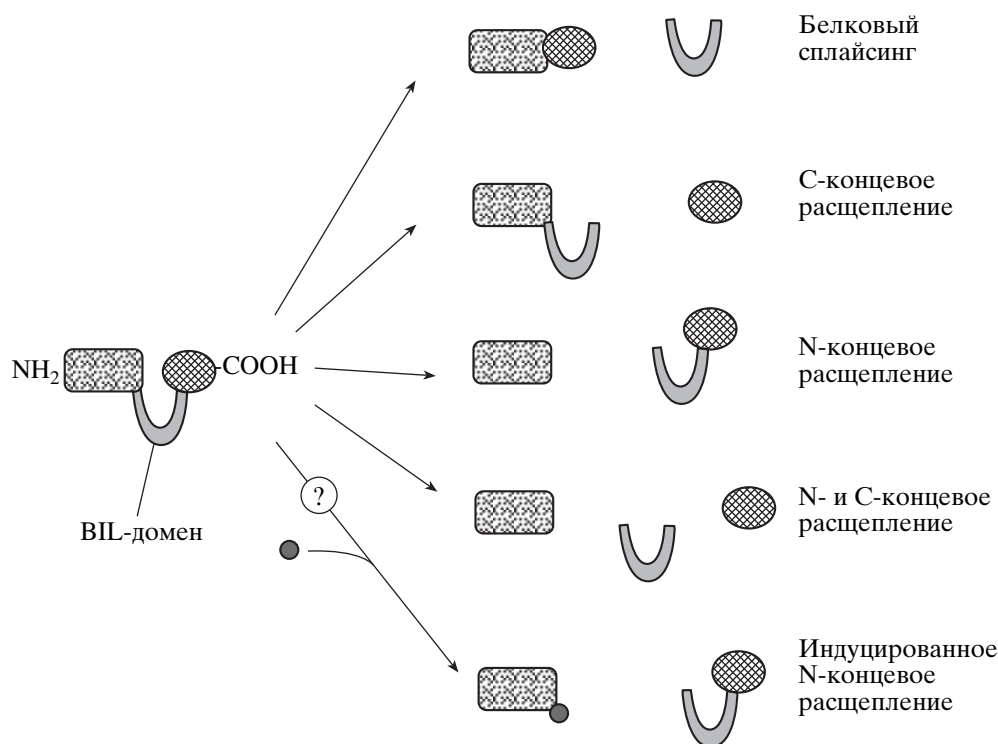


Рис. 9. Схема возможных преобразований белков, опосредованных активностью VIL-доменов.

VIL-домены В-типа (VIL-B) содержат аналогичные интеинам консервативные мотивы, а также дополнительно два специфических мотива длиной 10 и 11 а.о. На С-конце VIL-B предпоследней аминокислотой могут быть цистеин, серин или треонин, в то время как у практически всех интеинов в этом положении находится гистидин (присутствие цистеина на С-конце характерно для Hh-белков). Функции VIL-B заключаются в катализе отщепления или N-, или С-фланкирующей последовательности. Сплайсинг прилегающих полипептидов он не катализирует [52].

Различные возможные перестройки, опосредованные VIL-доменами, обобщены нами на рис. 9. Механизмы реакций, опосредуемых VIL-доменами, отличаются от механизма сплайсинга классических интеинов в деталях, однако принцип сходен. Отщепление на С-конце у VIL-A происходит, вероятно, в результате циклизации Asp на С-конце. VIL-B, который не имеет консервативной пары His-Asp на С-конце, катализирует, возможно, только N-концевое отщепление, а С-концевой разрыв может катализироваться каким-либо экзогенным агентом, как это происходит, например, у Hh-белков [51].

VIL-домены могут обладать довольно разнообразными функциями, поскольку они часто обнаруживаются в гипервариабельных белках бактерий. Например, VIL-домен найден в предшественнике адгезина FhaV у *Pseudomonas syringae*.

Считается, что адгезин FhaV, активируясь в результате расщепления, выполняет свои функции за счет ковалентного связывания с поверхностями-мишенями [51]. VIL-домены обнаружены также в некоторых генах других секреторных белков (*fhaB Mannheimia haemolytica*, *mafB Neisseria meningitidis* и т.п.) [51].

В целом активность VIL-доменов может являться предшественником РНК-сплайсинга [51]. В пользу этого свидетельствует тот факт, что в VIL-содержащих генах отсутствуют последовательности, сходные с интронами. Но функции VIL-доменов в клетке могут не ограничиваться только сплайсингом. Возможно, они обеспечивают высокую степень вариабельности поверхностных белков. Например, в некоторых поверхностных белках у *Neisseria meningitidis* (бактерии, вызывающие одну из форм менингита) и у *Ralstonia solanacearum* (патогенной бактерии растений) изменчивость С-концевого домена поверхностных вирулентных белков может определяться активностью VIL-домена [53, 54]. Расщепление по одному или двум сайтам может направленно изменять секретируемый или поверхностный белок. Кроме того, тот факт, что гомология некоторых мотивов VIL-доменов и Hh-белков высока, свидетельствует о возможности этерификации N-концевого экстеина (например, холестерина или другими нуклеофильными агентами) с последующим отщеплением его от исходного белка. Все

перечисленные модификации значительно увеличивают вариабельность секретируемых белков, что может повышать изменчивость антигенных детерминант и обеспечивать бактериям защиту от иммунной системы многоклеточных.

#### Автокаталитическая модификация ферментов

N–O-сдвиг инициирует процессинг у Hint-содержащих белков. Аналогичные реакции с последующим гидролизом пептидной связи протекают при созревании N-концевых нуклеофильных гидролаз, N-концевых аминотрансфераз и пируватзависимых ферментов [39, 55]. Интересно то, что их аутопроцессинг катализируется доменами, абсолютно отличными по последовательности и по трехмерной структуре от Hint-доменов. Поэтому мы рассмотрим эти превращения очень коротко.

Каталитическая активность ферментов первых двух семейств обеспечивается высокорезакционноактивным серином, треонином или цистеином на N-конце. Рассмотрим механизм аутопроцессинга на примере гликозиласпарагиназы из *Flavobacterium meningosepticum*, с которым сходен механизм созревания и гликозиласпарагиназы у других организмов, например, у человека. Зрелый фермент представляет собой гетеродимер, который кодируется одним геном. Неактивный белковый предшественник превращается в активную форму в результате автокаталитического отщепления пептидной связи перед Thr-152 с образованием отдельных  $\alpha$ - и  $\beta$ -субъединиц. Происходит это в результате N–O-сдвига с образованием эфирной связи, которая затем гидролизуется (рис. 8б) [39, 55]. В результате Thr-152 становится N-концевой аминокислотой  $\beta$ -субъединицы. Замена Thr-152 на аминокислоту другого типа (не Ser и Cys) полностью блокирует расщепление.

Аналогично инициируется созревание ряда других ферментов. Так, у бактерий активность декрбоксилаз и редуктаз ряда аминокислот обеспечивается N-концевой пирувоильной группой пирувоиловых ферментов, созревание которых протекает автокаталитически. Например, в белке-предшественнике гистидиндекрбоксилазы *Lactobacillus* между Ser-81 и Ser-82 происходит N–O-сдвиг, который в результате  $\beta$ -элиминации расщепляется на два пептида (рис. 8г). В результате реакции на N-конце фермента из Ser-82 образуется активный пируватный остаток [39, 55].

#### ВОЗМОЖНЫЕ ФУНКЦИИ ИНТЕИНОВ В КЛЕТКАХ

Значение интеинов в обеспечении жизнедеятельности клеток пока неясно. В начале функционирования интеинов пытались объяснить с позиции так называемых “эгоистичных генов” [45]. Ин-

теин за счет DOD-домена обеспечивает свое распространение по геному, а за счет Hint-домена – собственное вырезание с последующим лигированием экстеинов, что сохраняет биологическую активность белка-хозяина (т.е. не нарушает жизнедеятельность клетки-хозяина). Положение внутри белковой последовательности позволяет использовать механизм экспрессии в составе клеточных гена-носителя и его мРНК, что может свидетельствовать об “эгоизме” интеиновой последовательности.

Однако это заключение не всегда оправдано. Действительно, ряд признаков, и, в первую очередь, явление хоуминга, позволяет отнести интеины к мобильным элементам, таким как транспозоны и ретротранспозоны [32]. Но в отличие от последних, представленных в геноме множественными копиями, интеины встречаются в определенных местах определенных генов, и распространение их в геноме прямо коррелирует с количеством копий гена-носителя интеина.

Более вероятным кажется то, что свободный интеин (после сплайсинга) может выполнять в клетке пока неизвестные регуляторные и ферментативные функции. Такое предположение выдвинуто некоторыми исследователями [4]. Например, интеины могут регулировать активность ферментов-носителей. Созревание некоторых белков обеспечивается процессами *цис*- или *транс*-сплайсинга (например, DnaE у *Synechocystis* sp. PCC6803) [15, 32]. На активность фермента, безусловно, существенно влияет присутствие или отсутствие внутренней интеиновой последовательности [32]. Другими словами, интеин в данном случае может выступать как внутренний регулятор активности ферментов. Многие бактерии и многоклеточные организмы используют превращения, характерные для сплайсинга, для созревания или модификации своих белков (VIL-опосредованная модификация белков у бактерий; процессинг Hh-белков у многоклеточных). Кроме того, вырезанный интеин, аналогично активированным гликозиласпарагиназе или пирувоиловым ферментам, может катализировать определенные превращения в клетке за счет активированной N-концевой нуклеофильной группы [39]. Поэтому представление о исключительном “эгоизме” интеинов и аналогичных белков в настоящее время не представляется продуктивным. Скорее – это очередной способ регуляции и обеспечения жизнедеятельности клеток.

Автор выражает благодарность И.Я. Дубею за ценные предложения и замечания, высказанные при написании этого обзора.

#### СПИСОК ЛИТЕРАТУРЫ

1. Kane P.M., Yamashiro C.T., Wolczyk D.F., Neff N., Goebel M., Stevens T.H. 1990. Protein splicing converts

- the yeast TFP1 gene product to the 69 kDA subunit of the vacuolar H<sup>+</sup>-adenosine triphosphatase. *Science*. **250**, 651–657.
2. Anraku Y., Mizutani R., Satow Y. 2005. Protein splicing: its discovery and structural insight into novel chemical mechanisms. *IUBMB Life*. **57**, 563–574.
  3. Perler F., Davis E., Dean G., Gimble F., Jack W., Neff N., Noren C., Thorner J., Belfort M. 1994. Protein splicing elements: inteins and exteins a definition of terms and recommended nomenclature. *Nucleic Acids Res.* **22**, 1125–1127.
  4. Klabunde T., Sharma S., Telenti A., Jacobs W., Sacchetti J. 1998. Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Struct. Biol.* **5**, 31–36.
  5. Noren C., Wang J., Perler F. 2000. Dissecting the chemistry of protein splicing and its applications. *Angew. Chem. Int. Ed.* **39**, 450–466.
  6. Perler F. 2005. Protein splicing mechanisms and applications. *IUBMB Life*. **57**, 469–476.
  7. Yang J., Meng Q., Liu X. 2004. Inteин harbouring large tandem repeats in replicative DNA helicase of *Trichodesmium erythraeum*. *Mol. Microbiol.* **51**, 1185–1192.
  8. Pietrokovski S. 1994. Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci.* **3**, 2340–2350.
  9. Perler F. 2002. InBase, the inteин Database. *Nucleic Acids Res.* **30**, 383–384.
  10. Perler F., Olsen G., Adam E. 1997. Compilation and analysis of inteин sequences. *Nucleic Acids Res.* **25**, 1087–1093.
  11. Pietrokovski S. 1998. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci.* **7**, 64–71.
  12. Ding Y., Xu M., Ghosh I., Chen X., Ferrandon S., Lesage G., Rao Z. 2003. Crystal Structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J. Biol. Chem.* **278**, 3913–3914.
  13. Matsumura H., Takahashi H., Inoue T., Yamamoto T., Hashimoto H., Nishioka M., Fujiwara S., Takagi M., Imanaka T., Kai Y. 2006. Crystal structure of inteин homing endonuclease II encoded in DNA polymerase gene from hyperthermophilic archaeon *Thermococcus kodakaraensis* strain KOD1. *Proteins*. **63**, 711–715.
  14. Guhan N., Muniyappa K. 2003. Mycobacterium tuberculosis RecA inteин, a LAGLIDADG homing endonuclease, displays Mn<sup>2+</sup> and DNA-dependent ATPase activity. *Nucleic Acids Res.* **31**, 4184–4191.
  15. Ghosh I., Sun L., Xu M. 2001. Zinc inhibition of protein *trans*-splicing and identification of regions essential for splicing and association of a split inteин. *J. Biol. Chem.* **276**, 24051–24058.
  16. Choi J.J., Nam K.H., Min B., Kim S.J., Soll D., Kwon S.T. 2006. Protein *trans*-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*. *J. Mol. Biol.* **356**, 1093–1106.
  17. Clarke N. 1994. A proposed mechanism for the self-splicing of proteins. *Proc. Natl. Acad. Sci. USA*. **91**, 11084–11088.
  18. Shao Y., Xu M., Paulus H. 1996. Protein splicing: evidence for an N–O acyl rearrangement as the initial step in the splicing process. *Biochemistry*. **35**, 3810–3815.
  19. Mills K.V., Perler F.B. 2005. The mechanism of inteин-mediated protein splicing: variations on a theme. *Protein Peptide Lett.* **12**, 751–755.
  20. Xu M.Q., Coimb D.G., Paulus H., Noren J.C., Shao Y., Perler F.B. 1994. Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *EMBO. J.* **13**, 5517–5522.
  21. Sun P., Ye S., Ferrandon S., Evans T.C., Xu M.Q., Rao Z. 2005. Crystal structures of an inteин from the split *dnaE* gene of *Synechocystis* sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *J. Mol. Biol.* **353**, 1093–1105.
  22. Wang S., Liu X. 1997. Identification of an unusual inteин in chloroplast clpP protease of *Chlamydomonas eugametos*. *J. Biol. Chem.* **272**, 11869–11873.
  23. Perler F.B. 1999. A natural example of protein *trans*-splicing. *Trends Biochem. Sci.* **24**, 209–211.
  24. Khan M.S., Khalid A.M., Malik K.A. 2005. Inteин-mediated protein *trans*-splicing and transgene containment in plastids. *Trends Biotechnol.* **23**, 217–220.
  25. Flick K.E., Jurica M.S., Monnat R.J., Stoddard B.L. 1998. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*. **394**, 96–101.
  26. Gimble F.S., Duan X., Hu D., Quioco F.A. 1998. Identification of Lys-403 in the PI-SceI homing endonuclease as part of a symmetric catalytic center. *J. Biol. Chem.* **273**, 30524–30529.
  27. Dalgaard J.Z., Klar A.J., Moser M., Holley W.R., Chatterjee A., Mian I.S. 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an inteин that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res.* **25**, 4626–4638.
  28. Bakhrat A., Jurica M., Stoddard B., Raveh D. 2004. Homology modeling and mutational analysis of Ho endonuclease of yeast. *Genetics*. **166**, 721–728.
  29. He Z., Crist M., Yen H., Duan X., Quioco F.A., Gimble F.S. 1998. Amino acid residues in both the inteин splicing and endonuclease domains of the PI-SceI inteин mediate DNA binding. *J. Biol. Chem.* **273**, 4607–4615.
  30. Lykke-Andersen J., Garrett R.A., Kjems J. 1996. Protein footprinting approach to mapping DNA binding sites of two archaeal homing enzymes: evidence for a two-domain protein structure. *Nucleic Acids Res.* **24**, 3982–3989.
  31. Hu D., Crist M., Duan X., Quioco F.A., Gimble F.S. 2000. Probing the structure of the PI-SceI-DNA complex by affinity cleavage and affinity photocross-linking. *J. Biol. Chem.* **275**, 2705–2712.
  32. Pietrokovski S. 2001. Inteин spread and extinction in evolution. *Trends Genet.* **17**, 465–472.
  33. Adam E., Perler F. 2002. Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in *Escherichia coli* DNA gyrase subunit A. *J. Mol. Microbiol. Biotechnol.* **4**, 479–487.
  34. Hall T.M., Porter J.A., Young K.E., Koonin E.V., Beachy P.A., Leahy D.J. 1997. Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell*. **91**, 85–97.



35. Porter J.A., Ekker S.C., Park W.J., von Kessler D.P., Young K.E., Chen C.H., Ma Y., Wood A.S., Cotter R.J., Koonin E.V., Beachy P.A. 1996. Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell*. **86**, 21–34.
36. Komori K., Fujita N., Ichianagi K., Shinagawa H., Morikawa K., Ishino Y. 1999. PI-PfuI and PI-PfuII, intein-coded homing endonucleases from *Pyrococcus furiosus*. *Nucleic Acids Res.* **27**, 4167–4182.
37. Butler M., Goodwin T., Poulter R. 2005. Two new fungal inteins. *Yeast*. **22**, 493–501.
38. Butler M.I., Gray J., Goodwin T.J., Poulter R.T. 2006. The distribution and evolutionary history of the PRP8 intein. *BMC. Evol. Biol.* **6**, 42.
39. Bjornsdottir S.H., Blondal T., Hreggvidsson G.O., Eggertsson G., Petursdottir S., Hjorleifsdottir S., Thorbjarnardottir S.H., Kristjansson J.K. 2006. *Rhodothermus marinus*: physiology and molecular biology. *Extremophiles*. **10**, 1–16.
40. Khan M.S., Khalid A.M., Malik K.A. 2005. Intein-mediated protein trans-splicing and transgene containment in plastids. *Trends Biotechnol.* **23**, 217–220.
41. Hertveldt K., Lavigne R., Pleteneva E., Sernova N., Kurochkina L., Korchevskii R., Robben J., Mesyanzhinov V., Krylov V.N., Volckaert G. 2005. Genome comparison of *Pseudomonas aeruginosa* large phages. *J. Mol. Biol.* **354**, 536–545.
42. Nagasaki K., Shirai Y., Tomaru Y., Nishida K., Petrokovski S. 2005. Algal viruses with distinct intraspecies host specificities include identical intein elements. *Appl. Environ. Microbiol.* **71**, 3599–3607.
43. Suhre K., Audic S., Claverie J.M. 2005. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. *Proc. Natl. Acad. Sci. USA*. **102**, 14689–14693.
44. Ogata H., Raoult D., Claverie J.M. 2005. A new example of viral intein in Mimivirus. *Viol. J.* **2**, 1–7.
45. Gogarten J., Senejani A., Zhaxybayeva O., Olendzenski L., Hilario E. 2002. Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.* **56**, 263–287.
46. Liu X., Yang J., Meng Q. 2003. Four Inteins and three group II Introns encoded in a Bacterial ribonucleotide reductase gene. *J. Biol. Chem.* **278**, 46826–46831.
47. Paulus H. 2001. Inteins as Enzymes. *Bioorgan. Chem.* **29**, 119–129.
48. Paulus H. 2000. Protein splicing and related forms of protein autoprocessing. *Ann. Rev. Biochem.* **69**, 447–496.
49. Ingham P.W., McMahon A.P. 2001. Hedgehog signaling in animal development: paradigms and principles. *Genes Devel.* **15**, 3059–3087.
50. Roessler E., Belloni E., Gaudenz K., Vargas F., Scherer S.W., Tsui L.C., Muenke M. 1997. Mutations in the C-terminal domain of sonic hedgehog cause holoprosencephaly. *Hum. Mol. Genet.* **6**, 1847–1853.
51. Amitai G., Belenkiy O., Dassa B., Shainskaya A., Petrokovski S. 2003. Distribution and function of new bacterial intein-like protein domains. *Molec. Microbiol.* **47**, 61–73.
52. Dassa B., Haviv H., Amitai G., Petrokovski S., 2004. Protein splicing and auto-cleavage of bacterial intein-like domains lacking a C'-flanking nucleophilic residue. *J. Biol. Chem.* **30**, 32001–32007.
53. Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature*. **404**, 502–506.
54. Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlat M. 2001. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*. **415**, 497–502.
55. Albert A., Dhanaraj V., Genschel U., Khan G. Ramjee M.K., Pulido R., Sibanda B.L., von Delft F., Witty M., Blundell T.L., Smith A.G., Abell C. 1998. Crystal structure of aspartate decarboxylase at 2.2 Å resolution provides evidence for an ester in protein self-processing. *Nature Struct. Biol.* **5**, 289–293.