

УДК 577.21

## КОД ВНУТРИ КОДА: КАК КОДОНЫ ОСУЩЕСТВЛЯЮТ ТОНКУЮ НАСТРОЙКУ СВОРАЧИВАНИЯ БЕЛКА В КЛЕТКЕ

### Обзор

© 2021 А.А. Комар<sup>1,2,3,4</sup>

<sup>1</sup> Center for Gene Regulation in Health and Disease and Department of Biological, Geological and Environmental Sciences, Cleveland State University, Cleveland, Ohio 44115, USA; e-mail: a.komar@csuohio.edu

<sup>2</sup> Department of Biochemistry and Center for RNA Science and Therapeutics, Case Western Reserve University, Cleveland, Ohio 44106, USA

<sup>3</sup> Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA

<sup>4</sup> DAPCEL, Inc., Cleveland, Ohio 44106, USA

Поступила в редакцию 11.04.2021

После доработки 11.04.2021

Принята к публикации 28.04.2021

Генетический код устанавливает соответствие между последовательностью определенного триплета нуклеотидов в молекуле мРНК, называемого кодоном, и аминокислотой, добавляемой к растущей полипептидной цепи в процессе синтеза белка. Комбинации четырех азотистых оснований (А, G, U и C) приводят к образованию 64 триплетных кодонов, из которых 61 являются смысловыми (кодирующими аминокислоты), а 3 – стоп-кодонами (определяющими терминацию трансляции). У большинства организмов для синтеза белка используется 20 стандартных аминокислот. Таким образом, генетический код является вырожденным, так как большинство аминокислот (за исключением метионина и триптофана) кодируется более, чем одним (синонимическим) кодоном. Изначально предполагалось, что синонимические кодоны обладают полностью эквивалентными функциями, однако обнаружение того факта, что синонимические кодоны используются в мРНК с неодинаковой частотой, позволило предположить, что выбор положения определенных кодонов в мРНК может быть не случаен и может нести дополнительную функциональную нагрузку помимо кодирования аминокислотных остатков. Установление факта неэквивалентного использования кодонов в мРНК поставило вопрос о возможности наличия в генетическом коде дополнительной информации. Действительно, было обнаружено, что генетический код имеет «несколько слоёв», несущих такую дополнительную информацию, и что синонимические кодоны расположены в мРНК не случайно, а стратегически – чтобы гарантировать определенную кинетику трансляции, которая, в свою очередь, обеспечивает тонкую настройку процесса сворачивания белка в клетке и способствует упорядоченному котрансляционному сворачиванию, которое происходит в ходе поэтапного и последовательного структурирования отдельных участков полипептидной цепи, выходящих из рибосомы в различные моменты времени. В настоящем обзоре приведены основные данные, полученные в этой области, которые способствовали установлению роли синонимических кодонов (и их использования) в процессе сворачивания белков в клетке.

**КЛЮЧЕВЫЕ СЛОВА:** генетический код, использование/предпочтение кодонов, синонимические кодоны, кинетика трансляции, насцентные пептиды, котрансляционное сворачивание белка.

**DOI:** 10.31857/S0320972521080091

### ВВЕДЕНИЕ

Шестьдесят лет назад Crick, Barnett, Brenner и Watts-Tobin постулировали основные свойства генетического кода [1]. Они предположили, что группа из трех азотистых оснований (кодон) ко-

дирует одну аминокислоту, что код является неперекрывающимся, что последовательность оснований считывается с фиксированной начальной точки, что код не содержит каких-либо особых «запятых», которые могли бы указывать на выбор правильных триплетов, и что код является

Принятые сокращения: мРНК – информационная/матричная РНК; тРНК – транспортная РНК, РНКазы А – рибонуклеаза А; ЯМР – ядерный магнитный резонанс (nuclear magnetic resonance); ADAMTS13 – 13-й член семейства дисинтегринов и металлопротеиназ с тромбоспондиновым мотивом 1-го типа (a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13); CAT – хлорамфениколацетилтрансфераза (chloramphenicol acetyltransferase); ExAC – консорциум агрегации экзонов (the Exome Aggregation Consortium); FRET – резонансный перенос энергии флуоресценции (fluorescence resonance energy transfer); *MDR1* – ген множественной устойчивости к лекарствам 1 (multidrug resistance 1); PGK – фосфоглицераткиназа (phosphoglycerate kinase); P-gp – P-гликопротеин (P-glycoprotein); VWF – фактор фон Виллебранда (von Willebrand factor).

вырожденным, т.е., как правило, одна конкретная аминокислота может кодироваться несколькими (синонимическими) кодонами [1]. Последующие работы по изучению связывания аминоксил-тРНК с короткими тринуклеотидными матрицами позволили установить общие закономерности кодирования и выявить точную природу вырожденности генетического кода [2–8].

В дальнейшем было установлено, что синонимические кодоны, кодирующие одну и ту же аминокислоту, могут быть не полностью эквивалентными [9–19]. Было показано, что декодирование синонимических кодонов происходит с различными скоростями и эффективностью (и точностью), и что частота использования синонимических кодонов во многом определяет скорость их декодирования [20–22]. В этой связи следует отметить, что скорость и эффективность считывания кодона в значительной степени определяется доступностью соответствующих/родственных (cognate) тРНК, в то время как конкуренция с другими/неродственными тРНК оказывает влияние на точность трансляции [23–27]. Так как было обнаружено, что частоты использования кодонов обычно коррелируют с доступностью и концентрацией соответствующих тРНК [12, 13, 16], то было сделано предположение о том, что выбор кодона влияет на скорость элонгации [20–22] (действительно, как оказалось, концентрация тРНК является одним из основных факторов, определяющих время декодирования кодонов [28, 29]). Во многих последующих экспериментах было показано, что часто используемые кодоны подвергаются более быстрой трансляции по сравнению с теми, что используются реже. Это обусловлено большей доступностью (во время процесса считывания мРНК) соответствующих часто встречающихся родственных тРНК [20–22, 30, 31]. Кроме того, было обнаружено, что гены с высоким уровнем экспрессии содержат в своей последовательности более предпочтительные/часто используемые кодоны, в то время как низкоэкспрессируемые гены оказались обогащены редко используемыми синонимическими кодонами [17, 18]. И несмотря на то что набор оптимальных/часто используемых кодонов различается у разных биологических видов (эти изменения тем не менее оказались скоординированными с изменениями состава генов тРНК [13, 16]), был сделан вывод о том, что мРНК не только кодирует аминокислотную последовательность белка, но и определяет уровень его экспрессии [32–36]. Как следствие, была выдвинута теория совместной эволюции кодонов и тРНК, которая объясняла необходимость существования определенного баланса между точностью и эффек-

тивностью синтеза белка [32–36]. В рамках этой теории также было сделано предположение о том, что естественный отбор благоприятствует выбору оптимальных и часто используемых кодонов (в противовес редким) для обеспечения эффективной продукции белка [32–36]. Однако позднее стало понятно, что редко используемые кодоны также подвержены направленному отбору и обладают важной функцией, особенно в гомеостазе мРНК и регуляции её стабильности внутри клетки [37–44], а также при сворачивании белка в клетке (см. ниже).

Параллельно с попытками определить природу генетического кода и процесса декодирования было установлено, что аминокислотная последовательность белка содержит код сворачивания белка, т.е. информацию, необходимую для определения его уникальной трехмерной структуры [45]. Хотя точная природа кода сворачивания белка не была установлена (и остается не полностью разгаданной до настоящего времени), тем не менее был сделан вывод о том, что нативное свернутое состояние белка достигается не путем случайного перебора всех возможных конформаций на пути к нативной структуре (что потребовало бы огромного количества времени для его осуществления), но, скорее всего, достигается через ограниченный набор четко определенных путей с ограниченным числом доступных конформационных пространств (см. обзоры [46–51]).

Примерно в то же время (в начале 60-х и в 70-е гг. прошлого столетия) было также обнаружено, что сворачивание белка *in vivo*, по крайней мере, в случае некоторых белков, является котрансляционным процессом, (т.е. протекает во время синтеза белка на рибосоме) [52–58]. Что в то время было неясно, так это то, могут ли рибосома и/или процесс трансляции оказывать влияние на процесс сворачивания белка, и до какой степени конкретный путь сворачивания белка в клетке (если он существует) может обеспечивать высокую эффективность процесса сворачивания белка *in vivo*. Ответы на эти вопросы стали медленно возникать в течение последних 50 лет и продолжают появляться в наши дни.

Декодирование или трансляция молекул мРНК осуществляется рибосомами с добавлением каждой новой аминокислоты к растущей полипептидной цепи. Можно предположить, что природа процесса декодирования, или сама рибосома (как большой макромолекулярный комплекс), и/или скорость полимеризации полипептидной цепи могут влиять на конечную конформацию белка, воздействуя на конформационное пространство растущей полипептидной цепи и/или на кинетику процесса сворачи-

вания белка в процессе синтеза полипептидной цепи, или на то и другое. Необходимо отметить, что это предположение нашло свое подтверждение в ходе многочисленных недавних экспериментов, в которых было показано, что механизм декодирования, а также сама рибосома и, что важно, кинетика трансляции, которая определяет скорость полимеризации полипептидной цепи, могут оказывать влияние на процесс сворачивания белка и определять его эффективность *in vivo* (см. ниже). Однако таких данных не существовало до 1985 г., когда стереохимический анализ осуществляемой рибосомой реакции транспептидации, выполненный Лимом и Спириным, позволил предположить, что процесс декодирования может влиять на сворачивание белка, и что рибосома может способствовать формированию альфа-спирали в растущей полипептидной цепи [59, 60]. Таким образом, полученные данные позволили утверждать, что трансляция — это не просто процесс последовательного добавления аминокислот к растущему полипептиду, но и процесс, который может активно влиять на сворачивание белка.

Впоследствии и мы, и другие авторы показали, что различия в скоростях трансляции в различных участках мРНК, которые в основном обусловлены неравномерным и неслучайным использованием синонимических кодонов, могут влиять на сворачивание белка, обуславливая упорядоченное и последовательное образование дискретных структур во вновь синтезированных участках насцентной полипептидной цепи, и что кинетика синтеза белка может служить в качестве вторичного кода для тонкой настройки процесса сворачивания белка в клетке (см. обзор [61]). Более того, анализ распределения кодонов в молекуле мРНК позволил нам и другим авторам найти «отпечатки» структуры белка, закодированные в последовательности мРНК [61]. Стало ясно, что естественный отбор не только направлен на поддержание эффективной экспрессии белка, но и также благоприятствует определенному и неслучайному распределению кодонов в матричной РНК, и что мутационное давление и генетический дрейф способствуют расположению редко используемых кодонов в определенных местах, опосредуя, таким образом, выполнение ими важных функций, в частности в процессе сворачивания белка [61].

Настоящий обзор посвящен рассмотрению ряда ключевых результатов, свидетельствующих в пользу существования внутри генетического кода вторичного кода, который способствует тонкой регуляции сворачивания белка в клетке, а также рассмотрению роли редких кодонов в этом процессе.

## КОД СВОРАЧИВАНИЯ БЕЛКА

Основополагающие эксперименты по обратимой денатурации рибонуклеазы, выполненные Christian Anfinsen и его коллегами в 1950–1960-х гг. [45], а также последующие исследования других групп с использованием большого числа других белков показали, что аминокислотная последовательность белка содержит всю необходимую информацию для образования его уникальной трехмерной структуры [46–51]. Многие последующие теоретические исследования, а также работы по компьютерному моделированию сворачивания белка, включая недавние прорывные исследования в этой области, показавшие возможность точного предсказания структуры белка на основе его аминокислотной последовательности [62, 63], свидетельствуют о том, что ключевой постулат «проблемы сворачивания белка» является верным. В связи с этим необходимо отметить недавние исследования по *in silico* моделированию и предсказанию белковых структур, которые в ряде случаев оказались практически неотличимыми от структур, определенных с помощью рентгеновской кристаллографии [64], что является значительным достижением в области и важным доказательством постулата (догмы) Анфинсена.

Для объяснения процесса сворачивания белка была разработана концепция «энергетической воронки», предполагающая существование множества различных путей, которые приводят процесс сворачивания белка к нативной конформации, характеризующейся минимумом свободной энергии (см. обзоры [46–51]). В дальнейшем эта концепция была дополнена появлением новой модели, описывающей процесс сворачивания белка через образование отдельных кооперативных единиц сворачивания, так называемых фолдонов (см. обзоры [65–67]). Действительно, для нескольких белков было показано, что их сворачивание происходит через такие последовательные промежуточные состояния, а также высказано предположение о том, что кооперативная по своему существу природа фолдонов определяет быстрое формирование нативной структуры [65–67]. Гипотеза фолдонов также смогла объяснить предпочтительный выбор определенных энергетических состояний в образующихся кооперативных структурах, направленный в сторону образования нативных взаимодействий. Ограниченное количество и небольшой размер таких фолдонов (~15–35 аминокислотных остатков) обеспечивали также решение парадокса Левинтала [68].

Для многих, в особенности небольших, белков теоретические исследования, эксперименты

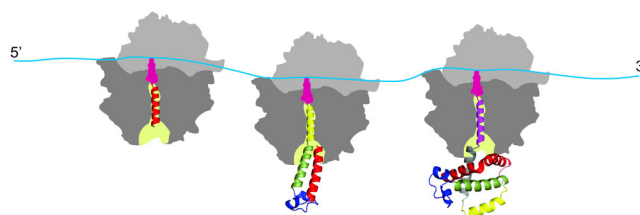
по компьютерному моделированию белковой структуры, а также эксперименты *in vitro* по разворачиванию и сворачиванию полипептидной цепи пришли в полное согласие друг с другом (работа [69] и указанные в ней ссылки). Эти исследования дали общие ответы на ряд ключевых вопросов, а именно: как сворачиваются белки и почему они сворачиваются таким образом?

Тем не менее во многих случаях (и особенно в случае больших белков) эти эксперименты не позволяли описать весь процесс целиком [70]. Кроме того, многие попытки добиться сворачивания (рефолдинга) в пробирке очищенных денатурированных белков имели лишь частичный успех. К тому же восстановление трехмерной структуры в пробирке для некоторых белков оказалось чрезвычайно медленным процессом [46], занимавшим часы или даже дни [71, 72]. В то же время было хорошо известно, что эта «проблема несоответствия» скорости сворачивания белка в пробирке и в клетке легко решается *in vivo*. Таким образом, стало понятно, что котрансляционная природа сворачивания белка в клетке может служить ключом к пониманию природного кода сворачивания белков. Как результат, исследователи в области теории сворачивания белка, компьютерного моделирования процесса сворачивания, а также исследователи, непосредственно вовлеченные в лабораторные эксперименты по изучению фолдинга, переключили свое внимание на изучение процесса фолдинга *in vivo*, для того чтобы лучше понять сходства и различия путей и механизмов сворачивания белка *in vitro* и *in vivo* (см. обзоры [73–78]).

В этой связи следует отметить, что котрансляционное сворачивание белка представляет собой процесс, который существенно отличается от процесса сворачивания (рефолдинга) белка *in vitro* [79–84]. Это процесс, который характеризуется тем, что nascentная полипептидная цепь в процессе её синтеза остается прикрепленной (через её C-концевой участок) к рибосоме, что приводит к ограничению её конформационного пространства и уменьшению числа возможных степеней свободы, тем самым существенно ограничивая число возможных интермедиатов фолдинга и, следовательно, снижая число возможных путей сворачивания белка [79–84]. Также необходимо отметить, что рибосома служит в качестве хаба [85] для многих молекулярных шаперонов. Кроме того, было показано, что самой рибосоме присуща «шаперон-подобная» активность ([86–88] и указанные в них ссылки). Молекулярные шапероны и катализаторы фолдинга могут взаимодействовать с растущей nascentной полипептидной цепью, выходящей из рибосомного канала, таким образом ускоряя медленные

этапы сворачивания белка и предотвращая неправильное сворачивание [73–78]. Тем не менее предполагается, что шапероны и катализаторы фолдинга в основном принимают участие в кинетическом разделении между правильным сворачиванием и агрегацией белка. Таким образом, считается, что они в основном влияют на выход правильно свернутого белка, а не на механизм его сворачивания [73–78]. Необходимо также отметить, что котрансляционное сворачивание белка начинается на самых ранних этапах синтеза полипептидной цепи на рибосоме, внутри рибосомного канала, с образования отдельных элементов вторичной структуры (таких как, например, альфа-спирали) [89–92] в верхних отделах канала, а также с образования ряда небольших элементов третичной структуры – в «вестибюле» (нижнем/расширенном участке) рибосомного канала [93–96] (рис. 1). Важно отметить, что котрансляционное сворачивание является «векторным» (направленным) процессом, происходящим от N- к C-концу растущего полипептида и, как правило, заключается в последовательном структурировании отдельных участков полипептидной цепи, выходящих из рибосомного канала [61, 79–84].

Несмотря на выявление ряда исключений из этого общего правила, когда было показано, что сворачивание определенного небольшого домена белка может протекать в направлении, противоположном по отношению к синтезу полипептидной цепи на рибосоме ([97] и указанные в ней ссылки), все же был сделан вывод о том, что для большинства белков синтез и сворачивание происходят параллельно и однонаправленно



**Рис. 1.** Котрансляционное сворачивание белка представляет собой поэтапный векторный процесс, протекающий в большинстве случаев от N- к C-концу растущей полипептидной цепи. Появление отдельных промежуточных интермедиатов сворачивания связано с локальными и общими скоростями трансляции. Процесс сворачивания начинается на самых ранних этапах синтеза полипептидной цепи на рибосоме с образованием некоторых элементов вторичной структуры (таких, как альфа-спирали) внутри рибосомного канала (в его верхних участках), а также некоторых элементов третичной структуры, формирующихся внутри «вестибюля» (нижнего/расширенного участка канала) на выходе из него. Третичная структура белка оказывается практически полностью сформированной к концу его синтеза

[61, 79]. Наконец, котрансляционное сворачивание белков тесным образом связано с элонгацией трансляции, которая не является однородным/непрерывным процессом ([61, 79, 98] и указанные в них ссылки), происходящим с постоянной/одной скоростью. Было высказано предположение о том, что локальные скорости биосинтеза белка эволюционно адаптированы таким образом, чтобы оптимизировать процесс фолдинга, позволяя обходить глубокие кинетические ловушки в процессе котрансляционного сворачивания ([61, 79, 98–107] и указанные в них ссылки). Таким образом, было постулировано, что самосборка белка *in vivo* скоординирована с синтезом белка и эволюционно оптимизирована для быстрого высокоэффективного достижения правильно свернутой конформации полипептидной цепи ([61, 79, 98–107] и ссылки в них). Мы и другие авторы предположили (см. ниже), что редкие кодоны (которые транслируются замедленно) могут находиться под давлением эволюционного отбора для обеспечения эффективного котрансляционного сворачивания белка, а места локализации трансляционных пауз могут определять то временное окно, в котором происходит локальное сворачивание фрагментов белковой структуры, особенно в тех участках, где процесс далек от равновесия ([61, 79, 98–107] и ссылки в них). Тем не менее следует отметить, что многие детали процесса котрансляционного сворачивания остаются лишь частично понятыми, а его механизм и путь остаются предметом интенсивных исследований в наши дни.

### КОТРАНСЛЯЦИОННОЕ СВОРАЧИВАНИЕ БЕЛКА: ПРОГРЕСС В ПОНИМАНИИ ЭТОГО ПРОЦЕССА

Прогресс, достигнутый в понимании котрансляционного сворачивания белков неразрывно связан с разработкой новых методов и подходов для его изучения (см. обзор [82]).

Работы по изучению котрансляционного сворачивания белков можно разделить на три основных периода. В начале 1960-х и в 1970-е гг. были получены первые данные, которые позволили предположить, что сворачивание полипептидов *in vivo*, по крайней мере, в случае некоторых белков, является котрансляционным процессом [52–58]. Это в основном были эксперименты, проводившиеся в равновесных условиях, для которых были сформулированы и определены основные требования к методам, используемым для изучения котрансляционного сворачивания белков, такие как:

1) существование легко измеряемых свойств (присущих свернутому белку), которые позволяли бы проводить оценку правильного сворачивания насцентных полипептидных цепей на рибосоме;

2) наличие уверенности в том, что конкретные исследуемые структурные особенности действительно связаны с насцентными цепями, а не с полипептидными цепями, связанными с рибосомами/полирибосомами неспецифически;

3) наличие уверенности в том, что изучаемые полипептидные цепи были синтезированы *de novo* для гарантии того, что измеряемые специфические свойства белка (отражающие его фолдинг) действительно являются результатом котрансляционного процесса.

Большая часть этих ранних экспериментов заключалась в выделении/фракционировании (из клеточных экстрактов) комплексов, состоящих из насцентных цепей, связанных с рибосомами, при помощи градиента плотности сахарозы. Затем проводили оценку структурных свойств насцентных полипептидных цепей посредством определения ряда их параметров: 1) специфической ферментативной активности [52–54], 2) распознавания определенных структур специфическими/конформационными антителами [55] или 3) определения образования правильных дисульфидных связей внутри и/или между насцентными цепями [56–58]. В этих первых экспериментах было принципиально показано существование котрансляционного сворачивания белков, однако понимание механизмов и масштабов этого явления было в то время весьма ограничено.

Интерес к изучению котрансляционного сворачивания затем в значительной мере ослаб, в связи с доминированием экспериментов по денатурации/ренатурации белков *in vitro*. Возобновление интереса произошло в конце 1980-х и в середине 1990-х гг. с осознанием того, что котрансляционное сворачивание, вероятно, характерно не только для нескольких отдельных случаев, но может быть присуще многим белкам в клетке (см. обзор [82]). Исследованиям в этой области способствовало массовое применение бесклеточных систем трансляции *in vitro*, таких как, например, лизат из ретикулоцитов кролика (RRL – rabbit reticulocyte lysate), который был разработан Pelham и Jackson в 1976 г. [108]; использование новых модельных белков, в частности, хемилюминесцентных и флуоресцентных; разработка новых кинетических подходов, которые проложили дорогу к изучению котрансляционного сворачивания белков в реальном времени.

В этом плане работа Колба, Макеева и Спирина представляет особый интерес, т.к. является

одной из первых попыток изучения котрансляционного сворачивания белка в реальном времени. Эти авторы разработали методику, позволяющую наблюдать ферментативную активность вновь синтезированной люциферазы светлячка в бесклеточной системе в люминометрической кювете в реальном времени [109]. Люциферазная активность, указывающая на правильное сворачивание белка, регистрировалась при образовании полноцепочечной молекулы белка в реакции трансляции [109]. Следует отметить, что добавление РНКазы А нарушало и трансляцию, и накопление активной люциферазы, указывая на то, что синтез и сворачивание люциферазы светлячка происходят одновременно. Несмотря на то что авторы не смогли обнаружить люциферазную активность в полипептидных цепях, связанных с рибосомой (было показано, что 12 C-концевых аминокислотных остатков имеют важное значение для активности фермента ([109] и ссылки в ней), тем не менее они обнаружили, что люцифераза светлячка становится активной сразу после высвобождения насцентных цепей [109]. Важно отметить, что столь быстрое проявление ферментативной активности было несовместимо со сценарием посттрансляционного сворачивания, поскольку рефолдинг фермента из денатурированного состояния проходил намного медленнее [109].

В то же время наша группа на кафедре молекулярной биологии биологического факультета МГУ им. М.В. Ломоносова совместно с лабораторией А.С. Спирина в Институте белка в Пушкино использовала метод связывания гема с глобиновыми цепями для изучения котрансляционного сворачивания цепей альфа-глобина [110, 111]. Способность кофакторов и лигандов, таких как гем, специфически связываться с растущими полипептидными цепями, может рассматриваться как показатель того, что конформация, необходимая для их связывания, уже появляется на рибосоме [82, 110, 111]. Изучая реакции трансляции *in vitro*, в присутствии [<sup>3</sup>H]-гемина и [<sup>14</sup>C]-лейцина или [<sup>35</sup>S]-метионина с помощью центрифугирования в градиенте плотности сахарозы, в условиях наличия или отсутствия пурамицина, мы показали, что связанные с рибосомой цепи глобина способны эффективно связывать гем [110, 111]. Кроме того, мы обнаружили, что неполные насцентные цепи глобина, прикрепленные к рибосоме, способны котрансляционно связывать гем, что указывало на то, что структура, позволяющая насцентной полипептидной цепи связывать гем, возникает еще до завершения синтеза альфа-глобина [111]. Множество других исследований, выполненных в тот период времени, предоста-

вило неоспоримые доказательства в пользу котрансляционного сворачивания белков (см. обзоры [61, 82]).

Начало третьего периода исследований котрансляционного сворачивания можно отнести примерно к началу 2000-х годов. Этот период продолжается и до сих пор и характеризуется взрывным ростом большого количества новых методов и подходов для изучения котрансляционного сворачивания белков, таких как одномолекулярная флуоресцентная спектроскопия и флуоресценция с временным разрешением [96], позволяющих изучать процесс сворачивания белка в реальном времени [96] и на уровне единичных молекул; ЯМР-спектроскопия и криоэлектронная микроскопия, которые, в свою очередь, позволяют определить структуру связанных с рибосомой насцентных цепей с атомарным разрешением. Разработка множества других подходов, сочетающих эксперименты *in vivo* и *in vitro* (бесклеточная трансляция), и существенный прогресс в разработке компьютерных подходов для моделирования фолдинга также приближают нас к более глубокому пониманию процесса котрансляционного сворачивания белков (см. обзор [82]).

Стало очевидно, что котрансляционное сворачивание характерно почти для всех белков как прокариотических, так и эукариотических клеток, как однодоменных, так и мультидоменных, а также состоящих из одной или нескольких субъединиц, цитоплазматических, секретиремых или мембраносвязанных, и что без понимания механизма и пути котрансляционного сворачивания белков всестороннее понимание кода сворачивания белка будет невозможным.

### КОД ИСПОЛЬЗОВАНИЯ КОДОНОВ ДЛЯ СВОРАЧИВАНИЯ БЕЛКА В КЛЕТКЕ: ЕГО ОСНОВНЫЕ ЧЕРТЫ

Изучение процесса котрансляционного сворачивания белка и особенно эксперименты, показывающие, что рефолдинг белков из денатурированного состояния происходит намного медленнее по сравнению с их котрансляционным сворачиванием (даже в условиях, когда такое сравнение проводится в одинаковых условиях, например, таких как клеточные экстракты, используемые для трансляции *in vitro*) [109], привели к предположению о том, что векторная природа котрансляционного процесса, а также дополнительная информация, которую несет молекула мРНК (помимо непосредственно кодирования аминокислотной последовательности), могут объяснить высокую эффективность



котрансляционного фолдинга *in vivo* [112–115]. Тем не менее стало ясно, что направленность котрансляционного сворачивания скорее всего играет роль лишь тонкой настройки (при этом очень важной, как будет видно из обсуждения ниже) в общем процессе фолдинга, поскольку некоторые белки, например, циркулярно перемутированные белки или белки, полученные в ходе химического синтеза по Меррифилду, оказываются свернутыми правильным образом и после изменения направления синтеза [46, 116].

Работы Лима и Спирина [59, 60], о которых упоминалось выше, позволили предположить, что считывание мРНК на рибосоме может облегчить формирование альфа-спиральной конформации в растущей насцентной полипептидной цепи. В связи с этим предположением возникло сразу несколько вопросов:

1) если рибосомы способствуют образованию альфа-спиральной конформации — то как тогда образуются другие вторичные структуры (другие типы спиралей, повороты, бета-структуры и др.)?

2) альфа-спирали, как известно, сворачиваются быстро [46]; может ли тогда формирование быстро сворачивающихся элементов структуры нуждаться в более высоких скоростях трансляции, а формирование медленно сворачивающихся элементов — требовать более медленных скоростей?

3) может ли существовать и быть необходимым определенным период времени для превращения альфа-спирали в другую (другие) вторичную структуру в процессе синтеза белка на рибосоме?

4) могут ли сайты трансляционных пауз разделять образование различных структурных элементов (таких как вторичные структуры одного или другого типа) во время синтеза белка на рибосоме?

5) может ли потребоваться более длительная задержка трансляции (трансляционная пауза) для формирования больших единиц сворачивания (например, доменов в мультидоменных белках)?

6) наконец, как изменения кинетики трансляции могут отразиться на сворачивании белков в клетке?

Примечательно, что возможность образования компактной (альфа-спиральной) структуры внутри рибосомного канала была экспериментально продемонстрирована через 30 лет после исходного теоретического предположения, выдвинутого Лимом и Спириным. Для этого были применены несколько подходов, включая FRET (перенос энергии флуоресцентного резонанса) [89]; определение доступности (к пегилиро-

ванию) остатков цистеинов, расположенных в разных участках насцентной полипептидной цепи (постулировалось, что длина участка растущего насцентного пептида, расположенного внутри рибосомного канала (покрываемого каналом) изменяется (увеличивается) в зависимости от формирования им альфа-спиральной структуры и, таким образом, изменяет момент доступности определенного цистеина для пегилирования вне канала) [90]; а также с помощью прямого определения структуры пептида в канале (см. обзор [91]).

Следует также отметить, что проведенные в середине 1970-х и в 1980-х гг. исследования показали, что элонгация трансляции — это неравномерный процесс [117–125]. Несмотря на выявление нескольких причин неравномерности трансляции (в частности, таких как структура мРНК [118], которая может препятствовать движению рибосомы), стало ясно, что неравномерное распределение синонимических кодонов является одним из ключевых факторов, контролирующих скорость элонгации трансляции [12–22, 124]. Как было отмечено ранее, было обнаружено, что частые/оптимальные кодоны, как правило, транслировались быстрее по сравнению с редкими [12–22].

Используя эти знания как отправную точку наших исследований, мы поставили перед собой задачу проанализировать выбор кодонов в контексте их распределения вдоль последовательности мРНК и изучить взаимосвязи между участками локализации определенных кодонов и кодируемыми ими структурами в белках [113–115]. В частности, нам было интересно выяснить, могут ли редкие кодоны обнаруживаться в пограничных участках между различными элементами вторичной структуры или на границах между доменами. Мы постулировали, что последовательные события процесса котрансляционного сворачивания (такие как формирование определенных вторичных и/или третичных структур, например, доменов) могут быть разделены редкими кодонами и сайтами трансляционных пауз [113–115]. Независимо от нас группа Alistair Brown из Института генетики университета Глазго (Великобритания) высказала предположение, что сворачивание доменов белков на рибосоме может быть разделено трансляционными паузами, и что такие участки замедленной трансляции могут служить в качестве «знаков препинания» в процессе котрансляционного фолдинга белков [112].

Следует признать, что эти ранние исследования не были достаточно полными (частично ввиду отсутствия достаточной информации о структурах и последовательностях белков). Тем

не менее мы пришли к заключению, что распределение редких кодонов (и их кластеров) вдоль молекул мРНК носит эволюционно консервативный характер, и что эти кодоны действительно могут разделять формирование определенных элементов вторичной структуры, таких как, например, альфа-спирали (анализ был проведен на нескольких однодоменных белках, таких как цитохромы *c*, миоглобины и глобины), а также элементов третичной структуры, таких как, например, домены (как было показано на примере мультидоменных белков, таких как кристаллины гамма-В и бета-В2, фосфоглицераткиназы (PGKs – phosphoglycerate kinases) и др.) [113–115, 126]. Многие последующие углубленные исследования [127–136] выявили сходные тенденции как на границах перехода между различными элементами вторичной структуры (например, бета-тяж → спираль; спираль → бета-тяж) [134], так и между доменами [127, 131–133]. Также стало очевидно, что в распределении кластеров редких кодонов вдоль мРНК существует определенная иерархия [61, 79, 98, 113–115, 127, 128]. Как оказалось, наиболее редкие кодоны (и их кластеры) часто кодируют границы между относительно большими структурными единицами (например, доменами), в то время как менее редкие кодоны – границы между более мелкими единицами структуры (например, мотивами, субдоменами). Было высказано предположение, что это различие отражает потребность в обеспечении более значительной задержки трансляции для независимого сворачивания больших единиц структуры по сравнению с небольшими структурными элементами [61, 79, 98, 113–115]. Мы также предположили, что ускорение трансляции в определенных участках мРНК является не менее важным фактором, поскольку быстро формирующиеся фрагменты структуры могут нуждаться в более высоких скоростях трансляции [61, 79, 98, 113–115]. Таким образом, было предложено рассматривать котрансляционное сворачивание как волновой процесс, протекающий через последовательные смены медленных и быстрых фаз, сопряженных с медленными и быстрыми фазами трансляции. Мы также предположили, что так называемый профиль частот использования кодонов, отражающий специфический характер распределения кодонов вдоль мРНК, может выступать в качестве кинетического путевода котрансляционного сворачивания белка [61, 79, 98, 113–115].

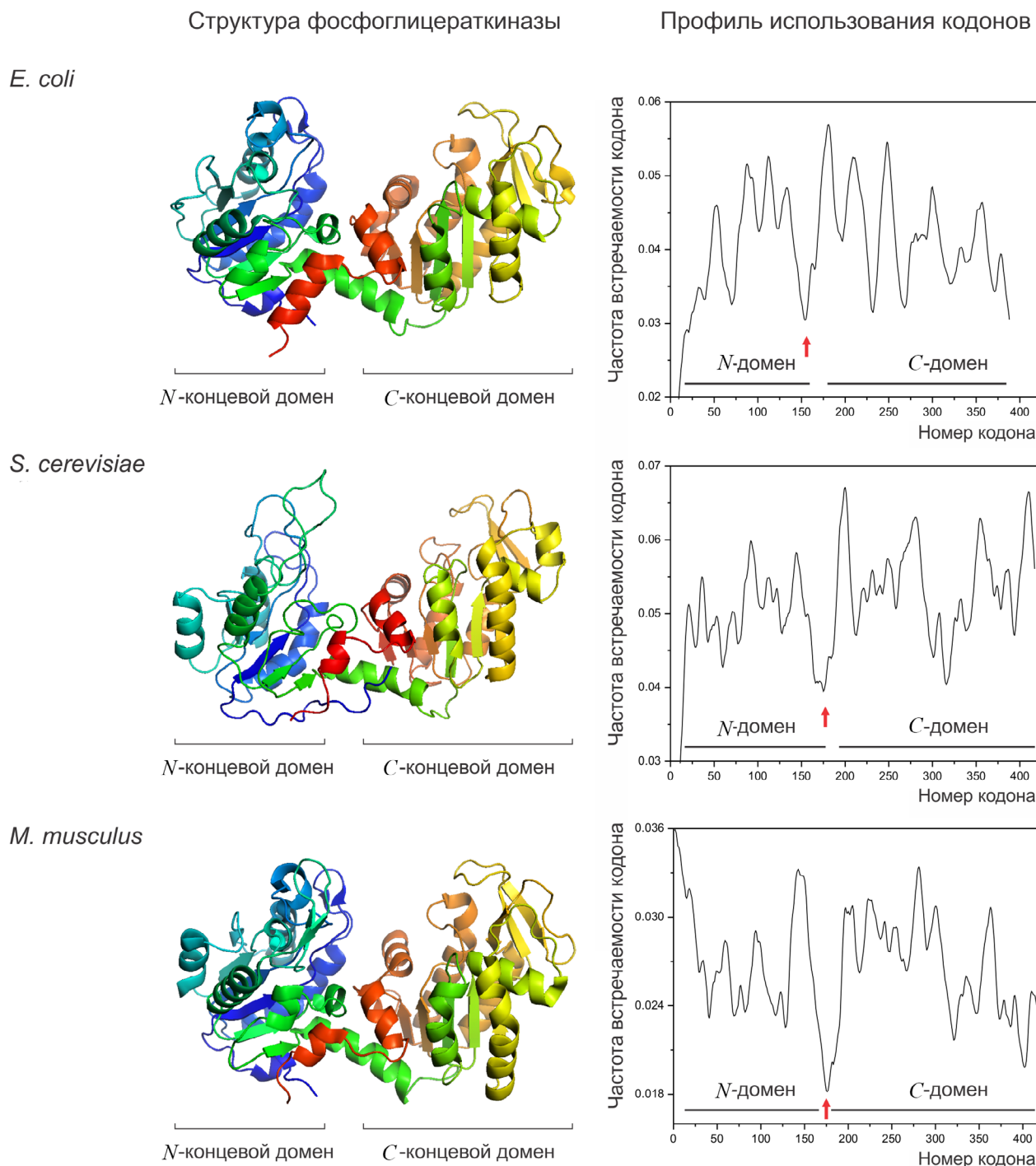
Основываясь на этих знаниях, мы задались вопросом о том, можно ли визуализировать связь между конкретной картиной распределения синонимических кодонов в мРНК и струк-

турой белка путем сравнения профилей использования кодонов структурно гомологичных белков. Мы предположили, что если такая связь существует, то структурно гомологичные белки из различных организмов (независимо от различий между предпочтениями кодонов у этих организмов) должны иметь сходные профили использования кодонов [115]. Мы обнаружили, что это действительно так [115]. Одно- и мультидоменные структурно гомологичные белки, такие как цитохромы *c*, миоглобины, гамма-кристаллины, фосфоглицераткиназы (рис. 2), как оказалось, имели очень схожие профили использования кодонов. В этих исследованиях нами было также продемонстрировано сохранение позиций кластеров наиболее редких кодонов, например, на границах доменов [79, 114, 115].

В нескольких последующих *in silico* исследованиях, в частности, по анализу хлорамфеникол-ацетилтрансфераз [131] и лакритинов глаза [136] из различных организмов, были получены сходные результаты. Следует также отметить, что совсем недавно было показано, что рибосомные профили (отражающие распределение транслирующихся рибосом на мРНК) оказались высококонсервативны у структурно и эволюционно родственных белков и белковых доменов, что дополнительно свидетельствует в пользу того факта, что кинетика трансляции мРНК у структурно родственных белков является высококонсервативным процессом, необходимым для обеспечения их правильного котрансляционного сворачивания [137].

Тем не менее, несмотря на широкое признание того факта, что часто используемые кодоны ускоряют трансляцию, а редкие кодоны вызывают транслационные паузы, получить экспериментальные доказательства того, что конкретный редко используемый кодон (или кластер редких кодонов) определяет соответствующую транслационную паузу на конкретном участке молекулы мРНК, оказалось достаточно затруднительным. Тем не менее в начале и середине 1990-х гг. мы смогли показать, используя одно- и мультидоменные белки, что участки замедленной трансляции могут действительно разделять этапы формирования отдельных структурных элементов, таких как альфа-спирали у глобинов [126] и/или домены у кристаллина гамма-В [138]. Интересно, что проведенное в то же время изучение кинетики индуцированного мочевиной равновесного разворачивания и сворачивания кристаллина гамма-В быка, показало, что сворачивание этого белка не только происходит через последовательное структурирование его доменов, но и также характеризуется дифференциальной кинетикой их сворачивания [139].





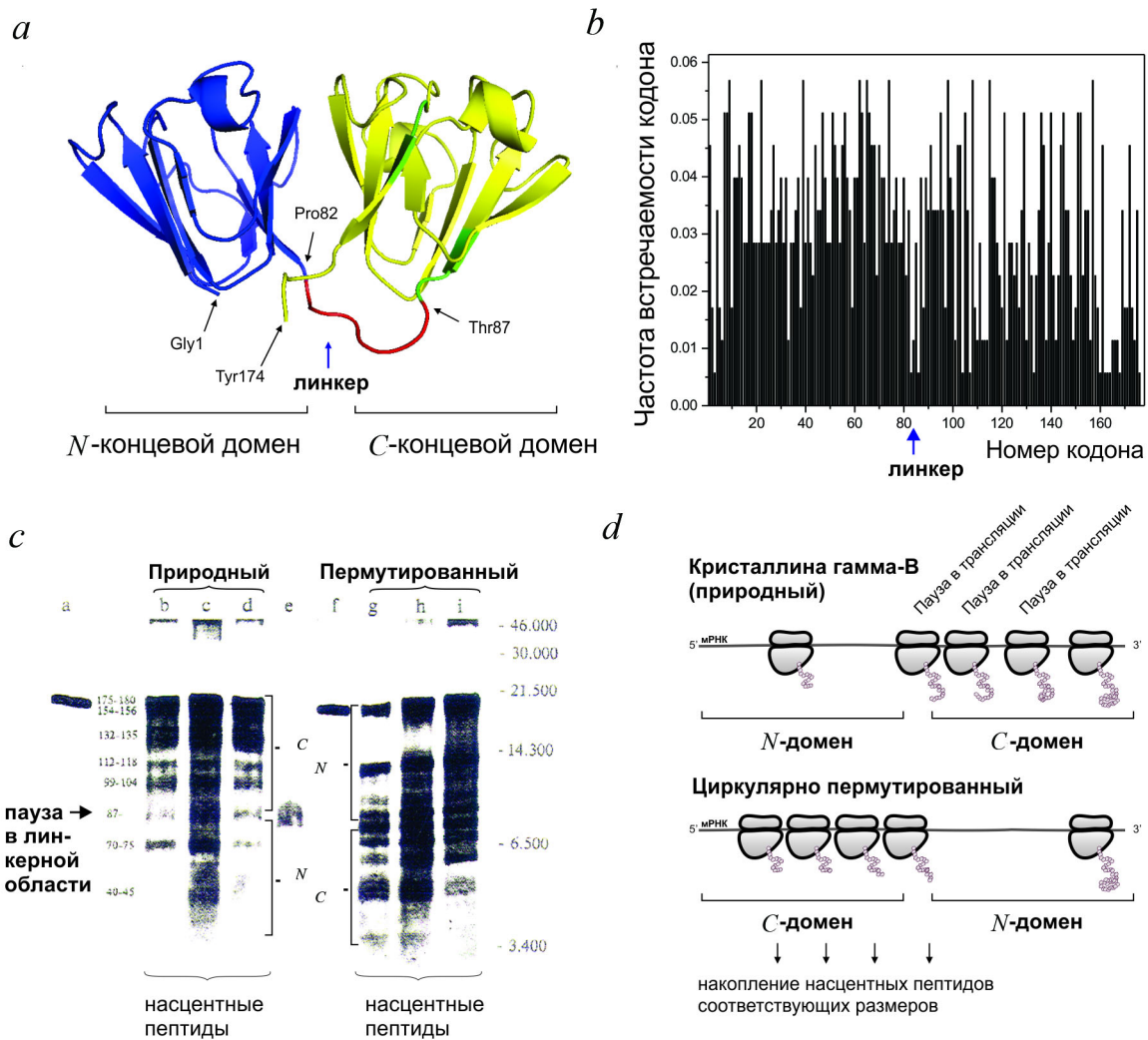
**Рис. 2.** Консерватизм структуры белка находит отражение в сходстве профилей использования кодонов. Модели третичной структуры фосфоглицераткиназ (PGK) из кишечной палочки *Escherichia coli* (PDB ID: 1ZMR) (сверху), дрожжей *Saccharomyces cerevisiae* (PDB ID: 3PGK) (посередине) и мыши *Mus musculus* (PDB ID: 2P9T) (внизу). Соответствующие профили использования кодонов указаны справа (обозначены границы между *N*- и *C*-концевыми доменами). Красными стрелками показан консервативный кластер редких кодонов в области границ доменов

Так, было показано, что сворачивание *N*-концевого домена происходит гораздо быстрее, чем *C*-концевого [139]. Проведенный нами впоследствии анализ процесса биосинтеза бычьего кристаллина гамма-В не только подтвердил

факт неравномерности трансляции мРНК этого белка с характерными паузами, но также позволил продемонстрировать (путем сравнения кинетики трансляции природного кристаллина гамма-В быка и его циркулярно пермутирован-

ного варианта с измененным порядком расположения *N*- и *C*-концевых доменов), что его *N*-концевой домен (в результате неравномерного распределения частых и редких кодонов в молекуле мРНК) подвергается более быстрой трансляции (о чем свидетельствовало меньшее число трансляционных пауз на этом участке

мРНК) по сравнению с *C*-концевым доменом [135] (рис. 3). Таким образом, мы пришли к заключению, что скорости трансляции могут быть действительно оптимизированы для настройки синтеза и сворачивания nascentных полипептидных цепей на рибосоме, и что для быстро сворачивающихся фрагментов белка



**Рис. 3.** Модульная (двухдоменная) структура кристаллина гамма-В и дифференциальная кинетика сворачивания его доменов отражены в различиях в использовании кодонов в его доменах и неравномерной трансляции белка. *a* – Модель третичной структуры кристаллина гамма-В быка (*Bos taurus*, PDB ID: 4GCR); *N*-концевой домен (быстро сворачивающийся) показан синим цветом, *C*-концевой (медленно сворачивающийся) – желтым; линкерный участок, связывающий два домена, показан красным цветом. *b* – Гистограмма использования/частоты кодонов: *N*-концевой домен (в среднем) кодируется более часто используемыми кодонами по сравнению с *C*-концевым; между двумя доменами есть четкая граница. *c* – [<sup>35</sup>S]-Авторадиограмма продуктов трансляции кристаллина гамма-В *in vitro* (из [138]), где *a* – кристаллин гамма-В спустя 15 мин трансляции; *b*–*d* – nascentные пептиды, выделенные из фракций полирибосом спустя 5, 10 и 15 мин трансляции белка гамма-В соответственно; *e* – отдельно транслированный *N*-концевой домен гамма-В; *f* – циркулярно пермутированный белок гамма (гамма-СР) спустя 15 мин трансляции; *g*–*i* – nascentные пептиды, выделенные из фракций полирибосом спустя 5, 10 и 15 мин трансляции гамма-СР. Основные пептиды обозначены цифрами в соответствии с их длиной (рассчитанной на основе молекулярного веса пептидов). Стрелками обозначена пауза в междоменном (линкерном) участке. *d* – Схематическое изображение ожидаемого распределения положения рибосом (пауз в трансляции) в случае кристаллина гамма-В дикого типа и циркулярно пермутированного белка. Увеличение времени нахождения рибосомы в определенных положениях на молекуле мРНК ведет к увеличению накопления nascentных цепей соответствующих размеров

действительно может быть необходима более быстрая трансляция соответствующих участков мРНК [139].

Следующим логичным вопросом был вопрос о том, насколько изменения в использовании синонимических кодонов в мРНК могут влиять на сворачивание белка в клетке?

В 1999 г. мы представили одно из первых доказательств того, что замены синонимических кодонов могут действительно влиять на движение рибосомы по мРНК, а также на сворачивание модельного белка CAT (chloramphenicol acetyltransferase, хлорамфениколацетилтрансферазы) в процессе его синтеза [140]. Мы показали, что замена 16-ти последовательно расположенных редких кодонов в гене *CAT* на частые кодоны приводит к ускорению движения рибосомы (в системе трансляции *in vitro*) на участке мРНК, содержащем замены этих кодонов, а также влияет на удельную активность фермента (в сравнении с белком дикого типа) [140]. Поскольку удельную активность белка можно рассматривать как меру его правильного сворачивания, то мы пришли к выводу о том, что замены кодонов оказали влияние на процесс сворачивания белка CAT [140].

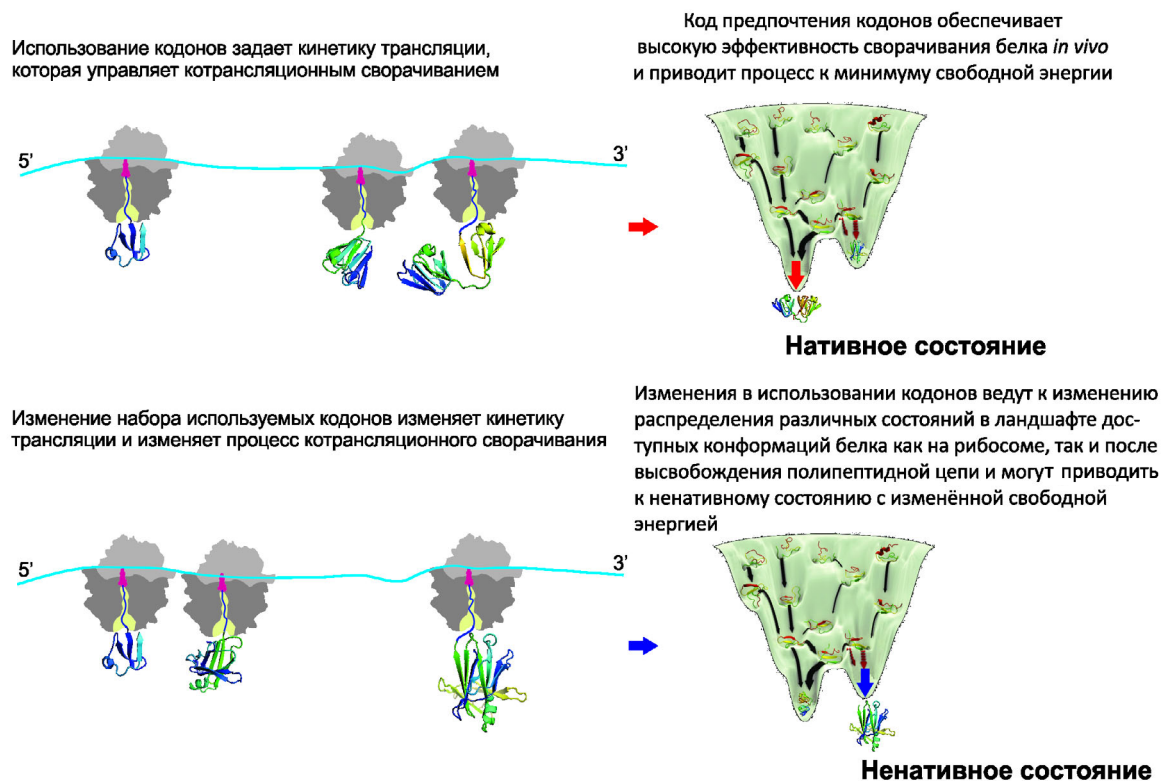
Во многих последующих экспериментах была подтверждена значимость использования синонимических кодонов для сворачивания белков и показано, что замены кодонов могут влиять на степень подверженности белков ограниченному протеолизу [141–143], изменять спектральные характеристики белков [144], изменять их подверженность агрегации [144–146], а также менять их удельную активность [141, 147], влиять на их внутриклеточные характеристики [148] и в конечном итоге приводить к возникновению заболеваний [149–152].

Кроме того, было высказано предположение, что выбор синонимичных кодонов может влиять на эффективность взаимодействия насцентных полипептидов с сигнал-распознающей частицей (SRP) [153], таким образом влияя на секрецию белка.

Следует отметить, что в гипотезе «кода предпочтения кодонов» [112–115] также было постулировано, что изменения кинетики трансляции должны в первую очередь приводить к изменению конформации прикрепленных к рибосоме насцентных цепей, что затем может приводить к изменению конечной конформации высвобождаемого полипептида и/или вызывать сдвиг равновесия между конформерами белка (правильно и неправильно свернутыми), что, в свою очередь, может привести, например, к повышенной агрегации белков и/или их деградации (ко- или посттрансляционной) или вызвать изменения

удельной активности белков и/или их функций (см. обзоры [61, 79]). Тем не менее прямые доказательства в поддержку этого постулата было получено лишь относительно недавно [154]. Для изучения того, как различное распределение синонимических кодонов влияет на кинетику трансляции, ко- и посттрансляционное сворачивание, а также конформацию белка и его стабильность, мы использовали кристаллин гамма-В быка, анализируя экспрессию этого белка *in vivo* в клетках *E. coli* и *in vitro* в полностью реконструированной системе бесклеточной трансляции. Мы сравнили экспрессию двух вариантов кристаллина гамма-В: один вариант характеризовался оптимальным использованием кодонов для трансляции белка в клетках *E. coli* (с профилем использования кодонов, сходным с таковым в клетках *B. taurus*, что, как ожидалось, должно приводить к более естественной кинетике трансляции), другой вариант — с неизменным составом кодонов, который не являлся оптимальным для трансляции белка в клетках *E. coli* [154]. Полученные нами данные о влиянии предпочтения синонимических кодонов на трансляцию мРНК кристаллина гамма-В показали, что выбор кодонов изменяет скорости трансляции (как общую скорость, так и локальную скорость считывания в определенных участках мРНК) и вызывает формирование альтернативных конформаций белка как в прикрепленных насцентных цепях, так и после высвобождения белка из рибосомы [154]. Важно подчеркнуть, что с помощью измерений в реальном времени на основе флуоресценции и FRET мы показали, что кинетика синтеза и котрансляционного сворачивания кристаллина гамма-В действительно изменяется в результате замен синонимических кодонов. Более того, нами впервые была выявлена значительная структурная гетерогенность очищенных полноразмерных препаратов синонимических вариантов кристаллина гамма-В, что было продемонстрировано с использованием прямого определения структуры методом 2D-NMR (двумерной ЯМР-спектроскопии), что также влияло на подверженность вариантов этих белков к агрегации [154]. Мы пришли к выводу, что синонимические замены в молекуле мРНК изменили распределение различных состояний в ландшафте доступных конформаций белка — как на рибосоме, так и после высвобождения полипептидной цепи [154] (рис. 4).

Эти эксперименты были проведены в сотрудничестве с Harald Schwalbe из университета Йоганна Вольфганга Гёте (Франкфурт, Германия) и Мариной Родниной из Института биофизической химии имени Макса Планка (Гёттинген, Германия), представив одно из самых убедитель-



**Рис. 4.** Генетический код управляет сворачиванием белка в клетке. Возникновение отдельных промежуточных интермедиатов сворачивания связано как с общей, так и с локальной скоростью трансляции, определяемой использованием синонимических кодонов. Изменение набора используемых кодонов вызывает изменение кинетики трансляции и процесса сворачивания белка как на рибосоме, так и после высвобождения полипептидной цепи, что также может приводить к изменению конечной конформации белка и отличному от нативного минимуму свободной энергии

ных доказательств гипотезы, согласно которой специфический характер распределения синонимических кодонов в молекуле мРНК может действительно служить в качестве вторичного кода для сворачивания белка в клетке [154].

#### КОД ПРЕДПОЧТЕНИЯ КОДОНОВ ДЛЯ СВРАЧИВАНИЯ БЕЛКА В КЛЕТКЕ: ЕГО ЗНАЧЕНИЕ

Открытие того факта, что замены синонимических кодонов могут не только влиять на уровни экспрессии белков, но также на их сворачивание, привело к углублению понимания мировым научным сообществом механизма влияния распределения синонимических кодонов на функции белков. Стало ясно, что замены синонимических кодонов могут играть роль в развитии заболеваний человека [149–154], а также потенциально влиять на безопасность и эффективность использования рекомбинантных/терапевтических белков [155–157]. К тому же, и что удивительно, было обнаружено, что даже еди-

ничная синонимическая мутация может нарушать сворачивание белка и его функционирование [141, 152, 158].

В 2007 г. Chava Kimchi-Sarfaty, Michael Gottesman и их коллеги показали, что одна единственная синонимическая мутация в гене *MDR1* (multi-drug resistance 1; ген множественной лекарственной резистентности 1), кодирующем белок P-gp (P-glycoprotein, P-гликопротеин) может вызывать изменение активности этого белка, таким образом влияя на чувствительность пациентов к действию различных лекарственных препаратов [141]. P-гликопротеин является трансмембранным белком, который выкачивает из клеток различные лекарства (см. [141] и ссылки в ней). В нормальных условиях он обеспечивает устойчивость клеток к экзогенным лекарственным веществам и/или эндогенным метаболитам [141, 149], однако у больных с онкологическими заболеваниями изменение активности этого белка может приводить к снижению эффективности химиотерапии, поэтому он представляет важную мишень в работе над доставкой лекарств [149]. Chava

Kimchi-Sarfaty и коллеги показали, что единичная синонимическая замена в гене *MDR1* (с.3435C>T, АТС>АТТ, р.Ile1145Ile) может приводить к изменению субстратной специфичности Р-гликопротеина [141]. Ранее было обнаружено, что концентрация в плазме и сыворотке крови, например, сердечного гликозида дигоксина, антигистаминового лекарства фексофенадина, который является агонистом  $\mu$ -опиоидного рецептора лоперамида, и некоторых других лекарств у людей с однонуклеотидным полиморфизмом с.3435C>T отличается (по сравнению с нормальными индивидами) после единичного или многократного (орального) приема этих лекарств, однако точная причина этих различий исходно не была выяснена [149]. В своей работе Kimchi-Sarfaty и коллеги объяснили причину изменения фармакокинетики Р-гр у людей, несущих этот полиморфизм, и впервые показали, что возникающие естественным образом вариации синонимических кодонов в определенном гене могут дать начало белковому продукту с той же аминокислотной последовательностью, но другими структурными и функциональными свойствами [141].

Впоследствии в сотрудничестве с Chava Kimchi-Sarfaty мы показали, что единичная синонимическая замена с.459G>A (GTG>GTA, р.Val153Val или р.Val107Val (107 – позиция аминокислоты после отщепления препропептида при секреции)) в гене *F9* (кодирующем фактор коагуляции крови FIX), которая была связана со снижением уровня свертывания крови и относительно легкой формой гемофилии В у ее носителей, влияет на синтез FIX и его конформацию, приводя к снижению уровня внеклеточного белка за счет его повышенной деградации [152]. Таким образом, мы смогли определить патогенетическую основу однонуклеотидной синонимической замены в гене *F9*, ассоциированной с гемофилией В [152]. Этот случай озадачивал исследователей, обнаруживших замену еще в 2008 г. [159], поскольку его нельзя было объяснить измененными свойствами мРНК (уровнем мРНК, сплайсингом/пропуском экзонов, удержанием интронов или измененной стабильностью мРНК), что обычно является причиной [150, 151, 160–163] возникновения многих заболеваний, ассоциированных с синонимическими мутациями, до тех пор, пока не стало ясно, что синонимические мутации могут влиять на биогенез белка.

Кроме того, также в сотрудничестве с Chava Kimchi-Sarfaty мы недавно показали, что единичная синонимическая замена с.354G>A (CCG>CCA, р.Pro118Pro) в гене *ADAMTS13* (дизинтегрин и металлопротеиназа с мотивом тром-

боспондина 1-го типа, 13-й член семейства), кодирующем большой мультидоменный секретлируемый белок, регулирующий тромбоз путем расщепления адгезивного гликопротеина крови, фактора фон Виллебранда (VWF – von Willebrand factor), может придать этому ферменту более высокую специфическую активность [158]. Эта замена является естественным вариантом, который присутствует в человеческой популяции (частота аллеля: 0,026 согласно 1000 Genomes, 0,0627 согласно ExAC) [158], и до настоящего момента она не была ассоциирована с патологическим фенотипом [158, 164]. На сегодняшний день вариант с.354G>A гена *ADAMTS13* представляет действительно уникальный пример, т.к. эта замена приводит к увеличению специфической активности белка [158]. Еще предстоит установить, дает ли эта мутация какое-либо преимущество её носителям.

Обнаружение того факта, что замены синонимических кодонов могут являться фактором риска для правильной функции белка (и влиять на структуру белка) добавляет дополнительные сложности в процесс производства рекомбинантных белков, включая белковые терапевтические средства. В настоящее время биофармацевтическая промышленность широко использует ряд доступных компьютерных программ и методов, которые учитывают предпочтение использования синонимических кодонов для направленного изменения генов с целью достижения их повышенной экспрессии (см. обзор [156]). Следует, однако, отметить, что эти инструменты обычно нацелены на увеличение эффективности трансляции (и обычно включают массовые замены большинства нечасто используемых кодонов на синонимические часто используемые) и обладают ограниченной (если вообще обладают) способностью оценки эффекта замен конкретных синонимических кодонов на сворачивание белка [156]. Тем не менее, как упоминалось выше, максимальное увеличение скорости и выхода белка при трансляции может вступать в определенные противоречия с процессом его синтеза и сворачивания, потенциально приводя к неправильному сворачиванию. Кроме того, все большее число работ показывает, что даже незначительные нарушения структуры белка могут вызвать изменения иммуногенных свойств белковых терапевтических средств и привести к образованию нейтрализующих антител [155, 165]. Эти данные представляют собой дополнительный фактор риска при производстве терапевтических средств на основе белков. В настоящее время оценка иммуногенного потенциала белковых терапевтических средств



представляет собой еще одну неудовлетворенную потребность в биотехнологической индустрии и в клинической практике [165].

Таким образом, становится очевидной необходимость разработки более совершенных стратегий оптимизации использования кодонов. Ряд компаний, включая, например, DAPCEL, Inc. [156], сейчас применяет вышеизложенные знания для разработки новых стратегий, направленных на продукцию (в любом желаемом организме-хозяине) правильно свернутых растворимых белков.

Тем не менее предстоит еще много работы. Необходимо отметить, что пока ещё не существует полного понимания точного влияния как единичных, так и множественных синонимических мутаций на сворачивание белка и его функционирование. Тем не менее, несмотря на отсутствие понимания многих деталей этого процесса, стало очевидным, что использование синонимических кодонов лежит в основе вторичного кода для сворачивания белков *in vivo* и осуществляет тонкую настройку процесса котрансляционного сворачивания, позволяя белкам обходить глубокие кинетические ловушки в процессе синтеза, обеспечивая быстрое и эффективное формирование белковой структуры в клетке. Таким образом, очевидно, что естественный отбор использует одновременно множество различных возможностей для оптимизации экспрессии генетической информации.

## ЗАКЛЮЧЕНИЕ

Настоящий обзор посвящен памяти Александра Сергеевича Спирина и его 90-летию юбилею, а также частично отражает его вклад в развитие области котрансляционного сворачивания белка.

Мой интерес к биосинтезу белка и, как я полагаю, интерес многих студентов и научных сотрудников кафедры молекулярной биологии биологического факультета МГУ им. М.В. Ломоносова, которую профессор Спирин возглавлял с 1972 по 2012 г., был в значительной степени вызван его лекциями по структуре рибосомы и механизму биосинтеза белка. Мне довелось посещать эти лекции в 1984 г., когда я был студентом кафедры молекулярной биологии. Они были поистине вдохновляющими, стимулирующими и вызвали множество вопросов. Мне посчастливилось продолжить мою учебу на этой

кафедре в качестве дипломника и аспиранта в группе профессора Владимира Владимировича Юркевича и затем в группе профессора Игоря Александровича Крашенинникова, проводя в том числе многие работы в тесном сотрудничестве с профессором Спириным. Я также имел возможность провести довольно много времени в лаборатории Спирина в Институте белка в Пушкино (в начале 1990-х гг.), проводя исследования по котрансляционному сворачиванию глобина. Этот опыт был бесценен для формирования моих научных взглядов и отношения к научной работе.

**Благодарности.** Настоящая работа была бы невозможна без исходного участия и вклада Ивана Аджубея и Игоря А. Крашенинникова, а также поддержки со стороны Славы Колба, Айгара Коммера, Валентина М. Степанова, Льва П. Овчинникова и Александра С. Спирина и последующих совместных исследований, проведенных с Райнером Янике, Клодом Райсом, а также в самое ближайшее время с Хавой Кимчи-Сарфати, Харальдом Швальбе и Мариной В. Родниной.

Я бесконечно признателен всем моим учителям, коллегам и сотрудникам за их чрезвычайно щедрые и вдохновляющие обсуждения и неоценимый вклад. Я также приношу извинения тем авторам, чьи работы или оригинальные публикации не были процитированы в этой короткой обзорной статье.

**Финансирование.** В последние годы работа моей лаборатории была поддержана грантами Human Frontier Science Program Organization [HFSP, грант №RGP0024/2010], the American Heart Association [AHA, грант 13GRNT17070025], the National Institutes of Health [NIH, гранты HL121779; HL151392 и GM128981], the Center for Gene Regulation in Health and Disease (GRHD) at CSU и биотехнологической компании DAPCEL, Inc.

**Конфликт интересов.** Я являюсь одним из основателей, а также вице-президентом по науке DAPCEL, Inc., биотехнологической компании, которая разрабатывает инновационные подходы для редизайна генов с целью производства правильно свернутых белков в любом желаемом организме-хозяине.

**Соблюдение этических норм.** Настоящая статья не содержит описания каких-либо исследований с участием людей в качестве объектов.

## СПИСОК ЛИТЕРАТУРЫ

1. Crick, F. H., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961) General nature of the genetic code for proteins, *Nature*, **192**, 1227-1232.
2. Nirenberg, M., and Leder, P. (1964) RNA codewords and protein synthesis. The effect of trinucleotides upon the binding of sRNA to ribosomes, *Science*, **145**, 1399-1407.
3. Leder, P., and Nirenberg, M. (1964) RNA codewords and protein synthesis. II. nucleotide sequence of a valine RNA codeword, *Proc. Natl. Acad. Sci. USA*, **52**, 420-427.
4. Bernfield, M. R., and Nirenberg, M. W. (1965) RNA codewords and protein synthesis. The nucleotide sequences of multiple codewords for phenylalanine, serine, leucine, and proline, *Science*, **147**, 479-484.
5. Trupin, J. S., Rottman, F. M., Brimacombe, R. L., Leder, P., Bernfield, M. R., and Nirenberg, M. W. (1965) RNA codewords and protein synthesis, VI. On the nucleotide sequences of degenerate codeword sets for isoleucine, tyrosine, asparagine, and lysine, *Proc. Natl. Acad. Sci. USA*, **53**, 807-811.
6. Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., et al. (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code, *Proc. Natl. Acad. Sci. USA*, **53**, 1161-1168.
7. Brimacombe, R., Trupin, J., Nirenberg, M., Leder, P., Bernfield, M., and Jaouni, T. (1965) RNA codewords and protein synthesis, 8. Nucleotide sequences of synonym codons for arginine, valine, cysteine, and alanine, *Proc. Natl. Acad. Sci. USA*, **54**, 954-960.
8. Söll, D., Ohtsuka, E., Jones, D. S., Lohrmann, R., Hayatsu, H., et al. (1965) Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids, *Proc. Natl. Acad. Sci. USA*, **54**, 1378-1385.
9. Goel, N. S., Rao, G. S., Ycas, M., Bremermann, H. J., and King, L. (1972) A method for calculating codon frequencies in DNA, *J. Theor. Biol.*, **35**, 399-457.
10. Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pavé, A. (1980) Codon catalog usage and the genome hypothesis, *Nucleic Acids Res.*, **8**, r49-r62.
11. Grantham, R., Gautier, C., and Gouy, M. (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type, *Nucleic Acids Res.*, **8**, 1893-1912.
12. Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes, *J. Mol. Biol.*, **146**, 1-21.
13. Ikemura, T. (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs, *J. Mol. Biol.*, **158**, 573-597.
14. Bennetzen, J. L., and Hall, B. D. (1982) Codon selection in yeast, *J. Biol. Chem.*, **257**, 3026-3031.
15. Hastings, K. E. M., and Emerson, C. P. Jr. (1983) Codon usage in muscle genes and liver genes, *J. Mol. Evol.*, **19**, 214-218.
16. Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms, *Mol. Biol. Evol.*, **2**, 13-34.
17. Sharp, P. M., Tuohy, T. M., and Mosurski, K. R. (1986) Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes, *Nucleic Acids Res.*, **14**, 5125-5143.
18. Sharp, P. M., and Li, W. H. (1987) The Codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res.*, **15**, 1281-1295.
19. Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H., and Wright, F. (1988) Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity, *Nucleic Acids Res.*, **16**, 8207-8211.
20. Sørensen, M.A., Kurland, C. G., and Pedersen, S. (1989) Codon usage determines translation rate in *Escherichia coli*, *J. Mol. Biol.*, **207**, 365-377.
21. Andersson, S. G., and Kurland, C. G. (1990) Codon preferences in free-living microorganisms, *Microbiol. Rev.*, **54**, 198-210.
22. Kurland, C. G. (1991) Codon bias and gene expression, *FEBS Lett.*, **285**, 165-169.
23. Kramer, E. B., and Farabaugh, P. J. (2007) The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition, *RNA*, **13**, 87-96.
24. Zaher, H. S., and Green, R. (2009) Fidelity at the molecular level: lessons from protein synthesis, *Cell*, **136**, 746-762.
25. Kramer, E. B., Vallabhaneni, H., Mayer, L. M., and Farabaugh, P. J. (2010) A comprehensive analysis of translational missense errors in the yeast *Saccharomyces cerevisiae*, *RNA*, **16**, 1797-1808.
26. Ribas de Pouplana, L., Santos, M. A., Zhu, J. H., Farabaugh, P. J., and Javid, B. (2014) Protein mistranslation: friend or foe? *Trends Biochem. Sci.*, **39**, 355-362.
27. Huang, Y., Koonin, E. V., Lipman, D. J., and Przytycka, T. M. (2009) Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage, *Nucleic Acids Res.*, **37**, 6799-6810.
28. Savir, Y., and Tlusty, T. (2013) The ribosome as an optimal decoder: a lesson in molecular recognition, *Cell*, **153**, 471-479.
29. Dana, A., and Tuller, T. (2014) The effect of tRNA levels on decoding times of mRNA codons, *Nucleic Acids Res.*, **42**, 9171-9181.
30. Bonekamp, F., Andersen, H. D., Christensen, T., and Jensen, K. F. (1985) Codon-defined ribosomal pausing in *Escherichia coli* detected by using the pyrE attenuator to probe the coupling between transcription and translation, *Nucleic Acids Res.*, **13**, 4113-4123.
31. Curran, J. F., and Yarus, M. (1989) Rates of aminoacyl-tRNA selection at 29 sense codons *in vivo*, *J. Mol. Biol.*, **209**, 65-77.
32. Bulmer, M. (1987) Coevolution of codon usage and transfer RNA abundance, *Nature*, **325**, 728-730.
33. Duret, L. (2000) tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes, *Trends Genet.*, **16**, 287-289.
34. Li, W. H. (1987) Models of nearly neutral mutation with particular implications for nonrandom usage of synonymous codons, *J. Mol. Evol.*, **24**, 337-345.
35. Shields, D. C. (1990) Switches in species-specific codon preferences: the influence of mutation biases, *J. Mol. Evol.*, **31**, 71-80.
36. Bulmer, M. (1991) The selection-mutation-drift theory of synonymous codon usage, *Genetics*, **129**, 897-907.
37. Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., et al. (2015) Codon optimality is a major determinant of mRNA stability, *Cell*, **160**, 1111-1124, doi: 10.1016/j.cell.2015.02.029.
38. Boël, G., Letso, R., Neely, H., Price, W. N., Wong, K. H., et al. (2016) Codon influence on protein expression in *E. coli* correlates with mRNA levels, *Nature*, **529**, 358-363.



39. Mishima, Y., and Tomari, Y. (2016) Codon usage and 3'-UTR length determine maternal mRNA stability in zebrafish, *Mol. Cell*, **61**, 874-885.
40. Bazzini, A. A., Del Viso, F., Moreno-Mateos, M. A., Johnstone, T. G., Vejnar, C. E., et al. (2016) Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition, *EMBO J.*, **35**, 2087-2103
41. Radhakrishnan, A., Chen, Y. H., Martin, S., Alhusaini, N., Green, R., and Collier, J. (2016) The DEAD-box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality, *Cell*, **167**, 122-132.
42. Hia, F., Yang, S. F., Shichino, Y., Yoshinaga, M., Murakawa, Y., et al. (2019) Codon bias confers stability to human mRNAs. *EMBO Rep.*, **20**, e48220.
43. Wu, Q., Medina, S. G., Kushawah, G., DeVore, M. L., Castellano, L. A., et al. (2019) Translation affects mRNA stability in a codon-dependent manner in human cells, *Elife*, **8**, e45396.
44. Medina-Muñoz, S. G., Kushawah, G., Castellano, L. A., Diez, M., DeVore, M. L., et al. (2021) Crosstalk between codon optimality and cis-regulatory elements dictates mRNA stability, *Genome Biol.*, **22**, 14.
45. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science*, **181**, 223-230.
46. Jaenicke, R. (1991) Protein folding: local structures, domains, subunits, and assemblies, *Biochemistry*, **30**, 3147-3161.
47. Fersht, A. R. (2008) From the first protein structures to our current knowledge of protein folding: delights and skepticisms, *Nat. Rev. Mol. Cell Biol.*, **9**, 650-654.
48. Bartlett, A. I., and Radford, S. E. (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms, *Nat. Struct. Mol. Biol.*, **16**, 582-588.
49. Finkelstein, A. V. (2018) 50+ years of protein folding, *Biochemistry (Moscow)*, **83** (Suppl 1), S3-S18.
50. Abaskharon, R. M., and Gai, F. (2016) Meandering down the energy landscape of protein folding: are we there yet? *Biophys. J.*, **110**, 1924-1932.
51. Ferina, J., and Daggett, V. (2019) Visualizing protein folding and unfolding, *J. Mol. Biol.*, **431**, 1540-1564.
52. Cowie, D. B., Spiegelman, S., Roberts, R. B., and Duerksen, J. D. (1961) Ribosome-bound  $\beta$ -galactosidase, *Proc. Natl. Acad. Sci. USA*, **47**, 114-122
53. Zipser, D., and Perrin, D. (1963) Complementation on ribosomes, *Cold Spring Harb. Symp. Quant. Biol.*, **28**, 533-537.
54. Kiho, Y., and Rich, A. (1964) Induced enzyme formed on bacterial polyribosomes, *Proc. Natl. Acad. Sci. USA*, **51**, 111-118.
55. Hamlin, J., and Zabin, I. (1972)  $\beta$ -Galactosidase: immunological activity of ribosome-bound, growing polypeptide chains, *Proc. Natl. Acad. Sci. USA*, **69**, 412-416.
56. Bergman, L. W., and Kuehl, W. M. (1979) Formation of intermolecular disulfide bonds on nascent immunoglobulin polypeptides, *J. Biol. Chem.*, **254**, 5690-5694.
57. Bergman, L. W., and Kuehl, W. M. (1979) Formation of an intrachain disulfide bond on nascent immunoglobulin light chains, *J. Biol. Chem.*, **254**, 8869-8876.
58. Bergman, L. W., and Kuehl, W. M. (1979) Co-translational modification of nascent immunoglobulin heavy and light chains, *J. Supramol. Struct.*, **11**, 9-24.
59. Lim, V. I., and Spirin, A. S. (1985) Stereochemistry of the transpeptidation reaction in the ribosome. The ribosome generates an alpha-helix in the synthesis of the protein polypeptide chain, *Dokl. Akad. Nauk. SSSR*, **280**, 235-239.
60. Lim, V. I., and Spirin, A. S. (1986) Stereochemical analysis of ribosomal transpeptidation. Conformation of nascent peptide, *J. Mol. Biol.*, **188**, 565-574.
61. Komar, A. A. (2019) Synonymous codon usage—a guide for co-translational protein folding in the cell, *Mol. Biol. (Mosk)*, **53**, 883-898.
62. Noé, F., De Fabritiis, G., and Clementi, C. (2020) Machine learning for protein folding and dynamics, *Curr. Opin. Struct. Biol.*, **60**, 77-84.
63. Gao, W., Mahajan, S. P., Sulam, J., and Gray, J. J. (2020) Deep learning in protein structural modeling and design, *Patterns (N Y)*, **1**, 100142.
64. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., et al. (2020) Improved protein structure prediction using potentials from deep learning, *Nature*, **577**, 706-710.
65. Lindberg, M. O., and Oliveberg, M. (2007) Malleability of protein folding pathways: a simple reason for complex behaviour, *Curr. Opin. Struct. Biol.*, **17**, 21-29.
66. Englander, S. W., and Mayne, L. (2014) The nature of protein folding pathways, *Proc. Natl. Acad. Sci. USA*, **111**, 15873-15880.
67. Englander, S. W., Mayne, L., Kan, Z. Y., and Hu, W. (2016) Protein folding—how and why: by hydrogen exchange, fragment separation, and mass spectrometry, *Annu. Rev. Biophys.*, **45**, 135-152.
68. Levinthal, C. (1969). How to fold gracefully, in *Mossbauer Spectroscopy in Biological Systems*, Proceedings of a Meeting held at Allerton House, Monticello, Illinois (Debrunner, P., Tsibris, J. C. M., and Münck, E., eds.) University of Illinois Press, Urbana, p. 22.
69. Gopan, G., Gruebele, M., and Rickard, M. (2020) In-cell protein landscapes: making the match between theory, simulation and experiment, *Curr. Opin. Struct. Biol.*, **66**, 163-169.
70. Gershenson, A., Gosavi, S., Faccioli, P., and Wintrode, P. L. (2020) Successes and challenges in simulating the folding of large proteins, *J. Biol. Chem.*, **295**, 15-33.
71. Chow, M. K., Amin, A. A., Fulton, K. F., Fernando, T., Kamau, L., et al. (2006) The REFOLD database: a tool for the optimization of protein expression and refolding, *Nucleic Acids Res.*, **34** (Database issue), D207-212.
72. Mizutani, H., Sugawara, H., Buckle, A. M., Sangawa, T., Miyazono, K. I., et al. (2017) REFOLDdb: a new and sustainable gateway to experimental protocols for protein refolding, *BMC Struct. Biol.*, **17**, 4.
73. Hartl, F. U., and Hayer-Hartl, M. (2009) Converging concepts of protein folding *in vitro* and *in vivo*, *Nat. Struct. Mol. Biol.*, **16**, 574-581.
74. Hingorani, K. S., and Gierasch, L. M. (2014) Comparing protein folding *in vitro* and *in vivo*: foldability meets the fitness challenge, *Curr. Opin. Struct. Biol.*, **24**, 81-90.
75. Balchin, D., Hayer-Hartl, M., and Hartl, F. U. (2016) *In vivo* aspects of protein folding and quality control, *Science*, **353**, aac4354.
76. Gruebele, M., Dave, K., and Sukenik, S. (2016) Globular protein folding *in vitro* and *in vivo*, *Annu. Rev. Biophys.*, **45**, 233-251.
77. Dahiya, V., and Buchner, J. (2019) Functional principles and regulation of molecular chaperones, *Adv. Protein. Chem. Struct. Biol.*, **114**, 1-60.
78. Jayaraj, G. G., Hipp, M. S., and Hartl, F. U. (2019) Functional modules of the proteostasis network, *Cold Spring Harb. Perspect. Biol.*, **12**, a033951.
79. Komar, A. A. (2009) A pause for thought along the co-translational folding pathway, *Trends Biochem. Sci.*, **34**, 16-24.
80. Gloge, F., Becker, A. H., Kramer, G., and Bukau, B. (2014) Co-translational mechanisms of protein maturation, *Curr. Opin. Struct. Biol.*, **24**, 24-33.

81. Thommen, M., Holtkamp, W., and Rodnina, M. V. (2017) Co-translational protein folding: progress and methods, *Curr. Opin. Struct. Biol.*, **42**, 83-89.
82. Komar, A. A. (2018) Unraveling co-translational protein folding: concepts and methods, *Methods*, **137**, 71-81.
83. Williams, N. K., and Dichtl, B. (2018) Co-translational control of protein complex formation: a fundamental pathway of cellular organization? *Biochem. Soc. Trans.*, **46**, 197-206.
84. Waudby, C. A., Dobson, C. M., and Christodoulou, J. (2019) Nature and regulation of protein folding on the ribosome, *Trends Biochem. Sci.*, **44**, 914-926.
85. Jha, S., and Komar, A. A. (2011) Birth, life and death of nascent polypeptide chains, *Biotechnol. J.*, **6**, 623-640.
86. Das, D., Das, A., Samanta, D., Ghosh, J., Dasgupta, S., et al. (2008) Role of the ribosome in protein folding, *Biotechnol. J.*, **3**, 999-1009.
87. Voisset, C., Saupe, S. J., and Blondel, M. (2011) The various facets of the protein-folding activity of the ribosome, *Biotechnol. J.*, **6**, 668-673.
88. Banerjee, D., and Sanyal, S. (2014) Protein folding activity of the ribosome (PFAR) – a target for antiprion compounds, *Viruses*, **6**, 3907-3924.
89. Woolhead, C. A., McCormick, P. J., and Johnson, A. E. (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins, *Cell*, **116**, 725-736.
90. Lu, J., and Deutsch, C. (2005) Folding zones inside the ribosomal exit tunnel, *Nat. Struct. Mol. Biol.*, **12**, 1123-1129.
91. Wilson, D. N., and Beckmann, R. (2011) The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling, *Curr. Opin. Struct. Biol.*, **21**, 274-282.
92. Tu, L., and Deutsch, C. (2017) Determinants of helix formation for a Kv1.3 transmembrane segment inside the ribosome exit tunnel, *J. Mol. Biol.*, **429**, 1722-1732.
93. Gilbert, R. J., Fucini, P., Connell, S., Fuller, S. D., Nierhaus, K. H., et al. (2004) Three-dimensional structures of translating ribosomes by Cryo-EM, *Mol. Cell*, **14**, 57-66.
94. Kosolapov, A., and Deutsch, C. (2009) Tertiary interactions within the ribosomal exit tunnel, *Nat. Struct. Mol. Biol.*, **16**, 405-411.
95. Tu, L., Khanna, P., and Deutsch, C. (2014) Transmembrane segments form tertiary hairpins in the folding vestibule of the ribosome, *J. Mol. Biol.*, **426**, 185-198.
96. Holtkamp, W., Kocic, G., Jäger, M., Mittelstaet, J., Komar, A. A., and Rodnina, M. V. (2015) Cotranslational protein folding on the ribosome monitored in real time, *Science*, **350**, 1104-1107.
97. Chen, X., Rajasekaran, N., Liu, K., and Kaiser, C. M. (2020) Synthesis runs counter to directional folding of a nascent protein domain, *Nat. Commun.*, **11**, 5096.
98. Komar, A. A. (2018) The Yin and Yang of codon usage, *Hum. Mol. Genet.*, **25**(R2), R77-R85.
99. Komar, A. A. (2007) SNPs, silent but not invisible, *Science*, **315**, 466-467.
100. Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., and Nussinov, R. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima, *J. Mol. Biol.*, **383**, 281-291.
101. Zhang, G., and Ignatova, Z. (2011) Folding at the birth of the nascent chain: coordinating translation with co-translational folding, *Curr. Opin. Struct. Biol.*, **21**, 25-31.
102. O'Brien, E. P., Ciryam, P., Vendruscolo, M., and Dobson, C. M. (2014) Understanding the influence of codon translation rates on cotranslational protein folding, *Acc. Chem. Res.*, **47**, 1536-1544.
103. Chaney, J. L., and Clark, P. L. (2015) Roles for synonymous codon usage in protein biogenesis, *Annu. Rev. Biophys.*, **44**, 143-166.
104. Jacobson, G. N., and Clark, P. L. (2016) Quality over quantity: optimizing co-translational protein folding with non-“optimal” synonymous codons, *Curr. Opin. Struct. Biol.*, **38**, 102-110.
105. Sharma, A. K., and O'Brien, E. P. (2018) Non-equilibrium coupling of protein structure and function to translation-elongation kinetics, *Curr. Opin. Struct. Biol.*, **49**, 94-103.
106. Samatova, E., Daberge, J., Liutkute, M., and Rodnina, M. V. (2021) Translational control by ribosome pausing in bacteria: how a non-uniform pace of translation affects protein production and folding, *Front. Microbiol.*, **11**, 619430.
107. Liu, Y., Yang, Q., and Zhao, F. (2021) Synonymous but not silent: the codon usage code for gene expression and protein folding, *Annu. Rev. Biochem.*, doi: 10.1146/annurev-biochem-071320-112701.
108. Pelham, H. R., and Jackson, R. J. (1976) An efficient mRNA-dependent translation system from reticulocyte lysates, *Eur. J. Biochem.*, **67**, 247-256.
109. Kolb, V. A., Makeyev, E. V., and Spirin, A. S. (1994) Folding of firefly luciferase during translation in a cell-free system, *EMBO J.*, **13**, 3631-3637.
110. Komar, A. A., Kommer, A., Krasheninnikov, I. A., and Spirin, A. S. (1993) Cotranslational heme binding to nascent globin chains, *FEBS Lett.*, **326**, 261-263.
111. Komar, A. A., Kommer, A., Krasheninnikov, I. A., and Spirin, A. S. (1997) Cotranslational folding of globin, *J. Biol. Chem.*, **272**, 10646-10651.
112. Purvis, I. J., Bettany, A. J., Santiago, T. C., Coggins, J. R., Duncan, K., et al. (1987) The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo*. A hypothesis, *J. Mol. Biol.*, **193**, 413-417.
113. Krasheninnikov, I. A., Komar, A. A., and Adzhubei, I. A. (1988) Role of the rare codon clusters in defining the boundaries of polypeptide chain regions with identical secondary structures in the process of co-translational folding of proteins [in Russian], *Dokl. Akad. Nauk SSSR*, **303**, 995-999.
114. Krasheninnikov, I. A., Komar, A. A., and Adzhubei, I. A. (1989) Frequency of using codons in mRNA and coding of the domain structure of proteins [in Russian], *Dokl. Akad. Nauk. SSSR*, **305**, 1006-1012.
115. Krasheninnikov, I. A., Komar, A. A., and Adzhubei, I. A. (1989) Role of the code redundancy determining cotranslational protein folding, *Biokhimiia*, **5**, 187-200.
116. Heinemann, U., and Hahn, M. (1995) Circular permutation of polypeptide chains: implications for protein folding and stability, *Prog. Biophys. Mol. Biol.*, **64**, 121-143.
117. Protzel, A., and Morris, A. J. (1974) Gel chromatographic analysis of nascent globin chains. Evidence of nonuniform size distribution, *J. Biol. Chem.*, **249**, 4594-4600.
118. Chaney, W. G., and Morris, A. J. (1978) Nonuniform size distribution of nascent peptides: the role of messenger RNA, *Arch. Biochem. Biophys.*, **191**, 734-741.
119. Lizardi, P. M., Mahdavi, V., Shields, D., and Candelas, G. (1979) Discontinuous translation of silk fibroin in a reticulocyte cell-free system and in intact silk gland cells, *Proc. Natl. Acad. Sci. USA*, **76**, 6211-6215.
120. Randall, L. L., Josefsson, L. G., and Hardy, S. J. (1980) Novel intermediates in the synthesis of maltose-binding protein in *Escherichia coli*, *Eur. J. Biochem.*, **107**, 375-379.
121. Abraham, A. K., and Pihl, A. (1980) Variable rate of polypeptide chain elongation *in vitro*. Effect of spermidine, *Eur. J. Biochem.*, **106**, 257-262.
122. Varenne, S., Knibiehler, M., Cavard, D., Morlon, J., and Lazdunski, C. (1982) Variable rate of polypeptide chain

- elongation for colicins A, E2 and E3, *J. Mol. Biol.*, **159**, 57-70.
123. Candelas, G., Candelas, T., Ortiz, A., and Rodríguez, O. (1983) Translational pauses during a spider fibroin synthesis, *Biochem. Biophys. Res. Commun.*, **116**, 1033-1038.
124. Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984) Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains, *J. Mol. Biol.*, **180**, 549-576.
125. Wolin, S. L., and Walter, P. (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA, *EMBO J.*, **7**, 3559-3569.
126. Krasheninnikov, I. A., Komar, A. A., and Adzhubei, I. A. (1991) Nonuniform size distribution of nascent globin peptides, evidence for pause localization sites, and a cotranslational protein-folding model, *J. Protein. Chem.*, **10**, 445-454.
127. Thanaraj, T. A., and Argos, P. (1996) Ribosome-mediated translational pause and protein domain organization, *Protein Sci.*, **5**, 1594-1612.
128. Thanaraj, T. A., and Argos, P. (1996) Protein secondary structural types are differentially coded on messenger RNA, *Protein Sci.*, **5**, 1973-1983.
129. Adzhubei, I. A., Adzhubei, I. A., Krasheninnikov, I. A., and Neidle, S. (1996) Non-random usage of "degenerate" codons is related to protein three-dimensional structure, *FEBS Lett.*, **399**, 78-82.
130. Oresic, M., and Shalloway, D. (1998) Specific correlations between relative synonymous codon usage and protein secondary structure, *J. Mol. Biol.*, **281**, 31-48.
131. Widmann, M., Clairou, M., Dippon, J., and Pleiss, J. (2008) Analysis of the distribution of functionally relevant rare codons, *BMC Genomics*, **9**, 207.
132. Clarke, T. F. 4th, and Clark, P. L. (2008) Rare codons cluster, *PLoS One*, **3**, e3412.
133. Chartier, M., Gaudreault, F., and Najmanovich, R. (2012) Large-scale analysis of conserved rare codon clusters suggests an involvement in co-translational molecular recognition events, *Bioinformatics*, **28**, 1438-1445.
134. Ma, X. X., Wang, Y. N., Cao, X. A., Li, X. R., Liu, Y. S., et al. (2018) The effects of codon usage on the formation of secondary structures of nucleocapsid protein of peste des petits ruminants virus, *Genes Genomics*, **40**, 905-912.
135. Newaz, K., Wright, G., Piland, J., Li, J., Clark, P. L., et al. (2020) Network analysis of synonymous codon usage, *Bioinformatics*, **36**, 4876-4884.
136. McKown, R. L., Raab, R. W., Kachelries, P., Caldwell, S., and Laurie, G. W. (2013) Conserved regional 3' grouping of rare codons in the coding sequence of ocular prosecretory mitogen lacritin, *Invest. Ophthalmol. Vis. Sci.*, **54**, 1979-1987.
137. Nissley, D. A., Carbery, A., Chonofsky, M., and Deane, C. M. (2021) Ribosome occupancy profiles are conserved between structurally and evolutionarily related yeast domains, *Bioinformatics*, doi: 10.1093/bioinformatics/ctab020.
138. Komar, A. A., and Jaenicke, R. (1995) Kinetics of translation of gamma B crystallin and its circularly permuted variant in an *in vitro* cell-free system: possible relations to codon distribution and protein folding, *FEBS Lett.*, **376**, 195-198.
139. Rudolph, R., Siebendritt, R., Nessler, G., Sharma, A. K., and Jaenicke, R. (1990) Folding of an all-beta protein: independent domain folding in gamma II-crystallin from calf eye lens, *Proc. Natl. Acad. Sci. USA*, **87**, 4625-4629.
140. Komar, A. A., Lesnik, T., and Reiss, C. (1999) Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. *FEBS Lett.*, **462**, 387-391.
141. Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., et al. (2007) A "silent" polymorphism in the *MDR1* gene changes substrate specificity, *Science*, **315**, 525-528.
142. Zhang, G., Hubalewska, M., and Ignatova, Z. (2009) Transient ribosomal attenuation coordinates protein synthesis and co-translational folding, *Nat. Struct. Mol. Biol.*, **16**, 274-280.
143. Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., et al. (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ, *Nature*, **495**, 111-115.
144. Sander, I. M., Chaney, J. L., and Clark, P. L. (2014) Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design, *J. Am. Chem. Soc.*, **136**, 858-861.
145. Hu, S., Wang, M., Cai, G., and He, M. (2013) Genetic code-guided protein synthesis and folding in *Escherichia coli*, *J. Biol. Chem.*, **288**, 30855-30861.
146. Kim, S. J., Yoon, J. S., Shishido, H., Yang, Z., Rooney, L. A., et al. (2015) Protein folding. Translational tuning optimizes nascent protein folding in cells, *Science*, **348**, 444-448.
147. Yu, C. H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., et al. (2015) Codon usage influences the local rate of translation elongation to regulate co-translational protein folding, *Mol. Cell*, **59**, 744-754.
148. Walsh, I. M., Bowman, M. A., Soto Santarriaga, I. F., Rodriguez, A., and Clark, P. L. (2020) Synonymous codon substitutions perturb cotranslational protein folding *in vivo* and impair cell fitness, *Proc. Natl. Acad. Sci. USA*, **117**, 3528-3534.
149. Komar, A. A. (2007) Silent SNPs: impact on gene function and phenotype, *Pharmacogenomics*, **8**, 1075-1080.
150. Sauna, Z. E., and Kimchi-Sarfaty, C. (2011) Understanding the contribution of synonymous mutations to human disease, *Nat. Rev. Genet.*, **12**, 683-691.
151. Hunt, R. C., Simhadri, V. L., Iandoli, M., Sauna, Z. E., and Kimchi-Sarfaty, C. (2014) Exposing synonymous mutations, *Trends Genet.*, **30**, 308-321.
152. Simhadri, V. L., Hamasaki-Katagiri, N., Lin, B. C., Hunt, R., Jha, S., et al. (2017) Single synonymous mutation in factor IX alters protein properties and underlies haemophilia B, *J. Med. Genet.*, **54**, 338-345.
153. Pechmann, S., Chartron, J. W., and Frydman, J. (2014) Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP *in vivo*, *Nat. Struct. Mol. Biol.*, **21**, 1100-1105.
154. Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., et al. (2016) Synonymous codons direct cotranslational folding toward different protein conformations, *Mol. Cell*, **61**, 341-351.
155. Kimchi-Sarfaty, C., Schiller, T., Hamasaki-Katagiri, N., Khan, M. A., Yanover, C., and Sauna, Z. E. (2013) Building better drugs: developing and regulating engineered therapeutic proteins, *Trends Pharmacol. Sci.*, **34**, 534-548.
156. Komar, A. A. (2016) The art of gene redesign and recombinant protein production: approaches and perspectives, *Top. Med. Chem.*, **2**, 1-17.
157. Alexaki, A., Hettiarachchi, G. K., Athey, J. C., Katneni, U. K., Simhadri, V., et al. (2019) Effects of codon optimization on coagulation factor IX translation and structure: implications for protein and gene therapies, *Sci. Rep.*, **9**, 15449.
158. Hunt, R., Hettiarachchi, G., Katneni, U., Hernandez, N., Holcomb, D., et al. (2019) A single synonymous variant (c.354G>A [p.P118P]) in ADAMTS13 confers enhanced specific activity, *Int. J. Mol. Sci.*, **20**, 5734.

159. Knobe, K. E., Sjorin, E., and Ljung, R. C. (2008) Why does the mutation G17736A/Val107Val (silent) in the *F9* gene cause mild haemophilia B in five Swedish families? *Haemophilia*, **14**, 723-728.
160. Gartner, J. J., Parker, S. C., Prickett, T. D., Dutton-Regester, K., Stitzel, M. L., et al. (2013) Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma, *Proc. Natl. Acad. Sci. USA*, **110**, 13481-13486.
161. Hamasaki-Katagiri, N., Salari, R., Wu, A., Qi, Y., Schiller, T., et al. (2013) A gene-specific method for predicting hemophilia-causing point mutations, *J. Mol. Biol.*, **425**, 4023-4033.
162. Hamasaki-Katagiri, N., Lin, B. C., Simon, J., Hunt, R. C., Schiller, T., et al. (2017) The importance of mRNA structure in determining the pathogenicity of synonymous and non-synonymous mutations in haemophilia, *Haemophilia*, **23**, e8-e17.
163. Katneni, U. K., Liss, A., Holcomb, D., Katagiri, N. H., Hunt, R., et al. (2019) Splicing dysregulation contributes to the pathogenicity of several *F9* exonic point variants, *Mol. Genet. Genomic Med.*, **7**, e840.
164. Edwards, N. C., Hing, Z. A., Perry, A., Blaisdell, A., Kopelman, D. B., F et al. (2012) Characterization of coding synonymous and non-synonymous variants in ADAMTS13 using *ex vivo* and *in silico* approaches, *PLoS One*, **7**, e38864.
165. Sauna, Z. E., Richards, S. M., Maillere, B., Jury, E. C., and Rosenberg, A. S. (2020) Editorial: immunogenicity of proteins used as therapeutics, *Front. Immunol.*, **11**, 614856.

## A CODE WITHIN A CODE: HOW CODONS FINE-TUNE PROTEIN FOLDING IN THE CELL

### Review

A. A. Komar<sup>1,2,3,4</sup>

<sup>1</sup> Center for Gene Regulation in Health and Disease and Department of Biological, Geological and Environmental Sciences, Cleveland State University, Cleveland, Ohio 44115, USA; e-mail: a.komar@csuohio.edu

<sup>2</sup> Department of Biochemistry and Center for RNA Science and Therapeutics, Case Western Reserve University, Cleveland, Ohio 44106, USA

<sup>3</sup> Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA

<sup>4</sup> DAPCEL, Inc., Cleveland, Ohio 44106, USA

The genetic code sets the correspondence between the sequence of a given nucleotide triplet in an mRNA molecule, called a codon, and the amino acid that it directs to be added to the growing polypeptide chain during protein synthesis. With four bases (A, G, U, and C), there are 64 possible triplet codons; 61 sense (encoding amino acids) and 3 nonsense (so-called, stop codons that direct termination of translation). In most organisms, there are 20 common/standard amino acids used in protein synthesis; thus, the genetic code is redundant with most amino acids (with the exception of Met and Trp) are being encoded by more than one (synonymous) codon. Synonymous codons were initially presumed to have entirely equivalent functions, however, the finding that synonymous codons are not present at equal frequencies in mRNA suggested that specific codon choice might have functional implications beyond amino acid coding. Observation of nonequivalent use of codons in mRNAs raised a possibility of the existence of auxiliary information in the genetic code. Indeed, it has been found that genetic code contains several layers of such additional information and that synonymous codons are strategically placed within mRNAs to ensure a particular translation kinetics facilitating and fine-tuning co-translational protein folding in the cell via a step-wise/sequential structuring of distinct regions of the polypeptide chain emerging from the ribosome at different points in time. This review summarizes key findings in the field that have identified the role of synonymous codons and their usage in protein folding in the cell.

**Keywords:** genetic code, codon usage, synonymous codons, translation kinetics, nascent peptides, co-translational protein folding