

O vigésimo primeiro e o vigésimo segundo aminoácidos: o código genético expandido

Caetano da Costa¹, Érica Rodrigues dos Santos², Eduardo Galembeck³

¹ Pós-doutorando em Ensino de Bioquímica na Universidade Estadual de Campinas, São Paulo

² Mestranda do Programa de Pós-Graduação Multiunidades em Ensino de Ciências e Matemática (PECIM) da Universidade Estadual de Campinas, São Paulo

³ Departamento de Bioquímica e Biologia Tecidual, Instituto de Biologia, Universidade Estadual de Campinas, São Paulo

Autor para correspondência: caedacosta@yahoo.com.br

Palavras-chave: aminoácidos, selenocisteína, pirrolisina, código genético



Desde que foi decifrado, o código genético traz as especificações de 20 diferentes tipos de aminoácidos componentes das proteínas. Contudo, a descoberta de dois novos aminoácidos geneticamente codificados de maneira excepcional levou à expansão do código e do número tradicional de aminoácidos universais. A partir da revisão das principais características do processo de síntese proteica, são descritos os aspectos essenciais da síntese e da incorporação da selenocisteína (21º aminoácido) e da pirrolisina (22º aminoácido), além de suas distribuições nos domínios da vida, funções e tipos de proteínas das quais fazem parte.

As proteínas são compostos fundamentais dos seres vivos. A maioria das proteínas é formada por combinações e sequências baseadas em apenas 20 aminoácidos. Esse conjunto de aminoácidos é amplamente conhecido e os estudantes têm contato com eles desde o ensino médio. A sequência de aminoácidos tem sua origem no DNA celular, sob a forma de nucleotídeos organizados em genes. A produção de proteínas ocorre no citoplasma, e é conhecida como síntese proteica ou tradução. Portanto, a expressão gênica requer a conversão de nucleotídeos em aminoácidos.

AMINOÁCIDOS INCOMUNS

Além dos 20 aminoácidos tradicionais ou comuns, certas proteínas apresentam um ou mais de dezenas de outros aminoácidos não convencionais ou incomuns, formados por meio de modificações pós-traducionais – reações que ocorrem depois do término da tradução, o que significa dizer que tais aminoácidos modificados não estão codificados no DNA, ou seja, não apresentam correspondência de nucleotídeos no DNA. Por outro lado, esse não é o caso de dois deles que, junto com os 20 convencionais, estão codificados no DNA. Esses dois aminoácidos são a **selenocisteína** que, a partir daqui, será chamada de **SeC**, o 21º aminoácido, descoberto em 1986 e a **pirrolisina**, que será aqui identificada como **PiL**, o 22º aminoácido, descoberto em 2002.

Ambos (SeC e PiL) são formados de maneira cotraducional, sendo incorporados às proteínas quando estas ainda se encontram em formação. A descoberta desses dois novos aminoácidos evidencia que o código genético pode ser expandido e, ainda mais surpreendente, que o término da tradução não significa necessariamente o seu final. Para compreender essa aparente contradição e as demais implicações dessa descoberta, é preciso antes entender o funcionamento do complexo processo de síntese proteica.

O PAPEL DA TRANSCRIÇÃO NA EXPRESSÃO GÊNICA

A fim de que um gene seja expresso, é necessário que, no mínimo, ocorra um produto

gênico – em geral, uma proteína. O primeiro passo para que isso ocorra é um processo chamado **transcrição**, que consiste na conversão de um trecho de uma fita de DNA em um trecho de RNA. Dessa forma, a informação genética é transferida ao RNA, por meio da ação de enzimas (RNA polimerases), formando o RNA mensageiro (mRNA). Em procariotos, o mRNA está pronto para a próxima etapa, a tradução. Em eucariotos, o processo é mais complexo, e envolve etapas adicionais.

Nos organismos eucarióticos, a maior parte do DNA está contida no núcleo, local onde se processa a transcrição. Neste caso, o mRNA formado é um precursor ainda imaturo, que necessita passar por uma etapa de processamento, produzindo um mRNA maduro. Além deste, ainda no núcleo, são sintetizados, a partir do DNA, outros tipos de RNA fundamentais para a tradução: os RNAs ribossômicos (rRNAs) e os RNAs transportadores (tRNAs). Assim, os diversos tipos de RNAs atravessam a dupla membrana nuclear por meio de transportadores específicos e chegam ao citoplasma, local onde efetivamente ocorre a tradução.

O CÓDIGO GENÉTICO

Na década de 1940, comprovou-se que a informação genética era transmitida aos descendentes por meio do DNA, e não das proteínas, como se acreditava anteriormente. A forma de ocorrência dessa transmissão começou a ser elucidada na década seguinte, por meio da célebre publicação de Watson e Crick (1953) a respeito da estrutura do DNA. A descoberta do mRNA em 1960 permitiu elucidar como os quatro nucleotídeos do DNA controlavam a sequência dos 20 aminoácidos presentes nas proteínas, mas gerou uma nova questão: como a sequência dos nucleotídeos do mRNA especificam a sequência de aminoácidos em um **polipeptídeo**? A resposta foi construída ao longo da década de 1960, culminando na entrega do Prêmio Nobel de Medicina em 1968 aos pesquisadores Robert W. Holley, Har Gobind Khorana e Marshall W. Nirenberg, que decifraram o código genético.

Polipeptídeo - estrutura formada por uma sequência de resíduos de aminoácidos. Ao associar-se uma função a um polipeptídeo, podemos chamá-lo de *proteína*. Assim, um polipeptídeo faz parte de uma proteína; por exemplo, as cadeias alfa e beta da hemoglobina são polipeptídeos. Portanto, a rigor, a hemoglobina é uma proteína, e não um polipeptídeo.

Os pesquisadores demonstraram que o código genético consiste em 64 trincas ou tripletes de nucleotídeos presentes no mRNA, chamados códon (Tabela 1). Como são 20 os aminoácidos, isso significa que cada aminoácido pode ser especificado por mais de um códon. Existem ainda códon específicos

que sinalizam o início (códon de iniciação) e o término (*stop* códon ou códon de terminação) das cadeias polipeptídicas. Com pouquíssimas exceções (SZYMANSKI; BARCISZEWSKI, 2007), as especificações do código genético são as mesmas em todos os organismos, dos vírus aos seres humanos.

Tabela 1.

Trincas do código genético e os seus significados. De um total de 64 trincas, 61 codificam aminoácidos (representados pelos seus símbolos de três letras) – sendo que a trinca AUG corresponde ao códon de iniciação, que codifica Metionina (símbolo em verde); apenas três trincas correspondem a *stop* códon ou códon de parada, mas nem sempre sinalizam o final da tradução, como será discutido adiante.

		Segunda Base				
		U	C	A	G	
Primeira Base	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG } Stop	UGU } Cys UGC } UGA } Stop UGG } Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G
						Terceira Base

O PROCESSO TRADUCIONAL

A tradução consiste na interpretação do código genético. Em linhas gerais, trata-se da conversão da informação genética contida no mRNA em sequências de aminoácidos, de acordo com as instruções do código. O processo ocorre no citoplasma (Figura 1), e envolve a participação de vários componentes, sendo os principais: ribossomos, de 40 a 60 tRNAs diferentes, pelo menos 20 enzimas de ligação dos aminoácidos aos tRNAs e várias proteínas solúveis (fatores) que atuam em cada etapa do processo.

A síntese de proteínas necessita de um enorme gasto de energia, envolvendo uma intrincada maquinaria celular. Cerca de um terço da massa seca total da maioria das células compõe-se de moléculas que participam diretamente da biossíntese proteica (SNUSTAD; SIMMONS, 2012). O processo ocorre nos ribossomos, estruturas macromoleculares complexas formadas por três tipos de RNA ribossômico (rRNA) e por proteínas, em partes aproximadamente iguais de RNA e de proteína. Os ribossomos podem ser considerados “bancadas” moleculares, as quais oferecem suporte e ferramentas para acomodar o mRNA.

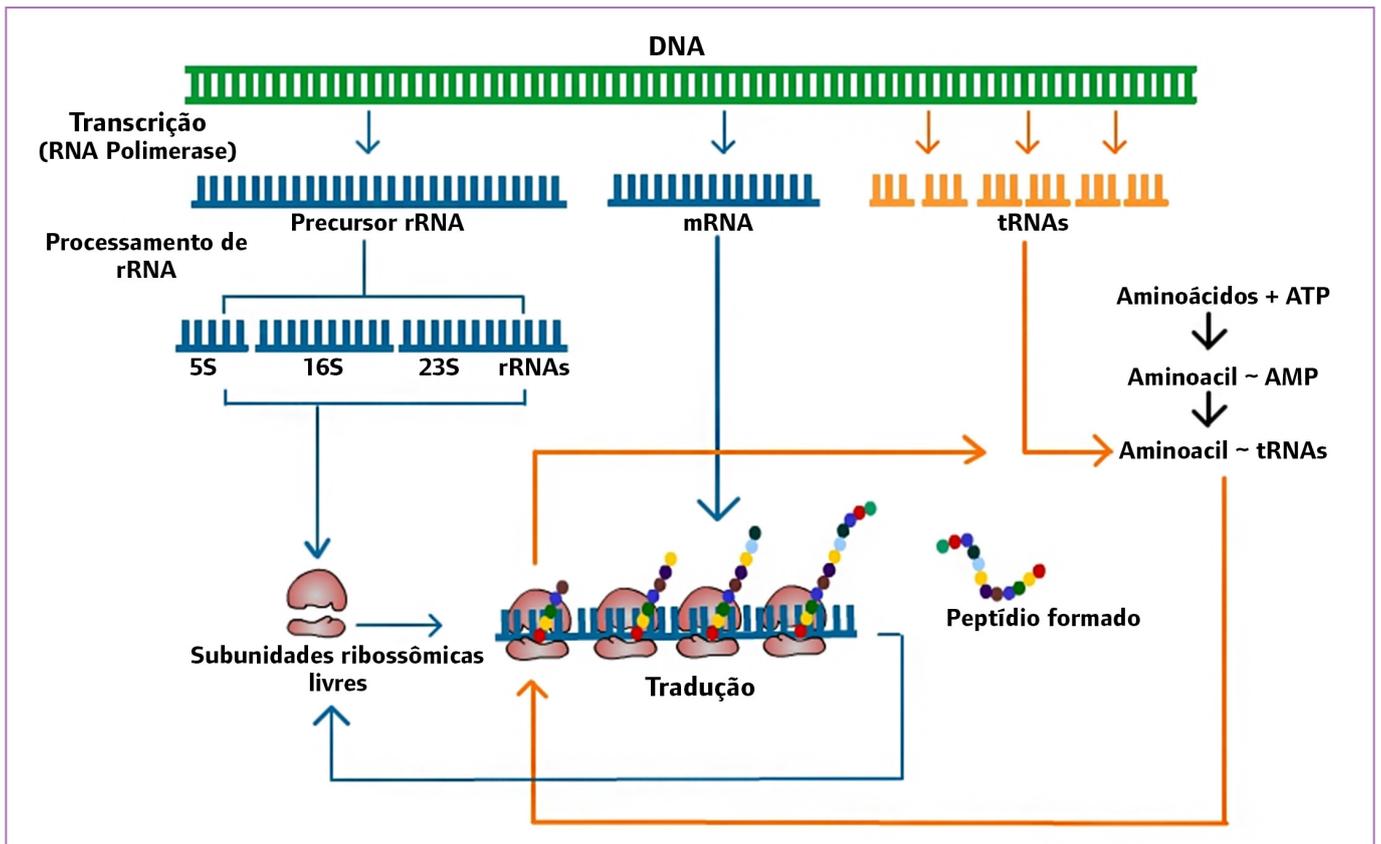


Figura 1. Visão geral da síntese proteica, mostrando os processos centrais (transcrição, processamento do mRNA e tradução) e as interações entre alguns dos principais componentes envolvidos: o DNA, os três tipos de RNA, os ribossomos e os aminoácidos (representados por pequenas esferas coloridas). [Observação: com exceção da Figura 8, todas as demais foram elaboradas por um dos autores (Santos, E.R.) com base nas imagens do capítulo 12 do livro-texto *Principles of genetics* (SNUSTAD; SIMMONS, 2012)]

OS RIBOSSOMOS E OS RNAs TRANSPORTADORES

Cada ribossomo é composto por duas subunidades, a grande e a pequena, que se associam antes de iniciar o processo de tradução. Embora apresentem variação de tamanho e de composição, a estrutura tridimensional dos ribossomos é basicamente a mesma tanto em procariotos quanto em eucariotos.

Do ponto de vista químico, já se sabia que uma interação direta entre os aminoácidos e os códons era inviável. Assim, em 1958, Francis Crick propôs que deveria existir alguma molécula que mediasse tal interação. Essas moléculas, inicialmente denominadas *adaptadoras*, foram identificadas como pequenas moléculas de RNA. Posteriormente, foram designadas RNA de transferência, ou RNA transportador (tRNA). Observe que sua estrutura é peculiar (Figura 2), com formato semelhante a uma cruz, contendo braços e alças (*loops*). Na extremidade de

um dos braços, cada tRNA contém uma trinca de nucleotídeos chamada *anticódon*, que é complementar à sequência do códon no mRNA. Dessa forma, permite um pareamento perfeito que medeia o acoplamento de certo aminoácido ao seu códon específico. Para cada um dos 20 aminoácidos, existem de 1 a 4 tRNAs.

Para que o tRNA execute a sua função, ele deve ligar-se a um aminoácido específico. Em outras palavras, o tRNA deve ser carregado com um aminoácido. Para tanto, deve ocorrer uma reação enzimática em duas etapas, catalisadas pelas mesmas enzimas, as aminoacil-tRNA sintetases, específicas para cada um dos 20 aminoácidos. A primeira etapa envolve gasto de energia sob a forma de ATP, para que ocorra a ativação do aminoácido; a seguir, o aminoácido ativado é ligado ao seu tRNA, formando os substratos para a síntese proteica nos ribossomos: os aminoacil-tRNAs. Por exemplo, se o aminoácido ativado é a alanina, forma-se alanil-tRNA.

Loops ou alças - regiões de formato circular que fazem parte da extremidade de dois braços do tRNA.

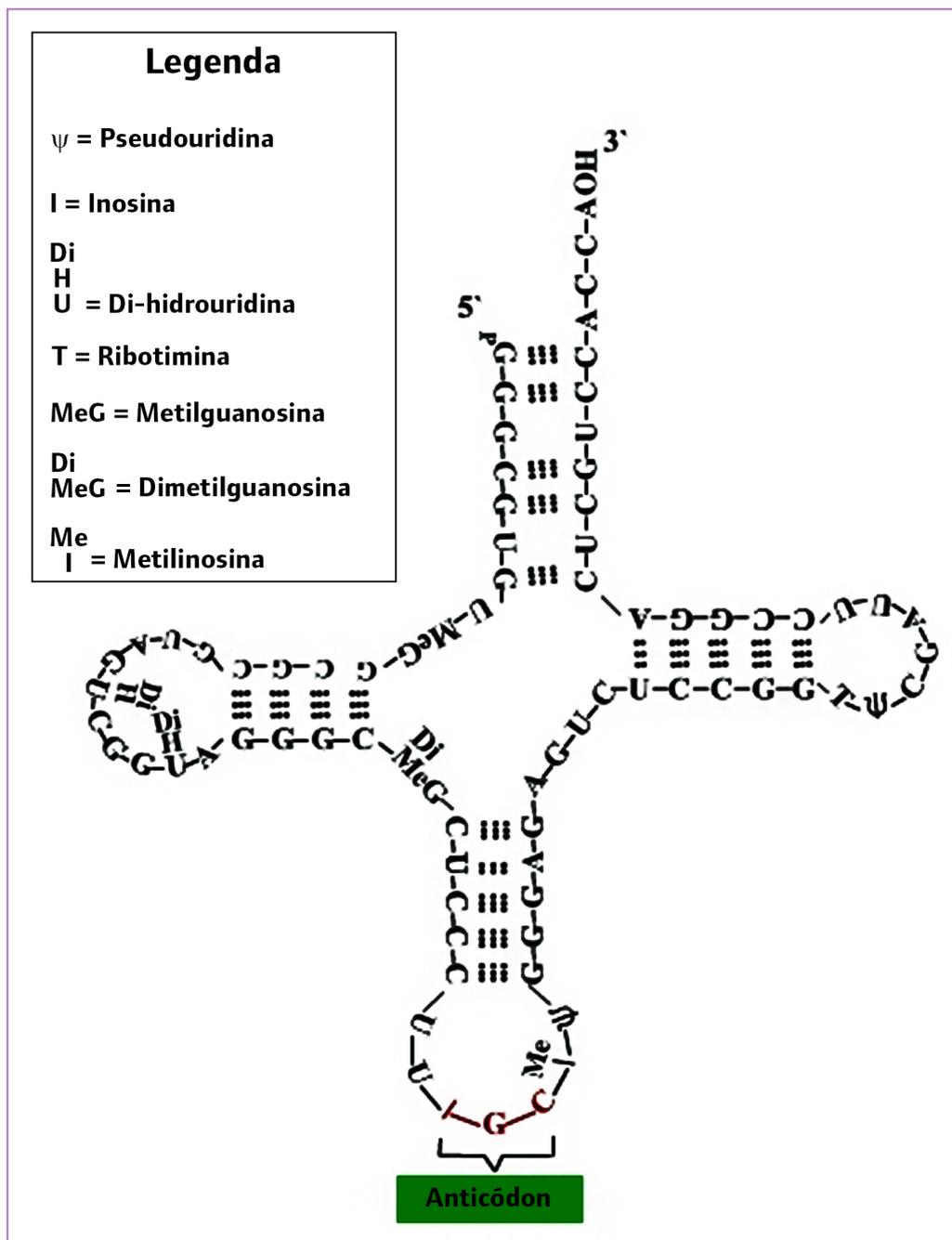


Figura 2. Estrutura de um tRNA, destacando a composição de nucleotídeos incomuns (vide legenda) e a localização do anticódon.

ORF - segmento de DNA contendo a sequência necessária à codificação de uma proteína.

Cada aminoacil-tRNA pode se ligar ao ribossomo por meio de três locais ou sítios de ligação (ver Figura 3): o **sítio A** (ou “aminoacil”) liga o aminoacil-tRNA que está por chegar, carregando o próximo aminoácido a ser incorporado à cadeia polipeptídica em formação; o **sítio P** (ou “peptídil”) liga o tRNA ao qual o polipeptídeo em formação está ancorado; e o **sítio E** (ou “de saída”) liga o tRNA que descarregou o aminoácido e que está prestes a deixar o ribossomo.

OS ESTÁGIOS DA TRADUÇÃO

O mRNA a ser traduzido consiste em uma sequência de códon: um de iniciação, seguido pelos que especificam aminoácidos, e um de terminação. Cada mRNA é derivado de uma outra sequência chamada de **ORF** (do inglês *Open Reading Frame*) ou quadro aberto de leitura. Presume-se que cada ORF codifique uma proteína.

O processo traducional pode ser dividido em três estágios: iniciação, alongamento e termi-

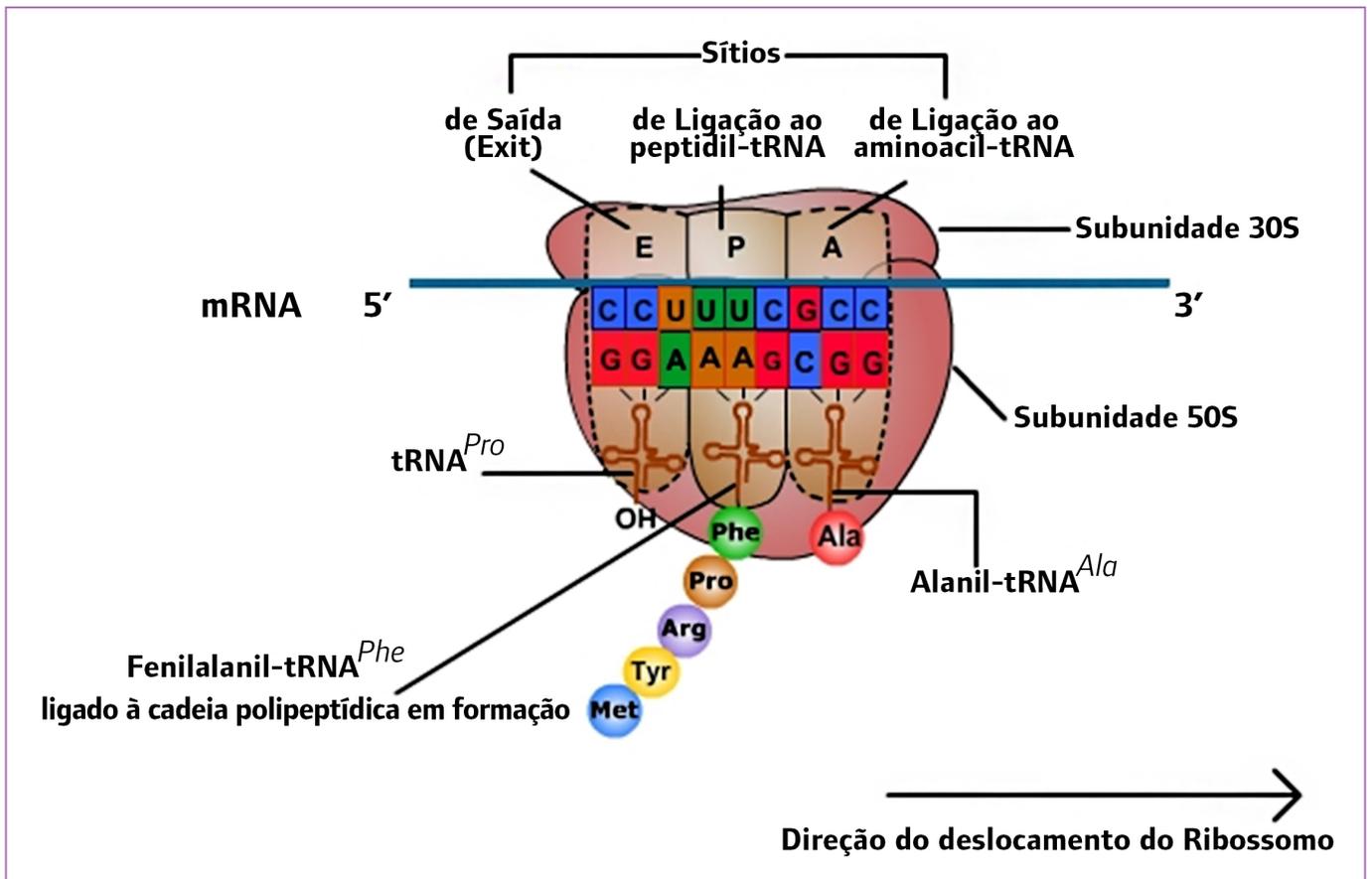


Figura 3. Detalhes da estrutura do ribossomo ligado ao mRNA e aos tRNAs, durante o processo de síntese proteica, destacando suas subunidades (30S e 50S) e os três sítios de ligação envolvidos (sítios E, P e A). Os pequenos retângulos coloridos representam os códons (sequência superior) e os anticódons (sequência inferior). Os aminoácidos estão representados por esferas coloridas, cada qual com símbolos de três letras (Ala = alanina, Phe = fenilalanina, Pro = prolina, Arg = arginina, Tyr = tirosina, Met = metionina).

nação. Em cada um deles, atuam diversas proteínas solúveis que participam como fatores de apoio. Serão priorizados os aspectos de cada estágio que sirvam para ilustrar os propósitos deste artigo, ou seja, para a compreensão da síntese dos aminoácidos SeC e PiL. Por essa razão, não serão exploradas certas diferenças da tradução entre procariotos e eucariotos.

A síntese é iniciada por um tRNA especial, carregando o aminoácido metionina, que responde ao códon de iniciação (geralmente AUG). Assim, no sítio P do ribossomo, o anticódon UAC emparelha-se com o códon AUG do mRNA, sinalizando o início da tradução (Figura 4).

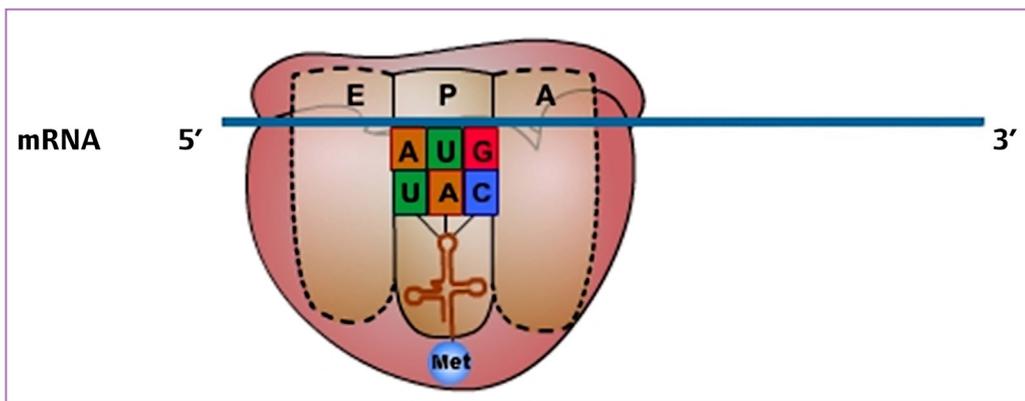


Figura 4. Configuração inicial da tradução, sinalizada pelo posicionamento do tRNA carregando a metionina no sítio P. São destacados apenas o códon de iniciação (AUG) e o seu anticódon (UAC) correspondente.

Resíduos de aminoácidos

- nome dado aos aminoácidos quando incorporados em uma proteína, devido à exclusão de uma molécula de água na reação de formação de cada ligação peptídica.

A partir daí, ocorre o estágio de alongamento, com a adição sucessiva dos aminoácidos. Cada aminoácido é adicionado em três etapas: (1) um aminoacil-tRNA liga-se ao sítio A do ribossomo; (2) a cadeia polipeptídica em crescimento é transferida do tRNA no sítio P para o tRNA que acabou de se ligar ao sítio A, formando uma nova ligação peptídica; (3) o ribossomo move-se ao longo do mRNA (translocação), de forma a posicionar o próximo códon no sítio A. Dessa maneira, durante a etapa (3), o tRNA contendo o polipeptídeo nascente e o tRNA que libe-

rou o aminoácido são deslocados dos sítios A e P para os sítios P e E, respectivamente. Portanto, logo após as três etapas, com a translocação do ribossomo, tem-se: no sítio E, o tRNA que logo será liberado; no sítio P, o tRNA com a sequência de **resíduos de aminoácidos** em crescimento; e o sítio A livre, pronto para receber o próximo aminoacil-tRNA (Figura 5). Assim, iniciam-se novamente as três etapas, consecutivamente, até atingir o estágio de terminação, em que um dos três códons de terminação (UAA, UAG ou UGA) ocupa o sítio A.

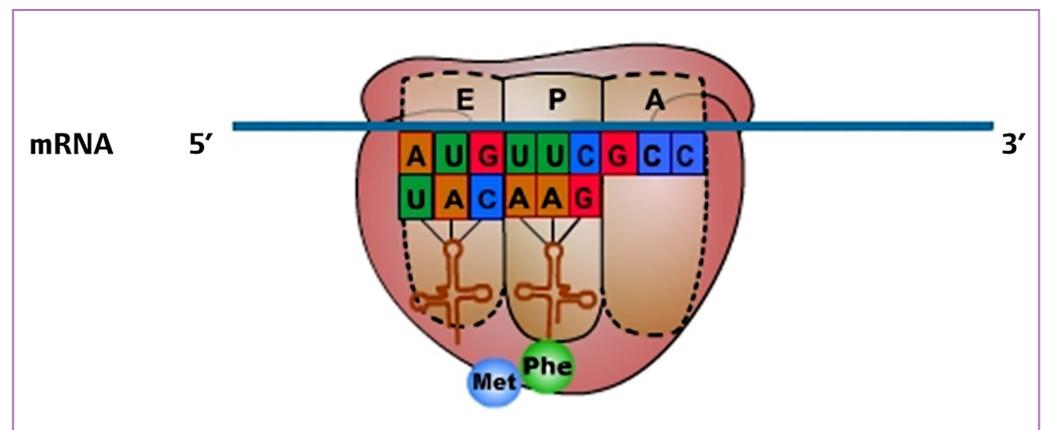


Figura 5.

Etapa do alongamento da cadeia polipeptídica, mostrando a cadeia peptídica já acrescida do aminoácido Phe (fenilalanina) ocupando o sítio P. Para realizar a adição de Phe, ocorreram as seguintes etapas sequenciais: i) seu aminoacil-tRNA (fenilalanil-tRNA) ligou-se ao sítio A; ii) a Met já ligada no sítio P foi transferida para o sítio A e ligada à Phe, formando o dipeptídeo fenilmetionina; iii) o ribossomo movimentou-se no mRNA, de modo que os tRNAs deslocaram-se para os sítios P e E, deixando o sítio A livre para receber o próximo aminoacil-tRNA.

Conforme descrito, pode-se considerar que as reais tradutoras do código genético são as aminoacil-tRNA sintetases, uma vez que são essas enzimas que promovem a ligação direta entre o anticódon e o aminoácido ativado ligado ao seu tRNA específico. Uma vez que se atinja um dos três códons de terminação, qualquer um deles é capaz de sinalizar o final da tradução. O processo realmente chega ao fim, uma vez que ocorre a desmontagem de toda a maquinaria: o polipeptídeo formado e a molécula de mRNA são liberados do ribossomo. Além disso, as subunidades maior e menor do ribossomo dissociam-se, e o mRNA é degradado.

Todos os 20 aminoácidos convencionais são adicionados à cadeia polipeptídica por meio dos estágios descritos. A seguir, será detalhado o que ocorre no caso da SeC e da PiL.

SÍNTESE E TRADUÇÃO DA SELENOCISTEÍNA (SEC)

A síntese e a incorporação da SeC são bem compreendidas em *E. coli*, e as considerações aqui descritas são válidas para este modelo bacteriano. A princípio, pensou-se que a SeC fosse produto de uma modificação pós-traducional, mas tal hipótese foi aos poucos perdendo força.

A SeC é sintetizada a partir do aminoácido convencional serina. Como ocorre com qualquer aminoácido tradicional, primeiro a serina é ativada e, em seguida, ligada ao tRNA. No entanto, aqui começam as particularidades – nos organismos que utilizam a SeC, existem dois tipos de tRNA que ligam serina: um que reconhece e é responsável pela incorporação de serina à cadeia polipeptídica

(o qual chamaremos tRNA^{Ser}), e outro, presente em concentrações menores, que é capaz de reconhecer apenas o códon UGA, que será denominado tRNA^{Ser(SeC)}. Este é um tRNA com características peculiares – além de reconhecer um único códon de terminação, difere dos demais tRNAs por ter maior comprimento e sofrer poucas modificações pós-transcricionais.

Portanto, a síntese de SeC ocorre no próprio tRNA^{Ser(SeC)}. Primeiro, a serina é ativada e ligada ao tRNA^{Ser(SeC)} pela seril-tRNA sintetase. O seril-tRNA^{Ser(SeC)} formado é usado como substrato para a síntese de SeC, da seguinte forma: a enzima selenocisteína sintase converte o seril-tRNA^{Ser(SeC)} em selenocisteil-tRNA^{Ser(SeC)}, em uma reação de duas etapas; em uma delas, há a participação de seleno-fosfato como doador de selênio, sintetizado às custas de ATP. Dessa forma, a inserção da SeC é garantida pelo contexto de supressão

do códon de parada (SZYMANSKI; BARCISZEWSKI, 2007), que permite a incorporação da SeC e a continuidade da tradução. Tal contexto pode ser caracterizado da seguinte forma: no mRNA que contém o códon de parada que codifica a SeC, existe uma sequência específica de inserção que flanqueia (marca) o referido códon de forma que seja reconhecido não como um sinal de término, mas como a especificação de um novo aminoácido a ser adicionado. Tal sequência é denominada SECIS – *SE*leno*Cys*teine *I*nsertion *S*equen*C*e. Outro importante elemento contextual para a incorporação da SeC é a participação de fatores específicos de alongamento – um deles é especial e auxilia a formação de um complexo entre o selenocisteil-tRNA^{Ser(SeC)} e a SECIS, o qual permite a inserção do tRNA carregado com a SeC no sítio A do ribossomo (ROTHER; KRZYCKI, 2010) – vide Figura 6.

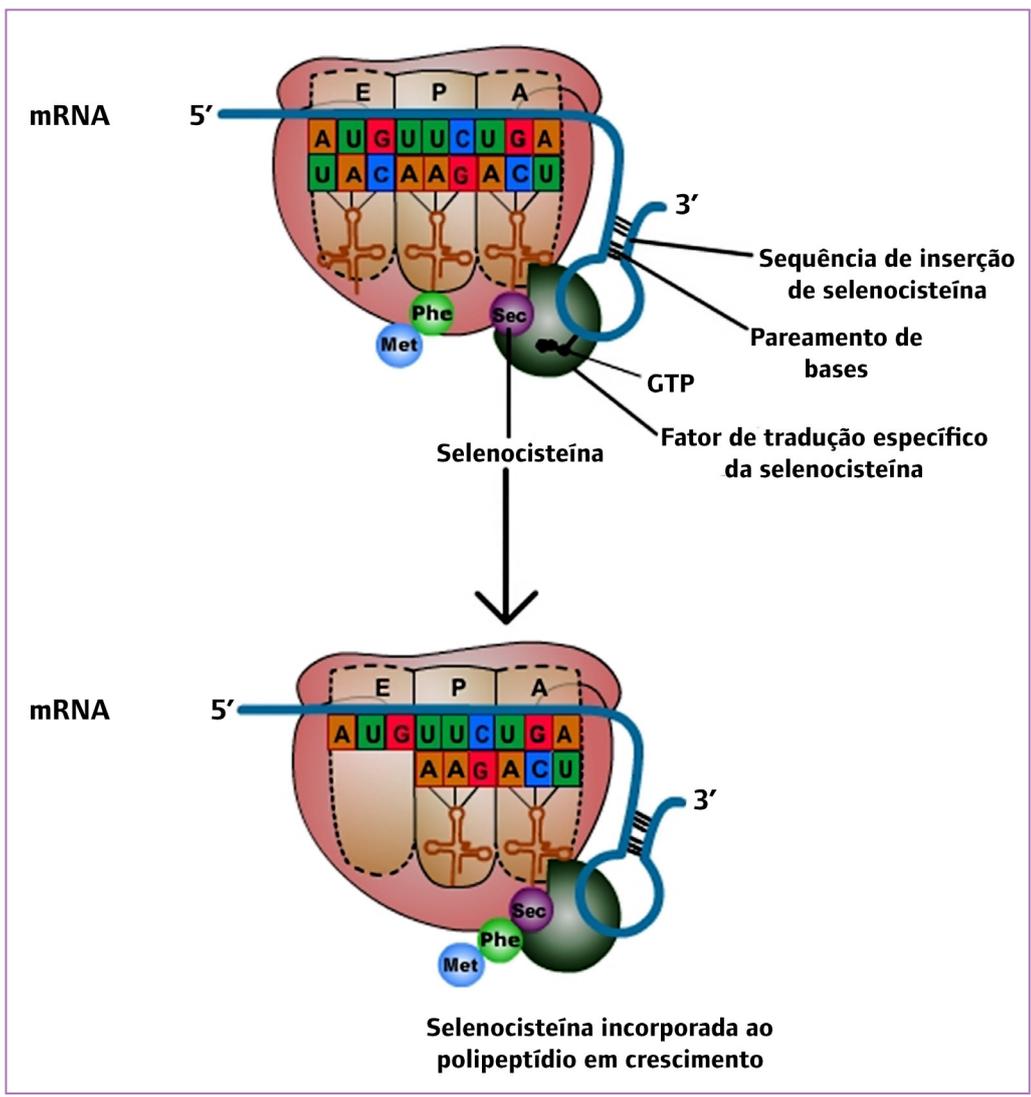


Figura 6. Incorporação da selenocisteína à cadeia polipeptídica em resposta à SECIS. O esquema superior representa a ligação do selenocisteil-tRNA ao sítio A, e destaca os elementos contextuais que permitem que o códon de terminação especifique a inserção da SeC (a esfera maior representa o fator de alongamento específico). No esquema inferior, observa-se a SeC (pequena esfera roxa) já incorporada à cadeia peptídica. A seguir, ocorre o deslocamento do ribossomo, deixando o sítio A livre para a próxima aminoacil-tRNA (vide Figura 5).

É interessante notar que, do ponto de vista estrutural, a SeC difere da cisteína unicamente pela presença de um átomo de selênio no lugar de um átomo de enxofre (Figura 7).

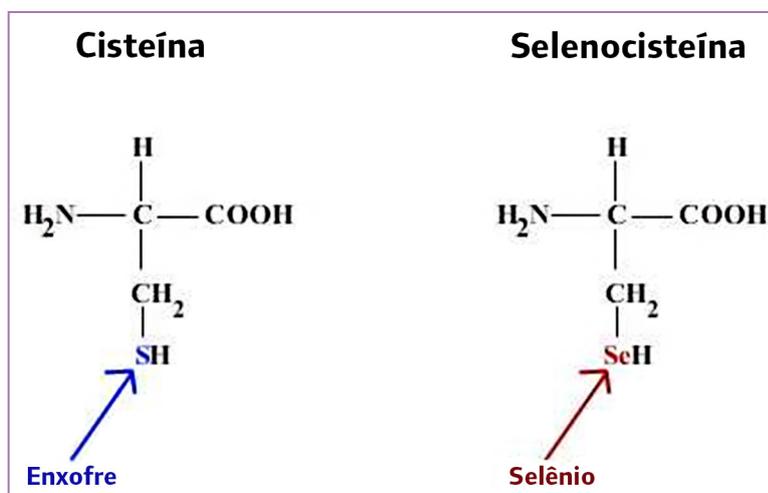


Figura 7. Estruturas da cisteína e da SeC, destacando o elemento diferencial entre as duas.

SÍNTESE E TRADUÇÃO DA PIRROLISINA (PiL)

Ao contrário da SeC, a PiL é sintetizada como um aminoácido livre, a partir de duas moléculas de lisina. Na verdade, trata-se de uma síntese em quatro etapas (RAGSDALE, 2011), ilustradas na Figura 8: a primeira consiste em uma reação enzimática que converte lisina em metil-ornitina, com participação da coenzima S-adenosilmetionina (SAM) e da enzima PylB. Uma outra enzima (PylC) catalisa a condensação da metil-ornitina com uma nova molécula de lisina, às custas de ATP, formando metil-ornitil-lisina. Este composto, por sua vez, é convertido em pirrolisina em uma complexa reação que ocorre em duas etapas – tal reação é catalisada pela enzima PylD, uma desidrogenase cuja atividade envolve desaminação (perda de NH₃), desidratação e ciclização (BORREL *et al.*, 2014). Assim, a PiL é sintetizada antes da ligação ao seu tRNA específico (tRNA^{PiL}). Portanto, existe uma pirrolisil-tRNA sintetase responsável pela ativação e ligação da PiL ao tRNA^{PiL}. Este é um RNA especial, diferindo dos demais por

ter o braço do anticódon um pouco mais longo (1 nucleotídeo a mais), além de apresentar outras partes mais curtas, com ausência de algumas bases nitrogenadas.

Assim como no caso da SeC, o mRNA contendo o códon de parada para PiL (UAG) apresenta uma sequência marcadora, a *PYLIS – PYrroLysine Insertion Sequence*, que garante a incorporação da PiL e a continuidade do processo traducional.

Em suma, quando o mRNA não apresenta sequências como *SECIS* ou *PYLIS*, os códons de terminação sinalizam de fato o fim da tradução.

SEC E PiL: DISTRIBUIÇÃO NOS DOMÍNIOS DA VIDA E FUNÇÕES

A SeC apresenta ampla ocorrência, sendo encontrada nos três domínios (**Arqueia**, Bactéria e Eucaria). Contudo, tal distribuição não é universal, pois estão ausentes em fungos, plantas vasculares e certos insetos, como besouros e bichos da seda (LABUNSKYY *et al.*, 2014).

Arqueia - um dos três domínios da vida, que compreende procariontes primitivos com características próprias, diferentes das bactérias comuns (alocadas no domínio Bactéria); o outro domínio é o Eucaria, que abrange os seres eucariontes (dos organismos microscópicos às plantas e animais).

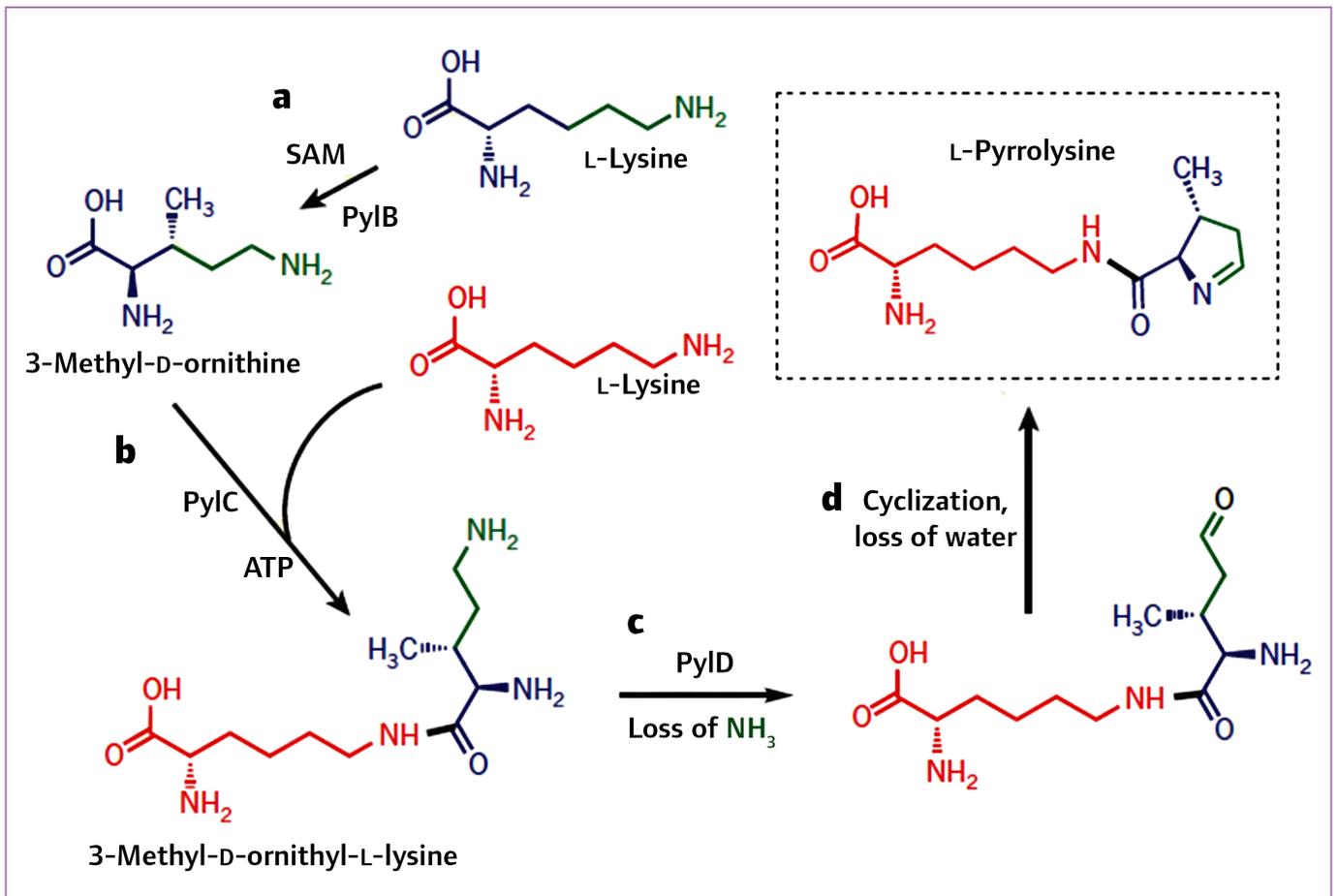


Figura 8.

Etapas propostas para a biossíntese da pirrolisina [Reproduzida com permissão da Macmillan Publishers Ltd: NATURE, Ragsdale, S.W. How two amino acids become one, v. 471, n. 7340, p. 583, copyright 2011]. Consistem em quatro reações enzimáticas sequenciais, com participação de duas lisinas e gasto de ATP (detalhes no texto).

Sítio ativo - região de uma enzima que acomoda o substrato e é responsável pela conversão do substrato em produto.

Apoptose - processo de morte celular programada que se inicia com sinalização mitocondrial.

Isoenzimas - enzimas que catalisam a mesma reação, e são expressas em locais intracelulares (compartimentos subcelulares) diferentes ou em tecidos diferentes.

Proteínas que contêm a SeC são coletivamente conhecidas como selenoproteínas, as quais exibem dezenas de tipos diferentes e distribuem-se em mais de 50 famílias identificadas, mas nem todas com função conhecida. São três as famílias bem estudadas (LABUNSKYY *et al.*, 2014; LU; HOLMGREN, 2009), nas quais a SeC está presente no **sítio ativo** de cada membro: i) glutatona peroxidases; ii) iodotironina deiodinases; e iii) tiorredoxina redutases. Em comum, as três famílias caracterizam-se por apresentar enzimas com diferentes localizações subcelulares e distribuição tecidual.

No caso das glutatona peroxidases, existem pelo menos cinco **isoenzimas** em seres humanos, que participam do sistema de defesa antioxidante das células e de vias

de sinalização que envolvem o peróxido de hidrogênio (H₂O₂). Já as iodotironina deiodinases participam da regulação da atividade dos hormônios tireoidianos, sendo que existem três isoenzimas conhecidas. Quanto às tiorredoxina redutases, são também três isoenzimas identificadas em mamíferos, envolvidas em vários processos fisiológicos, tais como a defesa antioxidante, a regulação de fatores de transcrição e a **apoptose**. É válido mencionar que a tiorredoxina redutase é a única enzima capaz de reativar a proteína tiorredoxina (LU; HOLMGREN, 2009), cujo papel é crucial para a síntese de DNA, pois mantém funcional a ribonucleotídeo redutase, enzima que converte ribonucleotídeos em sua forma desoxigenada.

Além das três famílias descritas, merece destaque a selenoproteína P, uma outra família que apresenta vários resíduos de SeC (10 ao todo) por molécula proteica. Essas selenoproteínas circulam no plasma humano e têm como principal função transportar e distribuir selênio aos tecidos (LU; HOLMGREN, 2009).

Em geral, outras funções atribuídas às selenoproteínas são: reparo de proteínas que sofreram dano oxidativo, controle da montagem do citoesqueleto celular e apoio ao enovelamento proteico (LABUNSKYY *et al.*, 2014). Recentemente, as selenoproteínas têm sido implicadas na diferenciação e função de enterócitos e adipócitos, além de um possível papel na secreção e ação da insulina (STEINBRENNER *et al.*, 2016).

Ao contrário do que ocorre com os 20 aminoácidos tradicionais, não existe um *pool* de SeC livre nas células e, no catabolismo proteico, a SeC é completamente degradada. Tudo isso ocorre basicamente devido à alta reatividade deste aminoácido, que poderia ser incorporado no lugar da cisteína ou gerar espécies reativas de oxigênio (LU; HOLMGREN, 2009).

Pool - quando associada a aminoácidos, essa palavra designa o conjunto de aminoácidos livres no interior das células.

Em contraste com a diversidade funcional e de ocorrência da SeC, a PiL tem sua distribuição restrita a certas arqueias metanogênicas e a um número bastante reduzido de bactérias. É componente de pouquíssimas proteínas, as quais atuam como enzimas do tipo metiltransferases, responsáveis pela síntese de metano a partir de metanol e/ou metilaminas (BORREL *et al.*, 2014).

Apenas como curiosidade, são conhecidos dois organismos (bactérias) que utilizam ambos os aminoácidos SeC e PiL: a *Desulfotobacterium hafniense* e a deltaproteobactéria simbiótica.

CONSIDERAÇÕES FINAIS

A descoberta da SeC e da PiL levou, no mínimo, a rever o conjunto dos 20 aminoácidos codificados. Embora representados por trinças nucleotídicas que normalmente funcionam como códons de parada, e assim estejam indiretamente codificados, é comum atribuir à SeC e à PiL as designações “21º” e “22º” aminoácidos, respectivamente, de acordo com a ordem cronológica de descoberta. Dessa maneira, pode-se considerar que o conjunto de aminoácidos universais ou natu-



rais contém 22 tipos diferentes, configurando o que a literatura especializada chama de expansão do código genético.

Além disso, os dados revelados pelo estudo dos dois novos aminoácidos indicam duas quebras de paradigma em relação ao processo de síntese proteica:

- 1) um códon de parada pode significar continuidade e não término – isso significa que uma proteína é produzida a partir de dois ORFs (normalmente, cada ORF corresponde no mRNA maduro a um trecho que inicia com um códon de início, seguido por uma série de códons até chegar ao códon de parada – ou seja, é esperado que uma ORF origine uma proteína);
- 2) a síntese de um aminoácido (SeC) pode ocorrer no próprio tRNA, sem envolver uma aminoacil-tRNA sintetase específica – portanto, sem necessidade de ativação e ligação daquele aminoácido.

Além dessas subversões apontadas, há outro aspecto da descoberta do 21º aminoácido que merece destaque: a necessidade de selênio como um elemento essencial. Ainda hoje, uma ideia comum entre os estudantes é que os principais elementos da matéria viva são carbono, hidrogênio, oxigênio, nitrogênio, fósforo e enxofre, memorizados como “CHONPS”, com as iniciais de cada elemento. Com a SeC, há muito tempo tal regra mnemônica deveria ser acrescida do selênio como um elemento fundamental para os seres vivos.

Como reflexão final, levanta-se a seguinte questão: uma vez que são três diferentes códons de terminação, haveria espaço para um vigésimo terceiro aminoácido no código genético? Aparentemente, a resposta é negativa. Análises computacionais baseadas nos três códons de parada, visando à busca e identificação de possíveis tRNAs que levassem a um 23º aminoácido, foram infrutíferas (LOBANOV *et al.*, 2006). Estes autores também não encontraram tRNAs atípicos (não vinculados a códons de parada) e concluem que, se o 23º aminoácido realmente existir, é provável que tenha uma distribuição extremamente restrita.

AGRADECIMENTOS

Agradecemos ao Prof. Bayardo Baptista Torres (IQ-USP) pela leitura criteriosa do manuscrito e sugestões que contribuíram para a clareza e precisão em várias partes do texto.

REFERÊNCIAS

- BORREL, G.; GACI, N.; PEYRET, P.; O'TOOLE, P. W.; GRIBALDO, S.; BRUGÈRE, J. F. Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette. *Archaea*, v. 2014, 2014. Disponível em: <file:///C:/Users/user/Downloads/ARCHAEA2014-374146.pdf>. Acesso em 03/07/2015.
- LABUNSKYY, V. M.; HATFIELD, D. L.; GLADYSHEV, V. N. Selenoproteins: molecular pathways and physiological roles. *Physiological Reviews*, v. 94, n. 3, p. 739-777, 2014.
- LOBANOV, A. V.; KRYUKOV, G. V.; HATFIELD, D. L.; GLADYSHEV, V. N. Is there a twenty third amino acid in the genetic code? *Trends in Genetics*, v. 22, n. 7, p. 357-360, 2006.
- LU, J.; HOLMGREN, A. Selenoproteins. *Journal of Biological Chemistry*, v. 284, n. 2, p. 723-727, 2009.
- RAGSDALE, S. W. How two amino acids become one. *Nature*, v. 471, n. 7340, p. 583-584, 2011.
- ROTHER, M.; KRZYCKI, J. A. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea*, v. 2010, 2010. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2933860/pdf/ARCH2010-453642.pdf> Acesso em 03/07/2015.
- SNUSTAD, D. P.; SIMMONS, M. J. Translation and the genetic code. In: SNUSTAD, D. P.; SIMMONS, M. J. *Principles of genetics*. 6th ed. Hoboken, NJ: John Wiley & Sons, 2012. p 285-319.
- STEINBRENNER, H.; SPECKMANN, B.; KLOTZ, L. O. Selenoproteins: antioxidant selenoenzymes and beyond. *Archives of Biochemistry and Biophysics*, v. 595, p. 113-119, 2016.

