

TEXTO 6

O CÓDIGO GENÉTICO

O Código Genético

O Sistema de Codificação Genética

A Natureza Tríplice do Código

A Decifração dos Códon

Colinearidade entre DNA e Polipeptídeo

Os Códon de Pontuação

O Código Genético é “Quase” Universal

As Consequências das Mutações

As consequências das mutações em microorganismos

As consequências das mutações nos organismos pluricelulares

Exercícios

Parte A: Revendo Conceitos Básicos

Parte B: Ligando Conceitos e Fatos

Parte C: Aplicando Conceitos

Parte D: Resolvendo Problemas

Apresentação

Se um gene é uma sequência de nucleotídeos no DNA, como essa sequência pode determinar a ordem dos aminoácidos nas proteínas? Até a década de 60, isso foi um mistério. Nessa época, no entanto, os biólogos conseguiram a incrível proeza de decifrar a linguagem da vida, um dos maiores feitos da Biologia do século XX. São as principais etapas do trabalho científico que culminou com a decifração do código genético que estudaremos nesse texto.

Essa apostila foi organizada pelos docentes do Instituto de Biociências da USP que ministraram e ministram as disciplinas “Biologia Molecular e de Microrganismos”, “Biologia Molecular” e “Fundamentos de Biologia Molecular” para os cursos de Bacharelado e Licenciatura em Ciências Biológicas. O texto foi adaptado e compilado de diversos livros didáticos e materiais instrucionais.

O Código Genético

O Sistema de Codificação Genética

Código genético é o conjunto de regras que governa a relação entre a sequência de nucleotídeos no DNA e a ordem dos aminoácidos nas proteínas. É nessa relação que reside a essência da expressão gênica e a decifração do código genético foi realmente um importante marco na história da Biologia.

Código é definido como “um sistema de símbolos (letras, números ou palavras) usado para representar um certo significado pré-estabelecido, às vezes secreto”. Por exemplo, o código Morse internacional é constituído por pontos, barras e espaços. Para quem não conhece o código, os pontos e barras não têm o menor significado, mas conhecendo-se as regras do código, os pontos e barras podem ser facilmente convertidos em frases.

O código genético permite à célula converter sequências de bases do DNA em cadeias polipeptídicas.

Existem 20 diferentes aminoácidos que devem ser especificados durante o processo de síntese de proteínas por algum tipo de combinação das quatro diferentes bases existentes no DNA. Quantas bases devem ser utilizadas para construir um código que permita que os 20 aminoácidos diferentes sejam especificados?

Se as palavras desse código tivessem uma base, apenas quatro aminoácidos seriam especificados. Um código baseado em conjuntos de duas bases permitiria 16 combinações diferentes. Já um código baseado em trincas de bases permitiria 64 combinações distintas de bases, o que seria suficiente para determinar os 20 aminoácidos. Como os sistemas celulares são geralmente econômicos, um código baseado em quatro letras seria muito improvável, pois permitiria 256 combinações diferentes. Assim, um código baseado em trincas seria a situação mais razoável. De fato, atribui-se ao físico russo George Gamow, por volta de 1954, a proposição de que um código baseado em três bases seria o mais provável.

Contudo, ainda fica a seguinte pergunta: Como são lidas essas trincas? Uma após a outra ou uma mesma base pode fazer parte de três trincas adjacentes? Nesta segunda opção diríamos que o código é sobreposto. (Fig.1)

Em um código sobreposto, cada base do DNA participaria da codificação de dois ou três aminoácidos. Este tipo de código imporiria severas restrições sobre as sequências de aminoácidos das proteínas. Assim, por exemplo, um aminoácido cujo código fosse CAG não poderia ser seguido por um outro cujo código tivesse como primeira base um T ou um C porque, por definição, a primeira base do próximo aminoácido teria que ser A ou G, dependendo da extensão da sobreposição.

Brenner foi o primeiro a propor que o código não era sobreposto, isto é, que cada base participaria na codificação de um único aminoácido.

Também devemos considerar que se o código fosse sobreposto, toda vez que uma base fosse alterada na molécula do DNA, um, dois ou três aminoácidos seriam alterados na proteína mutada. A evidência genética, contudo, não falava a favor dessa hipótese. Em estudos sobre mutações no vírus do mosaico do tabaco (TMV), a substituição

de um único nucleotídeo no genoma viral resultava geralmente na alteração de um único aminoácido na cápsula protéica do vírus, e não na alteração simultaneamente de três aminoácidos, como seria esperado se o nucleotídeo alterado fizesse parte de três códons ao mesmo tempo.

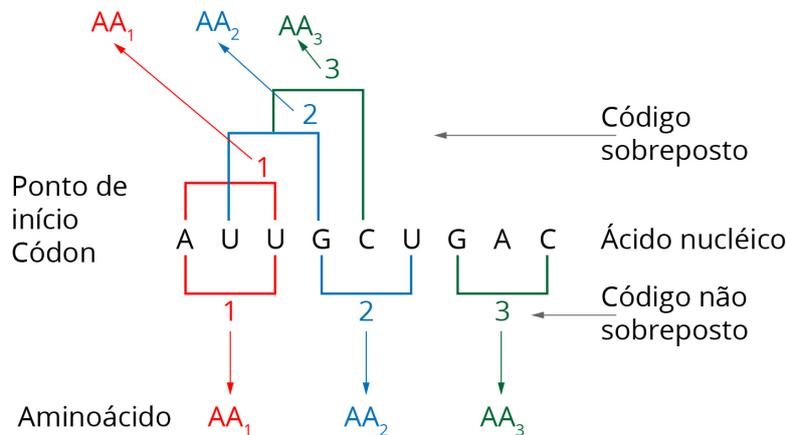


Figura 1. Comparação entre códigos de três letras sobreposto e não-sobreposto. No código não-sobreposto, a proteína é traduzida por códons que não compartilham nenhum nucleotídeo. Já no código sobreposto, cada códon compartilha um ou mais nucleotídeos, dependendo do grau de sobreposição, com os códons vizinhos.

A Natureza Tríplice do Código

Embora um código baseado em trinca fosse o mais adequado nos sistemas biológicos, apresentar uma evidência experimental de que isso era real não era uma tarefa trivial na década de 60. O trabalho de Benzer sobre mapeamento genético em fagos foi crucial no desenvolvimento desses experimentos. Benzer estudava mutantes chamados rII do fago T4. Mutantes rII eram capazes de lisar *E. coli* de uma linhagem B, mas eram incapazes de lisar *E. coli* das linhagens K. Ele verificou que os fagos com mutação de ponto revertiam muito frequentemente para o tipo selvagem. Ele poderia ter pensado que este fato seria devido à reversão da mutação original que resultaria novamente na sequência de bases do tipo selvagem. Isso é verdadeiro para a maioria das reversões, no entanto, mutações reversas não constituem a única maneira pela qual indivíduos com mutações de ponto no loco rII podem reverter ao tipo selvagem. Foi verificado que, em alguns casos, os revertidos carregavam uma segunda mutação ocorrida nas proximidades da mutação original. As duas mutações juntas eram responsáveis pela produção de um fenótipo selvagem, isto é, a segunda mutação suprimia o efeito da primeira e foi, por este fato, denominada mutação **supressora**. Os fagos que apresentavam as duas mutações juntas foram apelidados de **pseudoselvagens**. Mas como Benzer deduziu que os pseudoselvagens eram geneticamente diferentes dos selvagens? Foi verificado que, do cruzamento entre fagos selvagens e pseudoselvagens, resultavam, em elevada frequência, descendentes que exibiam o genótipo característico dos mutantes rII. Observe na figura 2 abaixo como foi possível obter mutantes rII por permutação, a partir do cruzamento entre linhagens selvagens e pseudoselvagens.

Crick e Brenner, em 1961, aproveitaram essas mutações supressoras para provar que o código genético era realmente composto por trinca, oito anos após a idéia ter sido proposta. Eles foram capazes também de demonstrar que a sequência de bases do DNA era lida a partir de um determinado ponto, trinca por trinca, de maneira não sobreposta. Isto significava que cada mensagem genética tinha uma estrutura de leitura fixa que poderia ser alterada pela adição ou remoção de uma única base. O polipeptídeo especificado pela mensagem conteria aminoácidos incorretos a partir do ponto em que a base foi adicionada ou removida. Um aspecto que deve ser ressaltado é o fato destas mutações terem sido induzidas por corantes do tipo acridina, que causam inserção ou deficiência de bases na sequência de nucleotídeos do DNA.

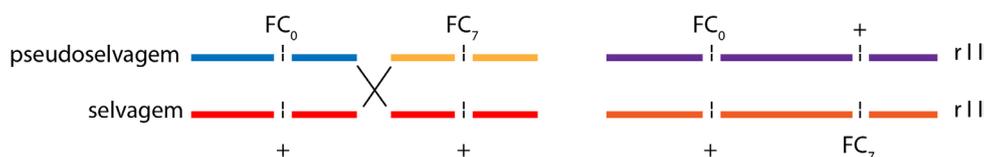


Figura 2. Cruzando-se um fago pseudoselvagem com um selvagem foi possível obter dois mutantes rII com uma única mutação cada, FC₀ e FC₇.

Crick e Brenner assumiram que se uma mutação fosse uma inserção de apenas uma base no DNA da região rII (representada por um sinal +), a sua supressora deveria ser uma deleção (representada por sinal -). Assim, se a sequência de bases do DNA estivesse sendo lida trinca por trinca (da esquerda para a direita), começando de um ponto determinado, a inserção de uma base determina que todas as trincas à direita da mutação fossem lidas de maneira diferente do original; esta seria, portanto, uma mutação do tipo **mudança do quadro de leitura**. Geralmente, isto resultaria num polipeptídeo não funcional. Para compreender melhor os efeitos de uma mutação que afeta o quadro de leitura, observe esta frase em inglês constituída por palavras de apenas três letras:

THE FAT CAT ATE THE BIG RAT

Deletando C:

C
THE FAT ATA TET HEB IGR AT

Inserindo A:

A
THE FAT ATA ATE THE BIG RAT

A inserção suprime o efeito da deleção pois restaura a fase de leitura. Mas sozinha, uma inserção também perturbaria completamente o sentido da frase:

Inserindo A:

A
THE FAT CAT AAT ETH EBI GRAT

Assim, a leitura correta das trincas seria interrompida após uma mutação (+), sendo, entretanto, restabelecida após uma mutação (-). Se a região alterada entre as duas mutações não for crítica, o polipeptídeo pode ser funcional.

Mutações supressoras devem estar sempre próximas das mutações que elas suprimem, uma vez que é pouco provável que um polipeptídeo contendo grande número de aminoácidos alterados permaneça funcional. Caso tivéssemos uma segunda mutação de sinal (+), ela não teria suprimido a mutação (-) e não haveria reversão para o tipo selvagem.

Considere a seguinte sequência de bases:

CAT CAT CAT CAT CAT CAT CAT C...

a. Note como a inserção de uma base altera a sequência:

Inserção de um G:

G

CAT CAG TCA TCA TCA TCA TCA TC...

b. Note como a inserção de uma segunda base qualquer mantém a sequência alterada:

Inserção de um A:

A

CAT CAG TCA TAC ATC ATC ATC ATC...

A natureza tríplice do código somente foi demonstrada quando foram obtidos mutantes contendo três mutações do tipo inserções ou três deficiências, pois esses mutantes triplos tinham o fenótipo selvagem. Este resultado só seria possível se o código fosse lido em trincas, pois somente neste caso a ordem correta da leitura seria restabelecida após três alterações de mesmo sinal.

c. Note como a inserção de uma terceira base qualquer restabelece a fase de leitura da sequência original:

Inserção de um T:

T

CAT CAG TCA TAC TAT CAT CAT CAT C...

Os experimentos revelaram também a impossibilidade de apenas 20 dos 64 códons possíveis serem usados para especificar aminoácidos. Neste caso, 44 códons não teriam sentido, isto é, não corresponderiam a aminoácidos. Se isso fosse verdadeiro, um destes códons sem sentido deveria muito provavelmente aparecer entre uma inserção e sua deficiência supressora, determinando a terminação da cadeia polipeptídica antes que a leitura correta fosse restabelecida. Em outras palavras, estes estudos já sugeriam que, como veremos em breve, o código genético era **degenerado ou redundante**.

Dizer que o código genético é degenerado significa que mais de um tipo de trinca de bases pode determinar o mesmo aminoácido na cadeia polipeptídica.

A Decifração dos Códons

Simultaneamente à demonstração da natureza tríplice do código genético, os cientistas iniciaram a tarefa da decifração dos códons. Códon é uma trinca de bases no RNA mensageiro que especifica um aminoácido.

Em 1960, a função do RNA mensageiro na síntese de proteínas foi estabelecida, mas parecia que o código só poderia ser decifrado quando fosse tecnicamente viável sequenciar todos os aminoácidos das proteínas e sequenciar todos os nucleotídeos de um gene, comparando-os. Entretanto, em 1961, foi dado o passo fundamental para a decifração do código. Nos quatro anos seguintes, todo o código foi decifrado em um dos maiores avanços da genética moderna.

Marshall W. Nirenberg e Heinrich Matthaei demonstraram que uma molécula de RNA mensageiro poderia ser traduzida “in vitro”. Isso foi feito introduzindo-se um RNA sintético com sequências de bases conhecidas em um extrato celular que possibilitava a síntese de proteínas. Os constituintes essenciais nesse sistema eram ribossomos, RNAts, sintetases do aminoacil-RNAts (enzimas que ligam o aminoácido ao RNAt), ATP e aminoácidos. Nesse sistema isento de células um RNA sintético é reconhecido como um RNA mensageiro e sua sequência de bases é traduzida em um polipeptídeo.

O experimento era feito colocando-se um mesmo RNA sintético em vinte frascos contendo o extrato celular, sendo que em cada um deles foi colocado um tipo de aminoácido marcado radioativamente. Após algum tempo verifica-se em quais dos frascos existem polipeptídeos radioativos. Por exemplo, quando o RNA sintético usado foi um poli-U, ou seja, um RNA em que todos os nucleotídeos tinham a uracila como base, foram detectados polipeptídeos radioativos apenas no frasco onde havia sido colocada fenilalanina radioativa. A conclusão foi que a trinca UUU codifica para fenilalanina. Essa foi a primeira trinca identificada por Nirenberg e Mathaei.

A introdução de poli-C e poli-A produziram polipeptídeos constituídos apenas pelos aminoácidos prolina e lisina, respectivamente. A síntese de poli G apresentou problemas e não pode ser testada “in vitro”.

Nirenberg, Mathaei e Ochoa produziram, em seguida, RNAs contendo sequências de duas ou mais bases unidas ao acaso. Foi impossível prever a sequência na qual as bases se ligariam e a decifração precisa dos códons não pôde ser completada. Por exemplo, uma sequência ao acaso de igual número de bases uracila e adenina produziria oito trincas possíveis: UUU, UUA, UAA, AAA, AAU, AUU, AUA, UAU em frequências iguais. Então, polipeptídeos contendo quantidades iguais de oito aminoácidos poderiam ser produzidos neste sistema de síntese de proteínas. Os autores prosseguiram com este tipo de análise diminuindo e aumentando a quantidade de cada uma das bases na mistura, tentando observar se a frequência dos aminoácidos se alterava. Embora isto pudesse dar algumas informações sobre os códons desses aminoácidos, não era possível relacionar trincas particulares com os aminoácidos.

Um avanço significativo na decifração do código ocorreu quando Khorana desenvolveu um método para sintetizar copolímeros com sequência conhecida. Ele fazia isso produzindo inicialmente dinucleotídeos com ordenação conhecida, por exemplo, dinucleotídeos UG com o U na extremidade 5' e o G, na 3'.

Inicialmente, a tradução de copolímeros alternativos, tais como, UGUGUGUG revelou que poli UG codificava para valina-cisteína-valina-cisteína. Daí, pode-se concluir que os possíveis códons para valina e cisteína eram UGU ou GUG, mas não era possível ainda concluir qual correspondia a qual aminoácido.

Khorana, a seguir, sintetizou polirribonucleotídeos com trinucleotídeos repetidos, tais como: UUGUUGUUGUUG, os quais codificaram para três tipos de polipeptídeos: polileucina, policisteína e polivalina.

Em 1964, Nirenberg e Leder fizeram uma descoberta que possibilitou a decifração de todos os códons. Eles descobriram que a adição de um RNA com apenas três nucleotídeos (um trinucleotídeo) ao sistema de síntese de proteínas resultava na ligação de um tRNA (ligado ao seu aminoácido específico) aos ribossomos. Esses complexos ribossomo-aminocil-tRNA podiam ser separados da mistura de reação por meio de filtração em filtro de nitrocelulose, ao qual os ribossomos aderem. (Fig. 3)

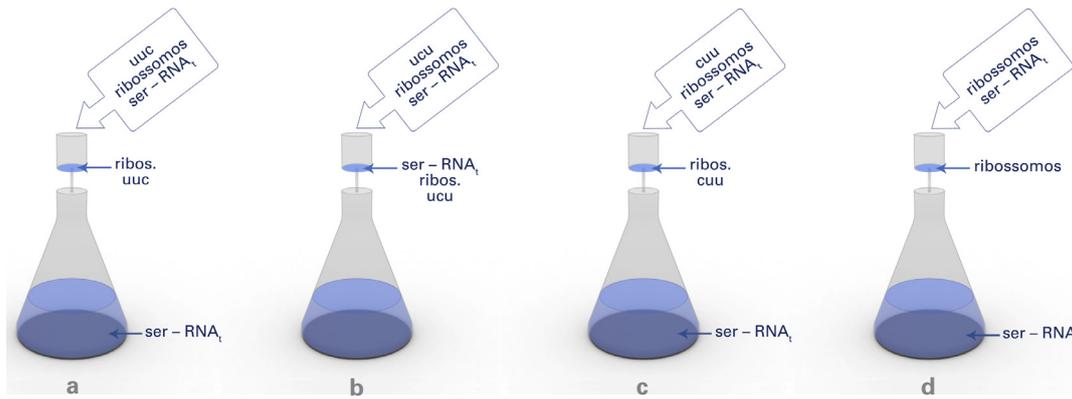


Figura 3. Representação esquemática do experimento de Nirenberg e Leder, em que se testou a retenção de aminoacil-RNAt, na presença de ribossomos e trinucleotídeos com sequência de bases conhecida. No exemplo, é mostrado que apenas a trinca UCU retém o complexo serina-RNAt (b), o que nenhuma outra das trincas testadas (a e c) ou apenas ribossomos (d) faz. Isso pode ser verificado utilizando-se aminoácidos radioativos e medindo a radioatividade retida no filtro após filtragem da mistura; apenas no frasco b o filtro apresentou radioatividade.

Diferentes trinucleotídeos foram preparados e, a seguir, foram realizados experimentos de ligação. Cada trinucleotídeo foi testado e foi verificado qual aminoácido radioativo surgia ligado ao complexo ribossomo-RNAt. Dentro de um ano, os aminoácidos tinham sido correlacionados com a maioria dos 64 códon possíveis. Somente os códon UAA, UAG e UGA permaneceram sem identificação. Mais tarde se demonstrou que essas trincas realmente não codificam aminoácido algum, sendo usadas como sinais de parada (códon pare). A figura 4 mostra as 64 trincas do código genético com seus aminoácidos correspondentes. Relembre que um códon é uma sequência de bases no mRNA e não no DNA. O anticódon está contido no RNAt que é complementar a cada códon. (Fig. 4)

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

Abreviaturas dos aminoácidos presentes em proteínas

- | | | | |
|-----------------|--------------------|------------------|------------------|
| Ala = alanina | Phe = fenilalanina | Ile = isoleucina | Ser = serina |
| Arg = arginina | Gly = glicina | Leu = leucina | Thr = treonina |
| Asn = aspargina | Gln = glutamina | Lys = lisina | Trp = triptofano |
| Asp = aspartato | Glu = glutamato | Met = metionina | Tyr = tirosina |
| Cys = cisteína | His = histidina | Pro = prolina | Val = valina |

Figura 4. A tabela do código genético.

As características fundamentais do código genético são mencionadas a seguir.

1. O código é inequívoco, ou seja, não é ambíguo: cada códon corresponde a apenas a um aminoácido e jamais a mais de um aminoácido.
2. O código é degenerado ou redundante: quase todos os aminoácidos têm mais de um códon: três têm 6 códons possíveis, 5 têm 4 códons, 1 tem 3 e nove têm 2 códons alternativos. Somente dois aminoácidos têm apenas 1 códon: a metionina e o triptofano.
3. O código contém 3 códons para os quais não foram encontrados aminoácidos; são eles: UAA, UAG e UGA. Esses códons foram chamados códons sem sentido (“nonsense” códons). Você verá mais tarde que eles são de fato códons de pontuação dando a instrução pare.
4. Códon alternativo para um aminoácido particular, normalmente, têm bases idênticas nas duas primeiras posições. A exceção óbvia para isso é quando há 6 códons possíveis para um aminoácido, nesse caso quatro dos códons têm as primeiras bases em comum.
5. Códon semelhante costumam especificar aminoácidos semelhantes. Por exemplo, os códons da treonina diferem dos códons da serina num único nucleotídeo na posição 5'. De modo semelhante, códons que têm U na segunda posição codificam para aminoácido hidrofóbico. Quando esses códons são alterados, tanto na extremidade 3' como 5', um aminoácido hidrofóbico ainda é incorporado na proteína. Devido a essa propriedade do código e também à sua degeneração, as células apresentam uma certa tolerância a substituições de um único nucleotídeo que nem sempre levam à produção de uma proteína não funcional.

Colinearidade entre DNA e Polipeptídeo

Comparando-se a sequência de nucleotídeos de um gene em procariontes com a sequência de aminoácidos em um polipeptídeo, nós podemos afirmar que o gene e a proteína são colineares, ou seja, a sequência de bases do gene corresponde exatamente à ordem dos aminoácidos na proteína. Cada gene representa um trecho contínuo da molécula de DNA e o seu comprimento é proporcional ao número de aminoácidos do polipeptídeo que ele codifica. Um peptídeo com N aminoácidos deve ser codificado por um gene que apresenta, no mínimo, 3N nucleotídeos.

A colinearidade entre o gene e a proteína foi inicialmente investigada por Yanofsky através de diversas mutações no gene da enzima sintetase A do triptofano de *E. coli*. A distância genética entre as mutações foi medida através da frequência de recombinação. A distância na proteína foi medida através do número de aminoácidos separando os sítios de substituição de nucleotídeos. Como podemos observar na figura 4, a ordem de diversos sítios de mutação distintos corresponde exatamente à ordem das substituições dos aminoácidos no polipeptídeo. Além disso, as frequências de recombinação são proporcionais às distâncias entre os aminoácidos substituídos na proteína. (Fig. 5)

O mapa de recombinação mostrou outros pontos importantes a respeito da organização do gene. Diferentes mutações podem fazer com que um mesmo aminoácido selvagem seja substituído por aminoácidos diferentes. Se duas dessas mutações não se recombinam, elas devem corresponder a substituições diferentes **no mesmo** nucleotídeo do DNA. Se elas se recombinam, mas afetam o mesmo aminoácido, elas devem representar mutações em posições diferentes no códon que determina o aminoácido alterado.

Em resumo, o trabalho de Yanofsky não só demonstrou a colinearidade entre a sequência de trincas e a sequência de aminoácidos especificada por ela, mas também demonstrou que um códon que foi alterado por mutação no meio de um polipeptídeo, frequentemente, resulta na substituição de um aminoácido. Uma mutação desse tipo é chamada mutação “missense” (sentido errado) e resulta em um polipeptídeo com um único aminoácido substituído. Este trabalho também mostrou, como proposto por Benzer, que pode ocorrer recombinação entre bases adjacentes. Até então, entendia-se a recombinação somente como a troca de alelos “inteiros” entre sequências homólogas, pois os genes eram vistos como unidades indivisíveis. Trabalhos dessa natureza evidenciaram que a unidade de recombinação genética era, de fato, o nucleotídeo, e não o gene.

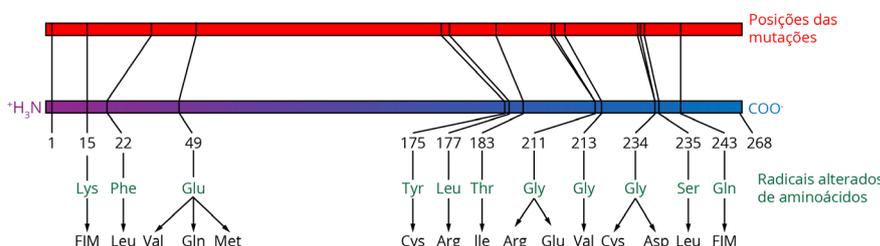


Figura 5. Colinearidade entre o mapa genético de mutações no gene da sintetase A do triptofano e as posições dos aminoácidos trocados na proteína dos mutantes

Os Códon de Pontuação

Ainda na década de 1950, Benzer lançou a sua hipótese de que uma classe de mutação observada por ele em fago T4 e denominada **ambivalente** era uma mutação do tipo sem sentido (*nonsense*), o que resultaria em um códon sem sentido no meio do RNAm, levando ao término prematuro da cadeia polipeptídica. Esses mutantes, no entanto, se multiplicavam normalmente em uma linhagem de *E. coli* denominada K permissiva, apresentando o fenótipo selvagem. Benzer imaginou que se o mutante ambivalente rIIa era capaz de crescer na linhagem permissiva de *E. coli* K é porque essas bactérias conseguiam ler os códon sem sentido, inserindo um aminoácido no ponto correspondente do polipeptídeo. Isto é, as linhagens K permissivas conteriam supressores dos códon sem sentido. Ele imaginou ainda que os supressores dos códon sem sentido poderiam ser RNAs com uma mutação no anticódon que os tornaria capazes de ler códon que normalmente não codificam para aminoácido.

Brenner aplicou essa idéia para outro grupo de mutantes do fago T4 chamados mutantes âmbar e foi capaz de provar que realmente existem códon sem sentido que interrompem a síntese de polipeptídeos.

Os mutantes âmbar tem mutações em certos pontos do genoma (um ponto diferente em cada tipo de mutante âmbar) que os impedem de crescer em linhagens normais de *E. coli*. No entanto, esses mutantes se comportam como selvagens quando crescem em linhagens K permissivas.

Brenner tinha vários mutantes do tipo âmbar que afetavam a síntese do polipeptídeo da cabeça do fago. Cruzando esses mutantes entre si, ele construiu um mapa genético fino da estrutura do gene, estabelecendo a ordem e as distâncias entre as mutações âmbar.

A seguir, ele determinou o tamanho do polipeptídeo da cabeça do fago produzido por cada um dos mutantes âmbar, quando eles infectavam bactérias não-permissivas. Lembre-se que nas bactérias permissivas os mutantes âmbar se comportam como selvagens produzindo polipeptídeos completos da cabeça do fago.

Os mutantes produziram polipeptídeos incompletos, cada um deles com um determinado comprimento. Ao ordenar os mutantes de acordo com o tamanho do polipeptídeo da cabeça produzido, Brenner verificou a existência de uma colinearidade entre esses tamanhos e as posições das mutações no mapa da estrutura fina do gene. (Fig. 6)

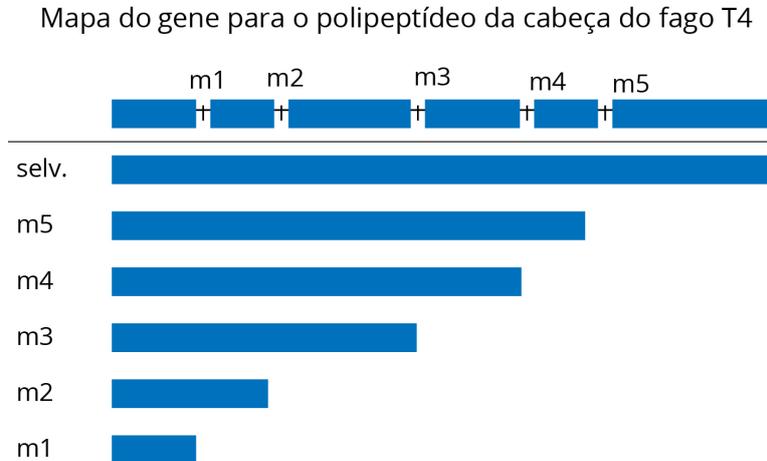


Figura 6. Colinearidade entre o mapa das mutações âmbar no gene que codifica para o polipeptídeo da cabeça do fago T4 e o tamanho dos fragmentos de polipeptídeo produzidos pelos mutantes em bactérias não-permissivas. No topo da figura está apresentada a sequência das mutações no gene para o polipeptídeo da cabeça do fago. Abaixo, os mutantes foram ordenados de acordo com o comprimento do polipeptídeo por eles produzido, o qual está indicado pela linha à direita do nome do mutante.

Esses resultados corroboraram a idéia de que as mutações âmbar ocasionam códon sem sentido que determinam a parada da síntese da proteína, resultando em fragmentos do polipeptídeo.

Brenner analisou as sequências de aminoácidos das extremidades das proteínas com término prematuro sintetizadas pelos diversos mutantes âmbar, e comparou-as com a sequência de aminoácidos da proteína sintetizada pelos fagos selvagens. Com isso ele pôde deduzir qual o aminoácido que deixou de ser adicionado na proteína de cada um dos mutantes e que provocou o término prematuro da síntese; teriam sido esses códon que mutaram, originando o suposto códon sem sentido típico dos mutantes âmbar.

Nos diversos mutantes, os aminoácidos que deixaram de ser adicionados nas proteínas mutantes foram glutamina, lisina, ácido glutâmico, tirosina, triptofano e serina. À primeira vista, não havia um padrão óbvio, mas Brenner deduziu que eles tinham algo em comum; pelo menos um dos códons codificadores desses aminoácidos podia ser obtido da trinca UAG pela substituição de uma única base. Assim, ele postulou que UAG era o códon de parada originado nas mutações âmbar; por isso, a trinca UAG tornou-se conhecida como **códon âmbar**. (Fig. 7)

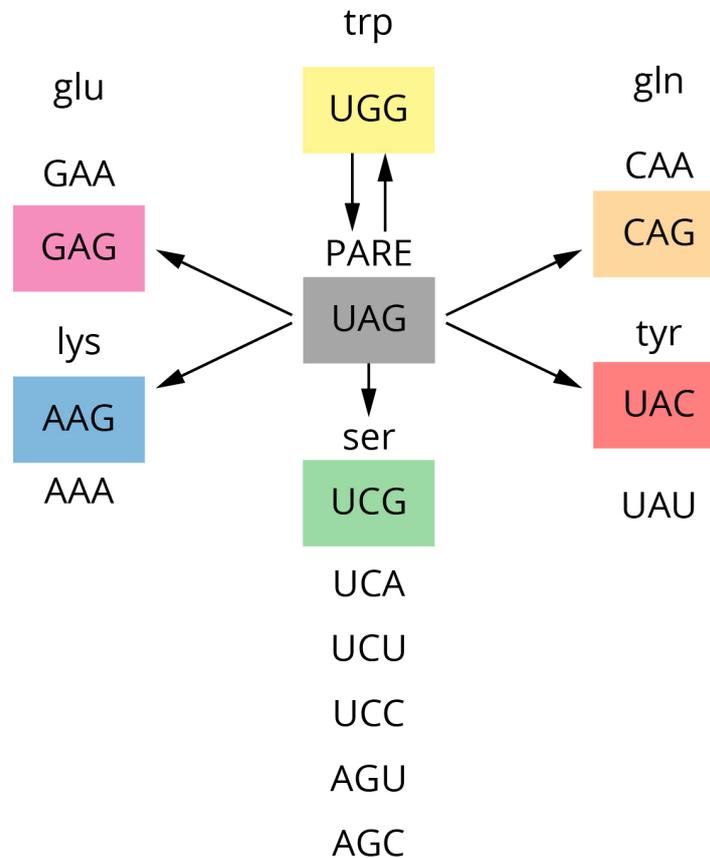


Figura 7. As substituições de aminoácidos que acompanham as reversões do mutante âmbar para o tipo selvagem. Os códons correspondentes estão assinalados.

Garen trabalhando com um sistema semelhante da fosfatase A de *E. coli* chegou à mesma conclusão. Em trabalhos posteriores, Brenner e Garen, trabalhando independentemente, mostraram que UAA era outro códon sem sentido; essa trinca ficou conhecida como **códon ocre**. Finalmente Brenner e Crick identificaram UGA como o terceiro códon sem sentido, que se tornou conhecido como **códon opala**.

A pergunta seguinte era se os códons sem sentido identificados em bactéria seriam usados normalmente pelos demais organismos para terminar seus polipeptídeos. Em outras palavras, seriam esses códons sem sentido os usados como **códon pare** na extremidade de qualquer RNAm? Isso foi demonstrado, bem mais tarde, pela técnica de sequenciamento de bases. Hoje não se tem mais dúvidas de que realmente aqueles três códons são o sinal **pare** para o término das cadeias polipeptídicas. Além disso, é comum se encontrar um ou mais códons sem sentido na extremidade de um RNAm.

A descoberta dos códons de término gerou a expectativa de que deveriam também existir códons de início da síntese das cadeias polipeptídicas, pois já existiam evidências de que a síntese de uma cadeia polipeptídica iniciava-se com o aminoácido metionina.

Em 1964, Sanger isolou de *E. coli* um RNAt ligado a um derivado do aminoácido metionina, o N-formilmetionina. Isso foi surpreendente, pois esse derivado tinha um grupo formil unido à extremidade amino da molécula do aminoácido. Em um polipeptídeo, os aminoácidos são conectados entre si através de uma ligação chamada peptídica que ocorre entre o grupo amina de um aminoácido e o grupo carboxila do seguinte.

O aminoácido N-formilmetionina apresenta o grupo amina bloqueado pelo formil, não podendo, portanto, ser usado para formar uma ligação peptídica com o grupo carboxila de um outro aminoácido. No caso, aminoácidos só poderiam ser acrescentados na extremidade carboxila do N-formilmetionina. Dessa maneira N-formilmetionina não poderia ocorrer no meio de um polipeptídeo, mas apenas em uma de suas extremidades.

Estudos realizados com lisozima do fago T4 sugeriram que a N-formilmetionina deveria estar no começo do polipeptídeo, isso porque havia evidências de que os aminoácidos conectavam-se uns aos outros com seus grupos amina voltados para o início do polipeptídeo e seus grupos carboxila dirigidos para a sua extremidade final.

O RNAt ao qual o N-formilmetionina estava ligado continha o anticódon 3' UAC 5', o códon para N-formilmetionina era, então, 5' AUG 3', o único códon que codifica para metionina.

Verificou-se que em *E. coli* havia dois tipos de RNAt que poderiam ligar-se à metionina. Ambos continham o mesmo anticódon. Entretanto, quando a metionina estava ligada a um determinado tipo de RNAt com anticódon UAC ela era prontamente formilada por uma enzima presente no sistema, resultando em um RNAt carregado com N-formilmetionina.

Sanger propôs, então, que todos os polipeptídeos iniciam com uma N-formilmetionina e que a tradução de um RNAm sempre começa com um códon AUG. Entretanto, nem todos os polipeptídeos isolados de *E. coli* começam com N-formilmetionina. Isso ficou esclarecido quando se descobriu que em um sistema da síntese de proteínas "in vitro" os polipeptídeos recém-formados tinham N-formilmetionina como primeiro aminoácido. Sabe-se, atualmente, que a N-formilmetionina é removida após a formação da proteína completa.

Se ocorrer um códon AUG no meio de um RNAm vai ser acrescentada metionina para o crescimento da cadeia polipeptídica. A N-formilmetionina não pode ser introduzida pois não tem um grupo amina livre para unir-se ao grupo carboxila do aminoácido precedente. Além disso, o RNAt que usualmente carrega a metionina só responde à presença de códons AUG internos e não ao códon AUG inicial. O RNAt que carrega a N-formilmetionina é diferente do RNAt que carrega metionina ao meio dos polipeptídeos, pois somente ele é capaz de reconhecer e se ligar ao ribossomo antes dele estar completamente organizado para prosseguir com o alongamento dos peptídeos.

Em bactérias, os códons GUG e UUG também podem ser códons de iniciação da tradução dos polipeptídeos. Embora eles normalmente codifiquem outros aminoácidos, quando eles ocorrem no início eles podem ser reconhecidos pelo RNAt carregado com a N-formilmetionina.

O Código Genético é "Quase" Universal

Os detalhes do código que foram demonstrados pelos experimentos in vitro são verdadeiros "in vivo"? Eles se aplicam aos organismos superiores, já que todos os trabalhos foram realizados com bactérias e fagos?

Hoje está largamente demonstrado que o código genético é praticamente universal. Na verdade, a universalidade do código pode ser considerada em diversos níveis. Assim, estudos sobre a sequência de bases do fago R17 mostraram que AUG é realmente o códon inicial desse vírus e que isso também é verdadeiro para leveduras e mamíferos. As observações sobre substituição de aminoácidos em mutantes de organismos tão diferentes como vírus do mosaico do tabaco, levedura e homem são compatíveis com o código estabelecido para *E. coli*.

Em extratos de *E. coli* contendo todo o material necessário para a transcrição e tradução, pôde ser sintetizada in vitro a proteína do vírus da poliomielite. RNA mensageiro para hemoglobina de coelho, foi traduzido em células de sapo, originando hemoglobina de coelho.

A universalidade do código genético sugere que ele deve ter sido estabelecido muito cedo na evolução.

As exceções à universalidade do código são raras e geralmente envolvem códons de parada. Por exemplo, em micoplasma, UGA codifica para triptofano e em certas espécies de ciliados UAA e UAG codificam para glutamina. As principais alterações do código ocorreram nos DNAs mitocondriais. Isto deve ter sido possível porque a mitocôndria é um sistema relativamente fechado, ou seja, RNAs transcritos no núcleo não são traduzidos por ribossomos mitocondriais e vice-versa. A maioria das alterações envolve os códons de parada, mas trocas no significado de alguns aminoácidos são também observadas, como podemos observar na tabela a seguir.

Tabela 1. Exceções ao código genético universal

Código Genético de Mitocôndrias			
Significado na mitocôndria			
Códons	Significado padrão	Mamíferos	Levedura
UGA	PARE	Triptofano	Triptofano
AUA	Isoleucina	Metionina	Metionina
CUA	Leucina	Leucina	Treonina
AGA	Arginina	PARE	Arginina

As Consequências das Mutações

Muitas mutações acarretam mudanças na sequência de nucleotídeos que não afetam o funcionamento do material genético dos organismos. Isso ocorre porque são mutações silenciosas que ocorrem na região de código dos genes, que ocorrem nos componentes não codificantes dos genes ou porque ocorrem fora dos genes, como estudaremos mais tarde. Em outras palavras, a maior parte dos genomas pode sofrer mutações sem qualquer efeito significativo no seu funcionamento. Já as mutações em regiões codificantes dos genes podem ter maior importância.

São as possíveis consequências das mutações de ponto (substituições) que afetam a sequência nucleotídica de um códon:

- a. A substituição pode ser sinônima, isto é, o novo códon especifica o mesmo aminoácido do códon não mutado. A mutação sinônima é também chamada de mutação silenciosa porque o gene mutado codifica para a mesma proteína codificada pelo gene não mutado, por causa da degeneração do código genético.
- b. A mutação pode ser não sinônima, isto é, ela gera um novo códon que especifica um aminoácido diferente. A proteína codificada pelo gene mutado terá, portanto, um aminoácido diferente. Muitas vezes isto não afeta a atividade biológica da proteína porque estas são capazes de tolerar certas mudanças em sua estrutura sem perda de sua função. Contudo, algumas alterações de aminoácidos, como, por exemplo, as que ocorrem no sítio ativo de uma enzima, têm um impacto maior e podem levar à perda de função da proteína. Uma alteração não sinônima é também chamada de mutação de sentido errado (“missense”).
- c. A mutação pode converter o códon que especifica um aminoácido em um códon de parada e por isso é chamada de mutação “nonsense”. Essa mutação sem sentido produz uma proteína menor porque a tradução do RNA mensageiro interrompe-se no códon mutado. O efeito desta mutação sobre a proteína dependerá do que foi perdido. Em geral, mutações desse tipo resultam em uma proteína não funcional.
- d. A mutação pode converter um códon de término de tradução em um códon que especifica um aminoácido. Isto resultará em uma proteína com aminoácidos adicionais. As proteínas poderiam tolerar estas mudanças sem o comprometimento de sua atividade biológica. No entanto, dependendo do número de aminoácidos adicionais, isto pode afetar a estrutura tridimensional da proteína, afetando sua função.

Como já estudamos, inserções e deleções também podem afetar muito a função dos genes. Se o número de nucleotídeos perdidos ou inseridos for três (ou múltiplo de três) um (ou mais códons) é perdido ou adicionado, produzindo efeitos variados sobre a estrutura da proteína codificada pelo gene. Se o número de nucleotídeos perdidos ou inseridos não for três ou múltiplo de três, haverá mudança no quadro de leitura (“frameshift”). Isto significa que todos os códons a jusante (a 3’) da mutação terão um quadro de leitura diferente daquele do gene não mutado. Isto resulta em alteração significativa na estrutura da proteína, já que parte do polipeptídeo codificado pelo gene mutado terá uma sequência de aminoácidos diferente da do polipeptídeo codificado pelo gene normal.

Os efeitos de mutações que ocorrem em regiões não codificantes do genoma são mais difíceis de serem previstos. Sítios no DNA de ligação a proteínas são suscetíveis a mutações de ponto, inserções e deleções que podem alterar a sequência de nucleotídeos envolvidos em interações DNA-proteína. Estas mutações poderiam inativar promotores ou sequências reguladoras da expressão gênica. Origens de replicação também poderiam perder a função se as mutações afetassem sítios relevantes para a ligação a outras proteínas envolvidas em complexos protéicos que participam da replicação do DNA.

Mutações também podem ocorrer em íntrons ou perto dos limites entre íntrons e éxons. Nestas regiões, mutações serão importantes se afetarem nucleotídeos envolvidos em interações RNA-RNA ou RNA-proteína que ocorrem no processo de *splicing*. Um resultado possível destas mutações é o de que o intron não seja removido do pré-RNA mensageiro. Consequentemente, isto equivaleria à inserção de novos códons, modificando o polipeptídeo traduzido.

As consequências das mutações em microorganismos

Muito do conhecimento sobre genética molecular que temos hoje resultou de experimentos realizados com microorganismos, principalmente as bactérias. Para se realizar experimentos genéticos em microorganismos, é importante que estejam disponíveis linhagens que apresentam mutações cujos efeitos fenotípicos sejam facilmente reconhecíveis experimentalmente. A maioria das mutações estudadas em microorganismos pode ser descrita com base nas propriedades de crescimento dos mutantes em diferentes meios de cultura. Assim, eles podem ser enquadrados nas seguintes categorias:

- a. **Mutantes auxotróficos** crescem apenas quando suplementados com um nutriente que não é necessário para a célula não mutada (prototrófica). Por exemplo, *E. coli* sintetiza normalmente seu próprio triptofano, graças às enzimas codificadas por cinco genes pertencentes ao operon triptofano. Bactérias que apresentem mutações nestes genes que levem à falta do aminoácido serão mutantes auxotróficos. Neste caso, o mutante não sobrevive em um meio de cultura não suplementado com triptofano.
- b. **Mutantes letais-condicionais** são mutantes que não resistem a determinadas condições de crescimento. Quando crescidos em condições normais (permissivas), eles parecem normais. Entretanto, se transferidos a condições restritivas, seu fenótipo anormal se manifesta. Mutantes sensíveis à temperatura são exemplos típicos de mutantes letais-condicionais. Eles crescem como células selvagens sob temperatura normal mas crescem mal, ou não crescem, quando a temperatura sobe acima de um limiar que pode variar para cada mutante.
- c. **Mutantes resistentes a inibidores** podem resistir aos efeitos tóxicos de um antibiótico ou de outro inibidor. Em alguns casos, a mutação modifica a estrutura da proteína que interage com o inibidor. Neste caso, o inibidor perde sua função porque já não pode interagir com a proteína alvo. Em outros casos, a mutação altera a proteína responsável pelo transporte do inibidor ao interior da célula.
- d. **Mutantes de regulação** possuem promotores ou outras sequências reguladoras alteradas. Nesta categoria estão os mutantes constitutivos, aqueles que expressam continuamente genes que, normalmente, estão ativos apenas em certas condições. Esses mutantes serão estudados posteriormente.

As consequências das mutações nos organismos pluricelulares

O primeiro ponto a ser considerado neste caso refere-se à distinção entre os casos de mutação que ocorre em uma célula somática e de mutação que está presente em um gameta. Como a célula somática não transmitirá seu material genético para os indivíduos da próxima geração, mas somente às células-filhas, uma mutação em uma célula somática é importante para o organismo no indivíduo em que ela ocorre, isto é, ela não tem impacto nas demais gerações. Na verdade, a maioria das mutações somáticas não é significativa mesmo que elas produzam a morte celular porque haverá um número enorme de células idênticas no mesmo tecido. Uma exceção seria quando uma mutação afeta uma célula de modo a contribuir, por exemplo, à formação de um tumor.

Mutações nas células que originam os gametas são mais importantes para a espécie, pois elas podem ser transmitidas aos organismos da geração seguinte e, portanto, podem estar presentes em todas as células dos organismos que herdaram a mutação. A maioria das mutações nos gametas, sejam as silenciosas ou as que ocorrem em regiões codificantes, não alterará significativamente o fenótipo dos organismos portadores. Mas aquelas que o fazem, podem ser divididas em duas categorias, em função do efeito que produzem sobre a atividade da proteína:

- a. **Mutações de perda de função:** são as que reduzem ou anulam a atividade proteica: A maioria dessas mutações que levam à perda de função são recessivas, já que o heterozigoto portará uma versão não mutada do gene que codifica para uma proteína funcional. Desta forma, a versão normal do gene “disfarça” o efeito da mutação. Existem, no entanto, exceções nas quais a mutação de perda de função pode agir como dominante se o produto do gene for realmente necessário em dose dupla.
- b. **Mutações de ganho de função:** são as que levam à expressão de um ou mais genes em tecidos onde eles normalmente não se expressam. Alternativamente, a mutação poderia levar à expressão exagerada de um gene, ou ainda à expressão de um produto gênico com funções distintas do produto original. Elas são menos comuns que as de perda de função e elas podem residir em regiões reguladoras do gene, e não necessariamente em regiões codificantes.

EXERCÍCIOS

Parte A: Revendo Conceitos Básicos

Preencha os espaços em branco nas frases de 1 a 8 usando o termo abaixo mais apropriado:

- (a) códon
- (b) códons sem sentido
- (c) códons sinônimos
- (d) degeneração
- (e) mudança de quadro de leitura
- (f) mutação de sentido errado
- (g) mutação sem sentido
- (h) mutação supressora

1. () é a característica do código genético na qual mais de um códon pode significar um mesmo aminoácido.
2. Trinca de bases no mRNA que codificam para um mesmo aminoácido na cadeia polipeptídica são ().
3. O resultado de mutações do tipo deleção ou inserção de bases, que afeta a leitura de todos os códon a partir do local onde ela ocorre, é chamado ().
4. Uma trinca de bases do mRNA que determina um aminoácido no polipeptídeo é um(a) ().
5. () é uma alteração no DNA que introduz um códon de parada no mRNA, acarretando a produção de um polipeptídeo mais curto.
6. Uma alteração no DNA que resulta na substituição de um aminoácido no polipeptídeo é chamada ().
7. Trinca de bases que não codificam para nenhum aminoácido e acarretam a parada da síntese do polipeptídeo são chamadas ().
8. Uma alteração no DNA que compensa os efeitos de outra alteração é chamada ().

Parte B: Ligando Conceitos e Fatos

Indique a alternativa mais apropriada para completar as frases de 9 a 12.

9. A primeira demonstração de que realmente o código genético era baseado em trinca de bases foi realizada
 - a. por Nirenberg, que demonstrou que a adição de um RNA com apenas 3 nucleotídeos determinava a ligação de um aminoácido específico ao ribossomo.
 - b. por Brenner, Crick e Garen, que mostraram serem três os códons de parada.
 - c. por Crick e Brenner, com a demonstração que mutações triplas do tipo inserções e deleções restauravam o quadro de leitura dos códons.
 - d. por Khorana, que demonstrou o papel de copolímeros alternativos na síntese de proteínas.
10. A decifração de todos os códons que determinam aminoácidos nos peptídeos só foi possível em experimentos de síntese de proteínas “in vitro” em que foram usados RNAm contendo
 - a. um único tipo de nucleotídeos.
 - b. trinucleotídeos repetidos.
 - c. somente três ribonucleotídeos.
 - d. variações na concentração das diversas bases.

11. Brenner identificou a sequência exata das bases em um códon de parada por meio de
 - a. cruzamentos entre mutantes com deficiência e mutantes ambivalentes, que indicavam que o mutante ambivalente deveria ter uma mutação que pararia a síntese do polipeptídeo.
 - b. evidências que indicavam que três dos códons jamais especificavam aminoácidos em um sistema de síntese de proteínas “in vitro”.
 - c. análises da sequência de aminoácidos dos fragmentos polipeptídicos produzidos por diferentes mutantes âmbar, comparadas com o selvagem.
 - d. análises da presença de metionina na extremidade inicial de qualquer proteína.
12. A identificação do códon de início da síntese dos polipeptídeos pôde ser feita pelo fato de
 - a. todos polipeptídeos bacterianos terem sempre uma N-formilmetionina no seu início.
 - b. o códon AUG ainda não ter sido decifrado no código.
 - c. as observações de Sanger terem mostrado que alguns RNAs transportadores apareciam ligados à N-formilmetionina, derivada do aminoácido metionina, e que ela não poderia ser incorporada jamais no meio de um polipeptídeo.
 - d. um mesmo RNAt poder se ligar à metionina ou à N-formilmetionina indistintamente.

Parte C: Aplicando Conceitos

13. Suponha que a sequência de nucleotídeos abaixo corresponda a um pequeno gene. (a cadeia complementar à que está aqui representada é a que serve de molde para transcrever o RNAm).

5'AAATTGATGACAAAGAGTCCATCACTTAATCCCGCCAAATAAACACAG 3'

- a. Qual a sequência de bases no RNAm produzido por este gene? Indique sua polaridade.
- b. Qual a sequência de aminoácidos no peptídeo codificado?
- c. Se a molécula de DNA sofrer uma substituição de nucleotídeo, como a representada abaixo, como fica o polipeptídeo codificado?

5'AAATTGATGACAAAGAGTCCATCACTTAATCCGGCCAAATAAACACAG 3'

- d. Se a molécula de DNA sofrer uma substituição de nucleotídeo, como a representada abaixo, como fica o peptídeo codificado?

5'AAATTGATGACAAAGAGGCCATCACTTAATCCCGCCAAATAAACACAG 3'

- e. Se ocorrer a deleção de um nucleotídeo, como a representada abaixo, como fica o polipeptídeo?

5'AAATTGATGACAAA_ AGTCCATCACTTAATCCCGCCAAATAAACACAG 3'

- f. Na sua opinião, dentre os tipos de mutação representados, qual deve ser o mais deletério para o funcionamento do peptídeo? Por quê?
- g. Suponha que tenha ocorrido, como representado abaixo, uma deleção de mais dois nucleotídeos além da indicada no item anterior; como fica o polipeptídeo correspondente?

5'AAATTGATGACAAA_ AGTC__TCACTTAATCCCGCCAAATAAACACAG 3'

- h. Se a molécula de DNA original sofrer uma substituição de nucleotídeo, como a representada abaixo, como fica o peptídeo codificado?

5'AAATTGATGACAAAGAGTCCATCACTTTTATCCCGCCAAATAAACACAG 3'

14. Discuta a veracidade da seguinte afirmação: “A troca de uma base na molécula de DNA leva obrigatoriamente a uma substituição de aminoácido na cadeia polipeptídica”. Sua resposta deve contemplar as características do código genético, as características dos genes eucarióticos e as características dos genomas dos organismos eucarióticos.
15. A troca de um único par de bases em um gene estrutural levou a uma alteração de aminoácido na cadeia polipeptídica. Essa alteração vai afetar obrigatoriamente a atividade da proteína correspondente?

Parte D: Resolvendo Problemas

16. O efeito de uma única mutação de ponto sobre a sequência de aminoácidos de uma proteína pode fornecer uma identificação precisa do códon usado para especificar um aminoácido particular. Assumindo mudança de uma única base em cada passo, deduza o códon tipo selvagem em cada um dos seguintes casos:
- Gln → Arg → Trp
 - Gln → Lys → Ile
 - Leu → $\begin{matrix} \nearrow \text{Trp} \\ \rightarrow \text{Val} \\ \searrow \text{Ser} \end{matrix}$
17. Um pesquisador determinou a sequência de aminoácidos de uma região de uma enzima da bactéria *E. coli*, a qual está representada a seguir:

- Ala - Pro - Trp - Ser - Glu - Lys - Cys - His...

Ele dispunha de uma série de mutantes que não mostravam atividade da referida enzima e conseguiu isolar de três deles a enzima inativa produzida, tendo obtido os seguintes resultados:

- Mutante 1: -Ala - Pro - Trp - Arg - Glu - Lys - Cys - His...
- Mutante 2: -Ala - Pro (a proteína terminava nesse ponto)
- Mutante 3: -Ala - Pro - Gly - Val - Lys - Asn - Cys - His...

Com base nessas informações responda:

- Qual é a natureza molecular de cada mutação?
 - Qual é a sequência de bases no DNA que especifica esta parte da proteína?
18. A adição de um nucleotídeo na cadeia de DNA e a deleção de outro a uma distância de 15 nucleotídeos, causa na proteína a seguinte alteração:
- original =
... - Lys - Ser - Pro - Ser - Leu - Asn - Ala - Ala - Lys
 - mutante =
... - Lys - Val - His - His - Leu - Met - Ala - Ala - Lys
- Determine as sequências de nucleotídeos dos RNAm que codificaram essas cadeias de aminoácidos.
 - Que nucleotídeo foi retirado e qual foi adicionado?
19. Uma cadeia de DNA de dupla-hélice tem a sequência de bases representada abaixo e produz “in vivo” uma cadeia polipeptídica com 5 aminoácidos.
- Qual fita do DNA é a transcrita para originar o RNA mensageiro?
 - Marque as extremidades 5' e 3' de cada uma das fitas.

TAC ATG ATC ATT TCA CGG AAT TTC TAG CAT GTA

ATG TAC TAG TAA AGT GCC TTA AAG ATC GTA CAT