

EVOLUÇÃO MOLECULAR

2006

Modificado por Tatiana Roman, a partir do texto da Prof. Tania de Azevedo Weimer (ULBRA)

INTRODUÇÃO

A evolução molecular compreende duas áreas principais de estudo: a evolução das macromoléculas e a reconstrução da história evolutiva dos genes. De forma didática, podemos dividir estas áreas em alguns tópicos mais específicos, tais como: evolução de macromoléculas em si, origem de moléculas biológicas básicas, origem e diversificação do genoma, evolução da estrutura gênica, evolução da síntese protéica, evolução dos códons e do código genético, e evolução do tamanho e conteúdo do genoma.

Evolução das macromoléculas e origem de moléculas biológicas básicas

Para entender a evolução das macromoléculas é preciso considerar algumas questões chave:

- a) Qual das macromoléculas originou-se primeiro: proteínas ou material genético e entre estes, DNA ou RNA?
- b) Como surgiram?
- c) O seu surgimento foi anterior ou concomitante a existência de uma estrutura celular completa?
- c) Independente da resposta da pergunta anterior, como funcionaram?

Em termos evolutivos, acredita-se que, antes de qualquer forma de vida, foi necessário o surgimento de algumas moléculas primordiais, capazes de fornecer condições para o surgimento da vida em si, ou então de outras moléculas intermediárias, necessárias para a existência da vida. Estas moléculas primordiais seriam macromoléculas surgidas a partir de um processo que pode ser chamado de “evolução química”, e que pode muito bem ser classificado como o “início da origem da vida”. Em outras palavras, podemos afirmar que antes da evolução da vida deve ter ocorrido uma evolução química.

Há duas dificuldades básicas para o entendimento da evolução química: a primeira é que é

difícil visualizar o meio molecular e os eventos que ocorreram num passado tão distante (estima-se de $4,0$ a $4,2 \times 10^9$ anos atrás); a segunda se refere à inexistência de fósseis moleculares ou, se estes existem, à sua raridade. De qualquer forma, devemos considerar que um primeiro passo evolutivo deve ter sido a origem e evolução de matéria orgânica numa atmosfera primitiva, sob condições abióticas, sendo o objetivo maior desta evolução o surgimento de uma célula moderna, com capacidade de auto-reprodução.

Mas, como esta célula poderia surgir a partir do caos original? Duas considerações importantes devem ser feitas:

a) os organismos vivos primitivos não necessitariam ser tão complexos como os atuais. As características básicas da vida (crescer e multiplicar) já deveriam existir no organismo primitivo, mas é razoável supor uma forma de vida mais simples, principalmente na ausência de competição com as formas mais sofisticadas posteriores.

b) a formação de moléculas orgânicas e estruturas subsequentes não devem ter sido o resultado de eventos completamente ao acaso. Esta consideração é apoiada pelo fato de que a evolução de nosso sistema solar oferece os pré-requisitos essenciais para o desenvolvimento e manutenção da vida:

- a quantidade de energia disponível a partir da irradiação solar é maior do que a fornecida por qualquer outra fonte;

- diferentes elementos (H, C, O, N, S, P, Ca) bem como compostos de C já estavam presentes antes e durante a formação do nosso sistema solar. Através de espectroscopia observa-se grande quantidade de moléculas orgânicas em nuvens estelares;

- a órbita da Terra fica a uma distância uniforme do sol, o que evita temperaturas extremas nas quais moléculas orgânicas são incapazes de se formar e/ou funcionar;

- a presença de água (excelente solvente estável) parece ser anterior ao começo dos registros geológicos;

- gases contendo H existiram durante um longo período inicial da história da Terra.

Assim, uma considerável pré-adaptação molecular deve ter ocorrido na “sopa primitiva” de compostos químicos, permitindo posteriormente a evolução da matéria orgânica. Embora não se possa ter certeza da natureza dos eventos moleculares originais, pode-se tentar reproduzir os mesmos a partir de reações atuais, sob condições controladas que imitem o ambiente pré-biótico,

deduzindo, desta forma, a origem da matéria orgânica.

Embora não se saiba ainda, com segurança, como era esta atmosfera pré-biótica (se fortemente redutora, fracamente redutora ou mesmo não redutora), experimentos sob diferentes condições têm mostrado a possibilidade de síntese orgânica em condições abióticas, tendo-se já obtido os seguintes compostos: aminoácidos, açúcares (ribose, glicose, desoxirribose), purinas, pirimidinas, porfirinas (que compõem hoje o núcleo do heme, da clorofila). A única certeza que se tem a respeito do ambiente primitivo é que não existiu oxigênio livre até o advento das bactérias fotossintéticas (cerca de $2,5-2,9 \times 10^9$ anos).

Na década de 1920, Oparin (bioquímico russo) e Haldane (geneticista inglês) propuseram que a Terra teria uma atmosfera redutora e que os compostos orgânicos formados nesta atmosfera deveriam ser similares aos utilizados pelos organismos vivos atuais. No entanto somente cerca de trinta anos mais tarde é que esta sugestão foi comprovada por dados experimentais. Miller, na década de 50, sintetizou moléculas orgânicas numa atmosfera redutora, em condições abióticas a partir de metano, amônia, hidrogênio e vapor d'água. Muitas moléculas foram formadas e embora as quantidades de cada composto, por reação, fossem pequenas, a quantidade total foi grande.

O material formado, no ambiente primitivo, pode ter-se acumulado em locais protegidos (por exemplo, da UV que degrada matéria orgânica), tais como fissuras de rochas ou profundidade de poços, concentrando-se em conseqüência. Uma vez concentrado houve chance para ocorrerem as polimerizações com formação de peptídeos, polissacarídeos e outros tipos de moléculas maiores.

Em meio aquoso as polimerizações dependem de agentes condensadores e muitos deles existiam na Terra primitiva: cianamida, cianogênio, ácido cianico, dicianamida. A formação de nucleotídeos, por exemplo, pode-se dar por fosforilação de adenosina, citosina e uridina sob a catálise de cianogênio. Em condições parcialmente anidras polimerizações também podem ocorrer, sob ação do calor, por evaporação de H_2O , na ausência de agentes condensadores. Pirofosfato pode se formar a partir de ortofosfato e pode ser usado para a formação de ADP e ATP.

Tem sido sugerido que as primeiras fontes de energia dos organismos seriam cadeias de polifosfatos e que o ATP seria introduzido só posteriormente. Lipmann (1965) propôs que a reação que usa PPi e o transforma em Pi liberando energia seria o fóssil molecular da forma primitiva de transferência de energia.

Fox e colaboradores, nos anos 50, obtiveram proteinóides (substâncias semelhantes a

proteínas) a partir de uma mistura de aminoácidos (AA). A proporção de AA nos proteinóides não se dava ao acaso nem era proporcional às quantidades de cada AA na mistura. Alguns ocupavam preferencialmente as extremidades terminais (N e C) dos proteinóides, havendo também interações preferenciais de AA.

- *Quem veio primeiro, o ovo ou a galinha?*

As interações entre ácidos nucléicos e proteínas são complexas, uns dependendo simultaneamente dos outros. Desta forma, é muito difícil deduzir qual deve ter sido a molécula evolutivamente mais primitiva, ácido nucléico ou proteína.

Duas hipóteses tentam explicar esta questão. De acordo com a primeira, os ácidos nucléicos teriam evoluído primeiro pela sua capacidade de auto-replicar e a partir dele surgiria um sistema protéico que facilitaria posteriormente a auto-replicação. A descoberta do RNA catalítico apoia esta hipótese. Para a segunda, os primeiros sistemas auto-replicativos seriam proteínas ou combinações ácido nucléico-proteína. Haveria interdependência, desde o início, entre proteínas e nucleotídeos. Apoiando esta hipótese, há a observação de moléculas de antibióticos, oligopeptídios circulares, produzidas pela bactéria *Bacillus brevis*, sintetizadas na ausência de mRNA. A enzima que catalisa a reação (adição sequencial de aminoácidos) serve como molde para a sequência de aminoácidos do peptídeo.

Embora não se tenha uma resposta definitiva a este problema, a maioria dos autores tende a aceitar preferentemente a primeira, uma vez que ao longo do tempo novas evidências surgiram que a favorecem, as quais serão discutidas na seção a seguir.

Origem e diversificação do genoma

Admitindo os ácidos nucléicos como anteriores às proteínas, qual deve ter sido o primitivo, DNA ou RNA?

A idéia de Haldane (1965) e Lipmann (1965) de que os primeiros genomas eram constituídos de RNA tornou-se mais consistente com a descoberta de que moléculas de RNA apresentam propriedades catalíticas. Em adição às suas propriedades como molde, as moléculas de RNA têm propriedades enzimáticas e assim são as únicas moléculas que podem funcionar como genótipo e fenótipo. Várias são as características do RNA que capacitam essa molécula como

precursor genômico:

- tanto a replicação como o código genético podem ter evoluído só com o RNA (e não só com o DNA), já que o RNA pode se auto-duplicar (como ocorre com vírus de RNA) e a síntese protéica é toda mediada por RNA (mRNA, rRNA, tRNA);
- a replicação do RNA é mais simples do que a do DNA; o requerimento de *primers* de RNA para a duplicação do DNA pode ser um fóssil molecular da replicação primitiva;
- desoxirribonucleotídeos são sintetizados a partir de ribonucleotídeos e a síntese da timina se dá a partir de uracil. Assim, essas duas substâncias podem ter surgido mais tarde na evolução;
- ribonucleotídeos (e não desoxirribonucleotídeos) têm um papel fundamental no metabolismo. ATP, CTP, GTP, AMP cíclico, NAD, NADP, por exemplo, são todos ribonucleotídeos, sugerindo que os principais aspectos do metabolismo já estavam estabelecidos antes da evolução do DNA;
- o RNA tem propriedades catalíticas, como exemplificado pela atividade de auto-splicing de alguns grupos de introns.

Todas estas evidências sugerem que a transição entre o mundo químico e o biológico foi feita através de algum tipo de ácido nucléico, provavelmente o RNA. Porém, acredita-se que o material genético primitivo tenha sido na verdade um precursor de RNA baseado em análogos de ribose, ao invés de ribose. Embora pudessem ser formados a partir da sopa primitiva, sabe-se que alguns tipos de nucleotídeos não deveriam ter sido gerados em grandes quantidades, pelo menos num primeiro momento. Além disso, a ribose pode ser sintetizada sob forma de dois estereoisômeros, D-ribose e L-ribose. Somente a D-ribose é usada para formar o RNA, embora ambas sejam formadas em quantidades iguais nos experimentos pré-bióticos. Um tipo de molécula análogo aos nucleotídeos de ribose, sintetizado mais facilmente e que não fosse inibida pela presença de diferentes estereoisômeros, mas que tivesse a propriedade de autoreplicação deve ter sido o precursor do RNA. Tal molécula poderia ter sido, por exemplo, o glicerol fosfato, que, além das características acima, também retém a capacidade de pareamento. Além disso, uma versão de ácido nucléico baseada somente em purinas deve ter precedido os ácidos nucléicos atuais. Posteriormente, o RNA poderia ter surgido para estabilizar o genoma precursor estruturalmente, e, ao provar ser superior a este funcionalmente, tê-lo substituído gradativamente.

A passagem do mundo de RNA para o de DNA só teria ocorrido após a evolução da síntese protéica, do surgimento do processo de transcrição reversa e do desenvolvimento de

processos enzimáticos capazes de converter precursores de ribonucleotídeos em precursores de desoxirribonucleotídeos. Uma vez formado o DNA este se manteve como repositário permanente da informação genética devido a três características que lhe conferem maior estabilidade:

- a falta do grupo OH⁻ no carbono 2' da ribose, o que torna a molécula menos suscetível à hidrólise;
- a estrutura em dupla hélice, que não só facilita a replicação, como é mais conservativa, facilitando o mecanismo de reparo, uma vez que a fita intacta serve para a correção da fita anômala;
- a substituição de U por T representa uma alta estabilidade genética (ou baixa taxa de mutação), devido a alta estabilidade tautomérica da T em comparação à U.

A pró-célula com material genético de DNA deve ter sido mais similar a células eucarióticas modernas do que a células procarióticas, e deve ter garantido a alta especialização evolutiva, onde o DNA fica responsável pela manutenção e transmissão da informação genética, e as proteínas pelas atividades enzimáticas, mantendo-se o RNA como a ligação entre esses dois tipos de moléculas.

Evolução da estrutura gênica

O gene eucarioto atual tem a estrutura composta por exons e introns, enquanto os dos procariotos não têm introns. Considerando que os exons codificam pedaços de proteínas que correspondem a módulos estruturais ou funcionais (domínios protéicos), duas possibilidades existem para explicar a estrutura atual dos genes: os introns teriam sido inseridos por algum processo de transposição, após a separação entre procariotos e eucariotos, tendo esta inserção separado, com precisão, as regiões codificadoras; a estrutura exons/introns faria parte do genoma primitivo e os introns teriam sido eliminados da linhagem procariota por economia genômica.

Gilbert (1987) sugere ser a segunda hipótese a correta, baseado no fato de que os introns, apesar de aparentemente sem função, são *hot spots* recombinacionais, devido ao seu grande tamanho. Este autor propõe que os primeiros genes, essencialmente de RNA, tinham estrutura com introns e exons. A partir deste gene primitivo diferentes “cópias” poderiam ser feitas, com a reorganização dos exons simples já existentes. Os introns atuariam como os elementos de reunião e separação, afastando exons adjacentes ou aproximando exons distantes, possivelmente através de um mecanismo de *proto-splicing*. Isto permitiria a obtenção de genes diferentes e mais complexos, conforme as combinações resultantes dos exons reunidos e separados. Este processo é chamado de

“*exon shuffling*” (“embaralhamento de exons”), e é aceito como o mecanismo que deve ter atuado na organização dos genes eucariotos. O fato de que os exons de um determinado gene codificam domínios da proteína que equivalem a módulos estruturais e/ou funcionais é a maior evidência de que tal embaralhamento foi responsável pela organização gênica. Além disso, existem vários exemplos de genes eucarióticos claramente formados de partes (exons) de outros genes, como genes de proteínas do sistema sanguíneo de vertebrados. Como os genes formados de RNA continham introns, a evolução da estrutura genômica para o DNA manteve a combinação exons/introns.

Por outro lado, a formação de genes por embaralhamento sugere que os introns estavam presentes antes da evolução da estrutura gênica. Desta forma, seria coerente afirmar que os introns não foram inseridos após o surgimento dos eucariotos, mas sim perdidos pelas bactérias após a separação dos dois grandes reinos. Em outras palavras, o último ancestral comum de pro e eucariotos deve ter tido introns em seus genes, com os procariotos perdendo-os subsequentemente devido a uma pressão seletiva forte para redução do tamanho genômico e para um aumento na eficiência da expressão gênica, mais apropriada para células pequenas e de crescimento e multiplicação muito rápidos. A presença de homologia nos domínios estruturais e funcionais de uma série de proteínas em pro e eucariotos, bem como a localização similar de exons e introns, é uma evidência que apóia fortemente esta hipótese.

Evolução da síntese protéica

Acredita-se que as primeiras proteínas tenham sido homopolímeros simples (di ou tripeptídeos), envolvendo poucos aminoácidos, mas com as seguintes funções básicas: reunião de RNAs em membranas, isolando o material genético e melhorando a função celular, consequentemente; função como canais ou poros de membranas; suporte da estrutura tridimensional das ribozimas. A transição de um mundo de RNA para um de RNA-proteína provavelmente ocorreu gradualmente: os aminoácidos (aa) iniciais poderiam ter sido ativados para formar pequenas cadeias peptídicas com função estrutural, uma vez que o RNA e cofatores seriam suficientes para realizar as funções químicas nas estruturas celulares primordiais. Posteriormente, polimerizações mais complexas de aa poderiam ter suplantado gradualmente as funções da maioria das ribozimas, inicialmente através de combinações RNA-proteína e, em seguida, através de estruturas protéicas completas.

Para a evolução da síntese protéica, alguns passos críticos devem ter surgido e se estabelecido. O primeiro é o perfeito carregamento do aminoácido pelo tRNA, processo este mediado pelas aminoacilsintetases. Estas enzimas fazem dois reconhecimentos críticos: escolhem um dos 20 aminoácidos e os respectivos tRNAs entre cerca de 40 destas moléculas, havendo, assim, a necessidade de coevolução enzima-tRNA, para o desenvolvimento da síntese correta. O reconhecimento do aminoácido específico pela sintetase é um problema difícil, porque muitos aminoácidos são quimicamente semelhantes. Este reconhecimento pode ocorrer tanto por uma forte discriminação primária (tirosil-sintetase, cisteil-sintetase), ou por uma baixa discriminação, seguida de revisão (*proofreading*) (valil-sintetase). Esta revisão é feita pelas próprias, sintetases que desenvolveram uma atividade enzimática adicional para garantir a correção do processo: catalisam a hidrólise de aminoácidos incorretamente ligados, reforçando assim a discriminação perfeita. Embora não haja regras estritas, parece haver uma forte correlação negativa entre a similaridade dos aminoácidos e a capacidade seletiva da enzima, e entre esta última e a capacidade de correção. Os aminoácidos que apresentam maiores diferenças químicas são mais facilmente reconhecidos e suas sintetases apresentam menor capacidade de correção secundária.

A discriminação sintetase/tRNA é mais fácil pois o tRNA é uma molécula maior, com mais sítios de reconhecimento. Além disso, os tRNAs coevoluiram com as sintetases (há uma só sintetase para todos os tRNAs de um mesmo aminoácido), enquanto os aminoácidos são fixos. Verificou-se, em experimentos laboratoriais, que a discriminação entre tRNAs e sintetases de espécies diferentes é menor que aquele intra-específico, demonstrando que o tRNA e a sintetase de uma mesma espécie coevoluiram para a discriminação máxima, enquanto entre espécies diferentes não houve coevolução. O reconhecimento sintetase/tRNAs é feito, principalmente, pelo anticodon, especificamente pelo nucleotídeo comum a todos os anticodons de um mesmo aminoácido. Ainda, um dos últimos 3 pares de base da haste aceptora também é reconhecido, existindo também uma base discriminadora, invariável entre tRNAs de aminoácidos iguais, localizada entre a alça aceptora e a extremidade CCA do tRNA. O carregamento correto dos tRNAs depende exclusivamente de uma atividade de revisão exercida pelas sintetases: eles se ligam facilmente ao sítio de reconhecimento da enzima, mas a sua permanência depende de uma revisão feita pela sintetase.

O segundo passo crítico é o reconhecimento perfeito entre aminoacil-tRNA e ribossoma. A seleção do ribossoma entre os diferentes tRNAs depende do reconhecimento pelo sítio A do

anticódon, e a frequência de erros neste processo é diretamente correlacionada com a frequência de malpareamento codon-anticodon. A evolução dos codons foi, então, também um processo crítico para o estabelecimento da síntese protéica como a conhecemos hoje.

Evolução dos codons e do código genético

O codon é altamente conservado, sendo idêntico em todos os organismos, já que qualquer modificação seria fortemente selecionada contra. As únicas exceções observadas na universalidade do codon são:

a) Nas mitocôndrias: as alterações estão relacionadas à redução do tamanho do genoma. Várias modificações ocorrem nos codons das mitocôndrias e todas elas são compatíveis com a redução do número de tRNAs que passa de 30-40 no genoma nuclear para 22 nas mitocôndrias de mamíferos e a 25 nas de leveduras. As modificações envolvem, principalmente, a utilização de uma U na primeira posição do anticodon para parear com qualquer base na terceira posição do codon. Tais modificações são:

Codon Universal		Codon das mitocôndrias		
		Mamífero	<u>Drosophila</u>	Levedura
UGA	Fim	Trp	Trp	Trp
AUA	Ile	Met	Met	Met
CUN	Leu	Leu	Leu	Thr
AGA	Arg	Fim	Ser	Arg
AGG	Arg	Fim	não usado	Arg

b) No núcleo: há apenas duas exceções e ambas envolvem codons de Fim: em protozoários ciliados os codons UAA e UAG codificam para glutamina, havendo apenas um codon de Fim (UGA). No procarionto *Micoplasma* o codon UGA codifica para triptofano.

Um aspecto importante do codon é a ocorrência de pareamento oscilante (*wobbling*), na 3ª posição (1ª posição do anticodon). Se o *wobbling* fosse eliminado, seriam necessários 61 tRNAs para reconhecer os 61 codons correspondentes aos 20 aminoácidos. Os nove aminoácidos com 2

codons têm um único tRNA, aminoácidos com 4 codons têm 2-3 tRNAs e aqueles com 6 codons têm 3-5 tRNAs. Algumas das oscilações comuns são:

Base na 1ª posição do anticodon	Base reconhecida na 3ª posição do codon
U (não modificada)	U,A,C,G
U* (modificada)	A,G
G	C,U
C	G
I (inosina)	U,C,A
A (rara)	U

Uma consequência desta oscilação é que alguns aminoácidos são representados por mais de um códon, havendo uma boa correlação entre o número de codons de um aminoácido e sua ocorrência nas proteínas. O uso dos aminoácidos é assim, pelo menos em parte, dependente da disponibilidade dos codons. Dados estereoquímicos sugerem que os codons são determinados por afinidades químicas naturais entre os aminoácidos e os anticodons particulares. Uma possibilidade, então, é que o código genético tenha se originado por interações diretas anticodon-aminoácido e que as sintetases, que hoje garantem a exatidão da aminoacilação, evoluíram mais tarde. Outra possibilidade, segundo Osawa e Jukes (1988), é que o codon tenha evoluído por interações codon-anticodon. Estes autores propuseram 23 anticodons como o número mínimo para parear com os codons atuais dos 20 aminoácidos. A partir deste número mínimo teria surgido o código genético primitivo, que seria quase idêntico ao código genético atual das mitocôndrias de mamíferos. Este teria evoluído por diferentes pressões de mutação GC/AT em duas direções principais: por um lado teria originado o código dos eucariotos (e deste o dos protozoários ciliados), e por outro o das eubactérias e, a partir destas, o das mitocôndrias, cloroplastos e micoplasma.

Há, no entanto, várias outras propostas para a evolução do códon. Crick e cols (1976) propuseram que na interação mRNA/tRNA primitiva, além do anticodon, participariam as bases adjacentes. Como as alças de anticodon atuais têm uma estrutura 3'-NR \overline{C} Y- 5'(N = qualquer base, R = purina, Y = pirimidina e \overline{C} Y = anticodon), o pareamento só seria possível com um codon primitivo com composição RRY. Eigen e Schuster (1978) advogam, dentro do mesmo raciocínio, um codon primitivo do tipo RNY. Corroboram a sua teoria o fato de que os aminoácidos mais prováveis de terem ocorrido no ambiente primitivo (glicina, alanina, ácido aspártico e valina) podem ser codificados por codons RNY, mas não RRY.

Para Fitch e Upper (1987) o codon primitivo teria composição NNN, que evoluiria para NYN codificando aminoácidos hidrofóbicos e NRN para aminoácidos hidrofílicos. Num segundo passo evolutivo haveria diferenciação da 1ª base, gerando codons YYN, RYN, YRN, RRN que especificariam quatro grupos diferentes de aminoácidos. Assim, por exemplo, todos os aminoácidos cíclicos teriam codons com pirimidina na 1ª posição. Num último passo evolutivo as pirimidinas seriam separadas em C e U e as purinas em A e G. Jurka e Smith (1987) relacionaram a origem do codon à biossíntese proteica: os aminoácidos centrais que são interconvertíveis e a partir dos quais todos os outros podem ser sintetizados, são todos codificados por codons RRN. Assim, é possível que este tenha sido o codon básico primitivo.

Outra idéia, bastante interessante, sugere que o código genético inicial, sendo essencialmente composto por RNA, tenha se baseado em apenas duas bases: A e U. O agrupamento destas bases em triplets consegue produzir 8 diferentes significados, incluindo um códon de Fim. As demais possibilidades de combinações são compatíveis com aminoácidos pré-bióticos, com exceção da asparagina, o que, por sua vez, pode ter sido compatível com as primeiras formas de vida. A introdução gradual de novas bases, especificamente C e G, poderia originar novas combinações formadas por duas letras ‘velhas’ e uma letra ‘nova’, permitindo a codificação de novos aminoácidos, sobre os quais a evolução agiria no sentido de manter os que causassem as menores mudanças fenotípicas. Quando a proporção de letras velhas e novas nos codons fosse a mesma, atingindo o código o máximo de redundância possível, novos codons formados apenas por letras novas surgiriam, formando mais alguns aminoácidos. O código atual poderia representar, então, um equilíbrio entre todas estas opções e o efeito das mesmas nas proteínas e na atividade celular. A hipótese de que originalmente duas e não três bases poderiam formar os codons, sendo posteriormente agregada uma terceira base, também é compatível com a idéia de que apenas A e U existiam como código genético inicialmente. O fato de que a maioria da informação genética atual reside nas duas primeiras letras dos codons apóia esta hipótese.

Evolução do tamanho e do conteúdo do genoma

Desde a origem da vida a partir da evolução química, houve por um lado, um grande aumento na complexidade gênica até o surgimento do gene eucarioto e de sua organização cromossômica, e por outro um aumento no conteúdo de DNA. Desta forma a quantidade de DNA

de mamíferos é cerca de 1000 vezes maior que a das bactérias. Este aumento não é, no entanto, perfeitamente correlacionado à complexidade do organismo (o chamado paradoxo do valor C). Embora um organismo mais complexo deva ter maior quantidade de genes no genoma, uma quantidade maior de DNA pode corresponder a uma maior proporção de DNA não codificador e não, necessariamente, de genes funcionais. Hoje sabe-se que isso é verdade, tanto que a presença de seqüências não codificadoras pode deixar em patamares similares quanto ao tamanho do genoma organismos tão diferentes como sapos e homens.

Os altos conteúdos de DNA de organismos superiores são decorrentes, principalmente, de duplicações gênicas ou de duplicações genômicas (poliploidias). Este último fator deve ter tido um papel muito importante na evolução das plantas. Nos animais a poliploidia pode ter ocorrido apenas antes da evolução dos cromossomos sexuais, uma vez que após o perfeito estabelecimento desta forma de determinação sexual, os tetraplóides que surgissem não se manteriam na população. Por outro lado, a duplicação gênica por *crossing-over* desigual deve ter tido um papel muito importante na evolução genômica, o que se pode deduzir pelo grande número de genes que formam *clusters* ou famílias. A duplicação gênica pode originar genes com funções relacionadas, mas novas (genes da família da globina β) ou aumentar o número de genes com a mesma função (genes de rRNA e histona).

Formação de novos genes:

a) duplicação de genes completos: se dois genes são formados por duplicação de DNA, um deles pode mutar adquirindo uma função inteiramente nova. Note-se que uma nova função gênica pode surgir através de relativamente poucos passos mutacionais. A mioglobina e as cadeias da hemoglobina apresentam muitas similaridades, sugerindo a sua origem a partir de uma cadeia ancestral comum. As funções de alguns pares de proteínas homólogas como a mioglobina e a hemoglobina são consideravelmente diferentes, enquanto outras como as das cadeias β e α globina permanecem essencialmente iguais.

b) elongação gênica: muitas proteínas têm seqüências internas repetidas e estas repetições correspondem, freqüentemente, a domínios funcionais ou estruturais, corroborando a hipótese da origem de genes complexos por recombinação de introns de genes mais simples. Maeda e cols (1984) demonstraram, por exemplo, que o alelo 2 da haptoglobina humana resulta da duplicação do

alelo 1 por *crossing-over* desigual em um heterozigoto para os subtipos F e S deste alelo.

c) genes híbridos: o produto de um "crossing-over" desigual em uma região de DNA contendo dois genes pode ser um novo gene formado por partes daqueles dois. A hemoglobina Lepore humana é formada pela região 5' da cadeia α e 3' da cadeia β . Este fenômeno parece ocorrer mais ou menos freqüentemente, uma vez que diferentes hemoglobinas Lepore têm sido descritas, o que deve ser decorrente do alto grau de similaridade (93%) entre essas cadeias globínicas. Genes híbridos, em geral, levam a efeitos deletérios, a menos que o gene original seja mantido, juntamente com o híbrido. Neste caso o gene composto pode evoluir para uma nova função.

Aumento do número de genes com mesma função

O genoma eucarioto contém vários tipos de DNA repetitivo: famílias multigênicas (genes com funções relacionadas), famílias supergênicas (conjunto de genes com um ou mais domínios de origem comum), genes moderadamente repetitivos (genes para rRNA e tRNA) e levemente repetitivo (genes para as cadeias α e β da globina e para as cadeias de imunoglobulinas), além de seqüências altamente repetitivas ainda sem função totalmente esclarecida (DNA satélite, família ALU humana),

O processo básico para a evolução das famílias multigênicas é a duplicação recorrente, com o aumento do número de genes passo a passo. Conforme as necessidades da célula, e do organismo em si, os genes duplicados podem assumir outras funções, como explicado no item "duplicação de genes completos". Neste caso, duplicação seguida de divergência em diferentes graus é o principal mecanismo responsável (entenda-se divergência como a diminuição da similaridade na seqüência, devida principalmente a mutações).

No caso dos agrupamentos gênicos com níveis variados de repetição, como ocorre com os genes dos rRNAs, a manutenção da identidade da seqüência certamente é fundamental para a realização das funções dos produtos codificados. Assim, a divergência molecular posterior à duplicação deve ser muito mais limitada do que a que ocorre com as famílias gênicas clássicas, como a das globinas, por exemplo. É importante salientar que a localização genômica é fundamental para a divergência: cópias agrupadas tendem a permanecer mais similares entre si do que cópias localizadas em locais distantes, como em diferentes cromossomos. Isto não exclui a possibilidade de eventos intermediários de divergência, como ocorre na família das histonas, onde repetições de

genes capazes de codificar diferentes histonas existem ao longo dos genomas.

Aumento do número de seqüências altamente repetitivas

Para a evolução de seqüências altamente repetitivas a duplicação gênica apenas não parece ser suficiente e outras explicações têm sido propostas. Uma delas é a evolução em concerto: *crossing-over* desigual entre genes ou cromossomos não homólogos levaria à repetição de um segmento de DNA e a um aumento ou redução (ao acaso) do número de genes. O *crossing-over* teria o efeito de homogeneizar as seqüências num agrupamento, que, na ausência de mutação, manteriam-se idênticas. Na realidade, mutações sempre ocorrem, de modo que mesmo em um agrupamento espera-se encontrar algumas variantes. Acredita-se que este mecanismo tenha sido o principal responsável pela origem dos grandes blocos de DNA repetitivo em tandem, e que ainda hoje influencie na organização do genoma como um todo.

No caso dos mini e dos microssatélites, o escorregamento das fitas (*strand slippage*) durante a replicação do DNA parece ser o principal mecanismo de origem e organização destas seqüências, podendo ser responsável diretamente pelo tamanho das mesmas. A principal evidência a favor desta hipótese advém do fato de que o escorregamento envolve em geral pequenas partes de DNA de cada vez. Porém, é possível que mecanismos de *crossing-over* desigual tenham influenciado o surgimento e origem dos minissatélites, uma vez que a principal fonte de variação, neste caso, é o número de cópias de uma mesma seqüência presente em determinado loco cromossômico.

No caso dos satélites, escorregamento de fitas pode ter sido importante nos passos iniciais da formação do satélite, enquanto que *crossing-over* desiguais atuariam em estágios mais tardios da evolução dos mesmos. A principal justificativa para esta idéia é o fato de que os satélites são em geral resultados de repetições de alta ordem de seqüências curtas, havendo, então, possibilidade de ocorrência dos dois mecanismos.

Uma outra origem para as seqüências repetitivas pode ser a transcrição reversa. A família ALU, do genoma humano, é uma família gênica que apresenta cerca de 1 000 000 de cópias espalhadas por todo o genoma. As porções 5' e 3' do gene 7 SL RNA (RNA citoplásmico abundante que atua na síntese protéica como partícula de reconhecimento de sinal) são homólogas à seqüência ALU, sugerindo ter esta se originado por deleção da porção central do gene 7 SL RNA. A

ALU parece assim ser uma classe gigante de pseudogenes e ter-se originado por transcrição reversa do mRNA processado, com posterior inserção do transcrito. Como as ALU mantêm as seqüências promotoras e são transcritas, mesmo não sendo funcionais, podem gerar mais cópias de si mesmas e inserir-se no genoma pelo mesmo processo de transcrição reversa, podendo então mover-se como um transposon. Este efeito cascata aumenta grandemente a chance de dispersão das seqüências ALU através do genoma. Além das ALU outros pseudogenes parecem originar-se por transcrição reversa de mRNAs processados. O genoma do camundongo, por exemplo, possui muitos pseudogenes de globina, alguns dos quais faltam introns.

Independente do mecanismo de origem, a grande questão que fica em relação ao DNA repetitivo é: qual a sua função no genoma? Acredita-se que, de alguma forma, estas seqüências são funcionalmente importantes; ou, então, talvez elas sejam mantidas porque suas altas taxas de mutação contribuem para o potencial evolucionário a longo prazo da população, servindo em última análise como reserva de variabilidade. Esta última hipótese poderia explicar a ocorrência do DNA repetitivo em tandem. Mas, e quanto às seqüências ALU e outros tipos de transposons? A única vantagem adaptativa aparente parece ser para elas próprias, já que o potencial mutagênico é imenso. Porém, é possível que estas seqüências representem “genes baratos”, que podem conter alguma informação capaz de se expressar mais ou menos independentemente em diferentes locais do genoma. De qualquer forma, o papel das seqüências altamente repetitivas tanto na evolução do tamanho, como na organização do genoma como um todo é incontestável e fundamental, e é provável que não só tenha agido em épocas passadas, mas que também esteja atuando atualmente, continuamente nos genomas dos diferentes organismos vivos.

Relógio molecular

Um dos aspectos interessantes da evolução ao nível molecular é que, para uma dada proteína, a taxa de substituição de aminoácidos e de nucleotídeos é aproximadamente constante entre linhagens bem como dentro das linhagens. Se as proteínas evoluem a taxas constantes elas podem ser utilizadas para estimar as taxas e tempos de divergência entre organismos e para construir filogenias. Isto é conhecido como **relógio molecular**. Sua existência foi primeiro sugerida por Zuckerkandl e Pauling (1965), após a análise da taxa de substituição de aminoácidos da Hb e do Cit C em diferentes linhagens de mamíferos e da verificação da constância das mesmas.

Atualmente, taxas de mutações nas seqüências do DNA genômico e das organelas (mitocôndrias e cloroplastos) vêm sendo utilizadas no entendimento das relações evolutivas entre os organismos.

Alguma controvérsia existe sobre o mecanismo do relógio molecular, alguns autores pondo dúvidas sobre sua validade. Gillespie (1984,1986) sugeriu que a taxa de substituição variaria ao acaso, com o tempo, em cada linha evolutiva. Neste caso, o relógio molecular se tornaria episódico, com períodos de substituições sendo separados por outros com poucas alterações.

Segundo Goodman (1981) as taxas de evolução seriam maiores após a ocorrência de duplicações gênicas e as proteínas evoluiriam mais rapidamente em períodos de irradiação adaptativa.

Nei (1987) fazendo uma longa revisão sobre o assunto, concluiu que, embora a taxa de substituição não seja estritamente constante e varie consideravelmente entre os organismos, ela pode ser considerada aproximadamente constante quando se considera o longo tempo evolutivo.

Uma maneira simples de testar o relógio molecular é estimar o número de substituições de aminoácidos ou de nucleotídeos em cada ramo de uma árvore evolutiva e relacionar com o tempo evolutivo avaliado.

Alguns exemplos, no entanto, têm surgido ao longo do tempo de exceções ao relógio molecular:

a) o padrão de substituição de aminoácidos embora constante com o tempo evolutivo varia consideravelmente entre as diferentes proteínas a depender de seus requerimentos funcionais. Proteínas como as histonas, extremamente necessárias à estrutura do cromossomo eucarioto mantiveram-se evolutivamente estáveis (dos 105 aminoácidos destas proteínas apenas dois são diferentes entre espécies tão distantemente relacionadas como a ervilha e o boi). Proteínas menos críticas como o fibrinopeptídeo suportam muito mais mudanças. Por outro lado, dentro de uma mesma proteína, a taxa de substituição varia com o domínio estrutural ou funcional. Assim, por exemplo, a taxa de substituição na superfície da molécula de Hb é cerca de 10 vezes mais alta do que no grupo heme (Kimura e Ohta, 1973).

b) a estrutura primária também parece interferir na taxa de mutação. Graur (1985) mostrou uma correlação negativa significativa entre a proporção de glicinas e a taxa de mutação. Como a glicina é o menor aminoácido, sua substituição por outro, levaria provavelmente a uma alteração na função, sendo então selecionada contra.

c) o tempo de geração dos organismos também interfere com a constância evolutiva.

Verificou-se que a taxa de substituição de AA é duas vezes maior em macacos do que no homem (medida em sítios/ano) e é maior em roedores que em primatas. Considerando as variações nos tempos de geração destes organismos, as maiores taxas poderiam ser decorrentes de mais ciclos de duplicação de DNA. Quando organismos com mesmos tempos de geração são comparados as taxas evolutivas se mantêm constantes. Desta forma o relógio molecular se aplica para espécies estritamente relacionadas.

d) taxa de substituição em organelas:

O genoma das mitocôndrias de mamíferos tem de 15000 a 17000 pares de bases, correspondendo a 1/10000 do genoma nuclear. Apresenta uma única seqüência, não repetida, codifica 13 proteínas, 2 rRNA e 22 tRNA, apresentando uma região controle; é um genoma estruturalmente muito estável.

O genoma das mitocôndrias de plantas apresenta muita variabilidade estrutural, sofre rearranjos, duplicações e deleções com freqüência, de modo que o tamanho varia de 40000 a 2500000 pares de bases; este genoma pode ser linear ou circular e a informação pode ser dividida em moléculas de DNA separadas, em círculos sub-genômicos; codifica 15-30 proteínas, 3 rRNA, um número variável de tRNAs e os genes estruturais parecem estar em múltiplas cópias.

O genoma dos cloroplastos possui de 120000 a 220000 pares de bases e é bastante estável. Na *Nicotiana tabacum* os cloroplastos codificam 37 tRNA (8 dos quais contêm introns), 8 rRNA, 45 proteínas (algumas contêm introns) além de 59 "open reading frames" de função desconhecida.

A taxa de substituição sinônima nas mitocôndrias de mamíferos foi estimada em $5,7 \times 10^{-8}$ subst./sítio/ano; este valor é cerca de 10 vezes maior que aquele estimado para as substituições sinônimas dos genes nucleares. As taxas de substituição não sinônima variam consideravelmente entre proteínas, mas são bem maiores que aquelas dos genes nucleares. As maiores taxas de mutação do genoma mitocondrial podem ser devidas a: baixa fidelidade na replicação do DNA mit.; ausência de reparo ou reparo ineficiente nas mitocôndrias; alta concentração de mutágenos na mitocôndria (como radicais livres, superóxidos). As mitocôndrias dos vegetais e os cloroplastos apresentam menores taxas de substituição que os genes nucleares. Assim não existe correlação entre as taxas de substituição e a evolução estrutural, sugerindo que os dois processos ocorrem independentemente; da mesma forma não existe correlação entre a evolução das organelas e a

evolução do genoma nuclear, o que indica que os dois processos também são independentes.

e) as taxas de mutação variam com a região do gene, em função das restrições seletivas:

O estudo da taxa de evolução ao nível do DNA é muito mais informativo do que aquele envolvendo seqüências protéicas devido a:

a) muitas mutações não se refletem em aminoácidos, em função dos codons sinônimos;

b) muitas seqüências de DNA não são codificadoras.

Estudos neste nível de investigação têm mostrado que a taxa de substituição de nucleotídeos é muito maior na 3^a posição do codon que nas demais; isto é possivelmente devido ao fato de que a maioria das substituições na terceira posição (mas não todas) são silenciosas. Para a maioria dos genes funcionais a taxa de substituições sinônimas é muito maior (cerca de cinco vezes) do que a de não sinônimas e enquanto estas variam enormemente de gene a gene as sinônimas têm, em geral, taxas muito semelhantes em todos os genes. Assim a taxa de mutação na 3^a posição do codon nos genes de histona é aproximadamente igual à do fibrinopeptídeo, embora a proporção de substituição de aminoácidos seja praticamente nula naquela e bastante alta nesta.

Estudos comparando a taxa de substituição de nucleotídeos entre genes funcionais, pseudogenes e regiões não codificadoras mostraram que a taxa de substituição de nucleotídeos segue a seguinte ordem decrescente: pseudogenes > mutações sinônimas nas regiões codificadoras > introns > regiões flanqueadoras > mutações não sinônimas nas regiões codificadoras. Este padrão é facilmente explicável pelas diferentes funções e importâncias das várias regiões do DNA. Os pseudogenes sendo não funcionais podem acumular mutações ao longo do tempo, já que não sofrem a ação da seleção; raciocínio semelhante se aplica às regiões codificadoras, no que se refere às mutações sinônimas. Os introns e regiões flanqueadoras por estarem envolvidos no processamento do transcrito de RNA e regulação da síntese suportam menos as mutações e finalmente modificações nas regiões que levam a alterações das proteínas são eliminadas por seleção natural, mantendo-se em níveis muito baixos.

f) o uso de codons sinônimos não se dá ao acaso; este fato foi observado em procariotos e eucariotos. Genes de um organismo e de espécies relacionadas mostram o mesmo padrão de escolha entre os codons sinônimos; os codons mais utilizados são aqueles reconhecidos pelos tRNAs mais abundantes. Além disto os desvios dos codons são maiores em genes fortemente expressos que em genes fracamente expressos.

g) transferência horizontal de genes: ocasionalmente ocorre transferência de informação genética entre diferentes *taxa* ao longo da evolução. A simbiose entre procariotos e eucariotos, que deu origem às mitocôndrias e aos cloroplastos, seria o primeiro exemplo deste fenômeno. Vírus e plasmídios podem carregar genes e atuar como agentes para a transferência de genes entre espécies relacionadas ou não. A leghemoglobina de soja e de outras leguminosas é tão semelhante à Hb de vertebrados que parece possível que o gene tenha sido transferido dos animais para estas plantas.