

# Espressione genica: traduzione (parte 1)

## Le caratteristiche del codice genetico

Il codice genetico è l'insieme delle regole attraverso le quali viene tradotta l'informazione codificata nel materiale genetico, costituito da acidi nucleici, in proteine costituite da sequenze amminoacidiche. Il processo si svolge all'interno della cellula, che traduce una sequenza di codoni (o triplette di basi) di RNA in una sequenza di amminoacidi durante la sintesi proteica. Quasi tutti gli esseri viventi usano il medesimo codice genetico, chiamato codice genetico standard; tuttavia vi sono alcune eccezioni. Sono state riportate infatti variazioni a livello di alcuni batteri. Diversità nel codice genetico sono state riscontrate anche a livello dei mitocondri. La traduzione è la seconda fase dell'espressione genica. La prima fase è la trascrizione, fase in cui una data sequenza di DNA, chiamata gene strutturale, viene usata come stampo per la creazione di un filamento complementare di RNA.

La sequenza di RNA si compone di gruppi non sovrapposti di tre basi ciascuno, chiamati codoni. Ad ogni codone corrisponde uno specifico amminoacido, si dice quindi che il codone codifica quell'amminoacido nel codice genetico. Le basi dell'RNA sono quattro: adenina, guanina, citosina ed uracile (nel DNA l'uracile è sostituito dalla timina). Esistono quindi  $4^3=64$  codoni possibili. 61 di essi codificano gli amminoacidi, mentre i restanti tre (UAA, UAG, UGA) codificano segnali di stop (stabiliscono, cioè, a che punto deve interrompersi l'assemblamento della catena polipeptidica). Poiché gli amminoacidi che concorrono alla formazione delle proteine sono 20, essi in generale sono codificati da più di un codone (con l'eccezione di triptofano e metionina). Il codice genetico è pertanto detto degenero e codoni distinti che codificano il medesimo amminoacido sono sinonimi. Ad esempio, la sequenza di RNA UUUACACAG si compone di tre codoni, UUU, ACA, CAG, cui corrispondono rispettivamente gli amminoacidi fenilalanina, treonina e glutammina.

La sintesi proteica applicata a questa sequenza produrrebbe quindi il tripeptide fenilalanina-treonina-glutammina. Il codice genetico standard è riportato nelle due tabelle seguenti. Nella prima sono indicati quali amminoacidi corrispondono ad ognuno dei 64 codoni; nella seconda sono indicati quali codoni codificano ognuno dei 20 amminoacidi ordinari. Nella genetica classica ai codoni di termine (stop) vennero assegnati i nomi dei geni in cui furono individuati per la prima volta - UAG venne chiamato ambra, UGA opale e UAA oca. La traduzione inizia in corrispondenza di un codone di avvio (start) ma, a differenza del codone di termine, questi non è sufficiente per avviare il processo di sintesi; in prossimità del codone di avvio devono infatti anche trovarsi alcune sequenze tipiche che permettono all'mRNA di legarsi ai ribosomi. Il codone di avvio più noto è AUG, che codifica anche la metionina. Altri codoni di avvio sono CUG, UUG e, nei procarioti, GUG e AUU. Molti codoni sono ridondanti, cioè due o più codoni

corrispondono allo stesso amminoacido. Questa ridondanza è limitata alla terza base del codone - ad esempio, sia GAA che GAG codificano la glutammina.

Un codone è detto quattro volte degenerare se qualsiasi nucleotide nella sua terza posizione codifica lo stesso amminoacido (ad esempio, UCA, UCC, UCG e UCU, tutti corrispondenti alla serina); è detto due volte degenerare se solo due delle quattro basi nella sua terza posizione codificano lo stesso amminoacido (ad esempio AAA ed AAG, corrispondenti alla lisina). Nei codoni due volte degeneri, i nucleotidi equivalenti nella terza posizione sono sempre o due purine (A/G) o due pirimidine (C/U). La ridondanza rende il codice genetico meno vulnerabile alle mutazioni casuali. Un codone quattro volte ridondante può subire qualsiasi mutazione alla sua terza posizione ed un codone due volte ridondante può subire una delle tre possibili mutazioni alla sua terza posizione senza che l'amminoacido da esso espresso - e quindi la struttura della proteina in cui l'amminoacido verrà inserito - cambi. Inoltre, dato che le mutazioni per transizione (da una purina all'altra o da una pirimidina all'altra) sono più probabili delle mutazioni per transversione (da purina a pirimidina o viceversa), l'equivalenza tra purine o tra pirimidine nei codoni due volte degeneri aggiunge un'ulteriore resistenza. Sono solo due gli amminoacidi codificati da un unico codone. Sono la metionina, codificata da AUG che è anche il codone di avvio, ed il triptofano, codificato da UGG.

Diverse variazioni al codice genetico standard sono state trovate nei mitocondri. Anche i protozoi ciliati presentano qualche modifica: in loro (come anche in alcune specie di alga verde) UAG e, spesso, UAA codificano la glutammina e UGA codifica la cisteina. Addirittura nei ciliati si pensa che il CODICE 6 appunto, corrispondente a queste riassegnazioni dei codoni stop si sia evoluta fino a sei volte parallelamente. In alcune specie di lievito CUG codifica la serina. In altre specie di batteri ed archeobatteri i codoni di arresto codificano invece amminoacidi non comuni: UGA codifica la selenocisteina e UAG la pirrolisina. È possibile che vi siano altri amminoacidi non-standard la cui codifica è ancora ignota. Inoltre si possono avere diversissime variazioni anche del codice genetico mitocondriale, che inoltre ha tassi di evoluzione maggiore a seguito dell'inefficienza dei meccanismi di riparazione del DNA. Ad esempio, oltre ai già citati vertebrati (CGM 2), anche gruppi filogeneticamente vicini a noi, come ad esempio le Ascidie, hanno un codice genetico mitocondriale particolare. A dispetto di queste variazioni, i codici genetici usati da tutte le forme di vita della Terra sono molto simili. Dato che i codici genetici possibili e potenzialmente adatti alla vita sono molti, la teoria dell'evoluzione fa pensare che questo codice genetico sia andato a definirsi molto presto nella storia della vita su questo pianeta.

Le pressioni che poi possono aver portato all'evoluzione di codici genetici non canonici sono state sicuramente secondarie, in particolare si pensa legate a fenomeni di resistenza al trasferimento genico o all'attacco virale. Questo è stato dimostrato attraverso diversi studi. Il codice genetico è la correlazione tra l'informazione genetica espressa in nucleotidi e l'informazione espressa in sequenza amminoacidica. Il codice genetico è un codice a triplette, ovvero un gruppo di tre nucleotidi (un codone) codificano per un amminoacido. Un codice a tre

lettere genera, però, 64 codici possibili; dal momento che gli aminoacidi sono 20, alcuni di questi vengono codificati da più di un codone. Questa molteplicità del codice è chiamata degenerazione del codice: tranne due eccezioni (AUG che codifica per la metionina e UGG che codifica per il triptofano), per ogni aminoacido è presente più di un codone. La degenerazione, in molti casi, segue delle regole secondo le quali:

- codoni con la prima base differenti codificano per lo stesso aminoacido o per altri simili;
- codoni che differiscono per la seconda base codificano per aminoacidi simili;
- codoni che differiscono per la terza base codificano per lo stesso aminoacido.

Una delle ipotesi che cerca di spiegare la degenerazione del codice asserisce che il numero di codoni per ogni aminoacido è correlato all'abbondanza relativa delle rispettive specie di tRNA (ipotesi dimostrata). Inoltre gli organismi diversi presentano un numero di codoni diverso nei confronti di ogni singolo aminoacido, cioè nelle diverse specie solo alcuni sono rappresentati, mentre gli altri sono di meno o addirittura assenti. In un primo momento venne proposto che per ogni codone esistesse un tRNA dotato di uno specifico anticodone. Se questo fosse vero esisterebbero almeno 61 diversi tRNA mentre altri 3 potrebbero essere complementari per i codoni di terminazione. In seguito, però, vennero prodotte prove che dimostravano che varie specie di tRNA a sequenza nota altamente purificati potevano riconoscere numerosi codoni diversi. Vennero anche scoperti casi in cui una base dell'anticodone non era una delle 4 basi canoniche ma una quinta, l'inosina.

Come tutte le altre basi minori del tRNA, l'inosina nasce da una modificazione enzimatica di una base presente in una catena di tRNA altrimenti completa. La base da cui essa deriva è l'adenina, in cui il carbonio in posizione 6 viene deaminato, formando il gruppo chetonico 6 dell'inosina. Nel 1966 Francis Crick ideò il concetto di appaiamento "tentennante" per spiegare queste osservazioni. Secondo questo concetto la base all'estremità 5' dell'anticodone non subisce le stesse limitazioni spaziali delle altre 2, il che permette di formare legami idrogeno con una qualsiasi delle molte basi posizionate all'estremità 3' del codone. Non tutte le combinazioni sono però possibili. Per esempio, una U nella posizione variabile (tentennante) può appaiarsi sia con l'adenina che con la guanina, mentre I può appaiarsi con U, C o A. Gli appaiamenti consentiti dalle regole del tentennamento sono quelli che permettono distanze ribosio-ribosio più vicine a quelle standard delle coppie di basi A:U e G:C. Le coppie purina-purina o pirimidina-pirimidina produrrebbero distanze ribosio-ribosio, rispettivamente, troppo lunghe o troppo corte. Le regole sul tentennamento non consentono ad un singolo tRNA di riconoscere 4 codoni diversi. Tre codoni possono essere riconosciuti solo quando l'inosina occupa la prima (5') posizione dell'anticodone.

