

ALMA MATER STUDIORUM – UNIVERSITA' DI BOLOGNA  
CAMPUS DI CESENA  
SCUOLA DI INGEGNERIA E ARCHITETTURA  
CORSO DI LAUREA IN INGEGNERIA BIOMEDICA

# **Analisi quantitativa dell'espressione genica mediante real-time RT-PCR**

Elaborato in  
Biochimica

Relatore  
Prof Emanuele D. Giordano

Presentata da  
Alice Andalò

Correlatrice  
Dott.ssa Alice Pasini

III sessione

Anno Accademico 2013/2014



# Indice

<b>Indice</b> .....	<b>i</b>
<b>SOMMARIO</b> .....	<b>iii</b>
<b>1. INTRODUZIONE BIOLOGICA ALL'ESPRESSIONE GENICA</b> .....	<b>1</b>
1.1 Acidi nucleici .....	1
1.1.1 <i>DNA</i> .....	2
1.1.2 <i>RNA</i> .....	4
1.2 Espressione genica .....	6
1.2.1 <i>Replicazione del DNA</i> .....	8
1.2.2 <i>Trascrizione</i> .....	10
1.2.3 <i>Sintesi proteica</i> .....	12
<b>2. OPERAZIONI PRELIMINARI ALL'ANALISI DELL'ESPRESSIONE GENICA</b> .....	<b>15</b>
2.1 Colture cellulari.....	15
2.2 Estrazione RNA.....	16
2.3 Dosaggio allo spettrofotometro.....	18
2.4 Elettroforesi su gel di agarosio .....	19
2.5 Retro-Trascrizione .....	20
<b>3. REAL-TIME PCR</b> .....	<b>23</b>
3.1 Reazione a catena della polimerasi .....	23
3.2 La fluorescenza .....	26
3.3 Real-Time PCR .....	30
3.3.1 <i>SYBR Green</i> .....	30
3.3.2 <i>Sonde</i> .....	32
3.4 Applicazioni.....	35
3.5 Hardware.....	35
3.6 Software .....	38
3.7 Progettazione e ottimizzazione della reazione di PCR.....	40

3.7.1 Qualità dell'RNA .....	40
3.7.2 Protocollo termico di PCR .....	40
3.7.3 Disegno dei Primer .....	41
3.7.4 Importanza dei controlli .....	42
3.7.5 Riproducibilità, Sensibilità e Specificità .....	42
<b>4. ANALISI DEL SEGNALE .....</b>	<b>45</b>
4.1 Curva di amplificazione .....	45
4.1.1 Linea Threshold .....	46
4.1.2 Ciclo Soglia .....	47
4.2 Curve di Melting .....	47
4.3 Analisi quantitativa .....	49
4.4 Quantificazione assoluta .....	50
4.4.1 Curve Standard .....	50
4.5 Quantificazione relativa .....	52
4.5.1 Metodo del $\Delta CT$ .....	54
4.5.2 Metodo Livak o del $\Delta\Delta CT$ .....	55
4.5.3 Metodo Pfaffl .....	57
4.6 MIQE .....	59
4.7 Significatività statistica .....	59
<b>5. CASO DI STUDIO: ANALISI DI ESPRESSIONE GENICA TRAMITE REAL-TIME PCR .....</b>	<b>61</b>
5.1 Scopo e metodi .....	61
5.2 Analisi dell'espressione genica .....	62
<b>Conclusioni .....</b>	<b>65</b>
<b>BIBLIOGRAFIA .....</b>	<b>66</b>
<b>SITOGRAFIA .....</b>	<b>68</b>

# SOMMARIO

Il seguente lavoro di tesi deriva dall'approfondimento dell'esperienza di tirocinio curriculare svolto presso il laboratorio di Ingegneria Cellulare e Molecolare (ICM) "S. Cavalcanti", della Scuola di Ingegneria e Architettura dell'Università di Bologna. Lo scopo della tesi è la descrizione dell'analisi di espressione genica, effettuata tramite real-time Polymerase Chain Reaction (PCR).

L'analisi di espressione genica permette di descrivere il fenotipo cellulare, cioè la manifestazione delle proteine che vengono effettivamente prodotte dalla cellula nell'istante in cui si sta eseguendo l'indagine. Essa si affianca alle analisi genetiche che indagano invece il genotipo cellulare, cioè il corredo genetico che descrive tutte le proteine potenzialmente producibili da una cellula.

Misurare l'espressione di un gene consiste nel dare una valutazione quantitativa della presenza di trascritti, cioè le molecole di RNA messaggero, o dei tradotti, cioè le specifiche proteine da essi codificati. Il lavoro di tesi si limita all'analisi dei profili di espressione genica basata sulla quantificazione dei livelli di molecole trascritte di mRNA. I geni contengono le informazioni per produrre mRNA, ma in ogni momento ogni cellula produce mRNA solo da una frazione di geni. Se un gene viene usato per produrre mRNA, è considerato acceso (espresso), altrimenti risulta spento. Molti fattori determinano lo stato di accensione/spegnimento di un gene, e quindi la possibile presenza/assenza della proteina corrispondente. Gli esperimenti di analisi di espressione genica spesso implicano la misura della quantità relativa di mRNA specifici derivanti da due o più condizioni sperimentali. Questo perché livelli alterati di una specifica sequenza di mRNA, indice di una variazione dei livelli della corrispondente proteina, fanno ipotizzare un cambiamento delle necessità cellulari. Uno dei metodi utilizzati per esaminare il livello di espressione di geni selezionati è la reverse transcriptase-Polymerase Chain Reaction (RT-PCR), la quale permette l'amplificazione di sequenze specifiche di mRNA. Questa procedura unita alla tecnica di real-time PCR consente un'accurata quantificazione del livello di espressione di geni di interesse.

Il primo capitolo di questo lavoro di tesi introduce i concetti di biologia necessari per comprendere il fenomeno dell'espressione genica. Il secondo capitolo descrive i metodi e le tecniche di laboratorio utilizzate per ottenere il cDNA, il

materiale genetico che verrà amplificato nella real-time PCR. Nel terzo capitolo si descrive la tecnica di real-time PCR, partendo da una descrizione della PCR convenzionale fino a delineare le caratteristiche della sua evoluzione in real-time PCR. Si prosegue con la spiegazione del principio fisico alla base della tecnica e delle molecole necessarie (*fluorofori* e *sonde*) per realizzarla; infine si descrive l'hardware e il software dello strumento. Il quarto capitolo presenta le tecniche di analisi del segnale che utilizzano metodi di quantificazione assoluta o relativa. Infine nel quinto capitolo è presentato un caso di studio, cioè un'analisi di espressione genica con real-time PCR condotta durante l'esperienza di tirocinio presso il laboratorio ICM.

# 1. INTRODUZIONE BIOLOGICA ALL'ESPRESSIONE GENICA

Il processo attraverso cui l'informazione è trasferita dai geni alle proteine è detto espressione genica: esso è responsabile della specializzazione delle cellule di uno stesso organismo (fenotipo) dotate dello stesso corredo genetico (genotipo) [Campbell et al, 2008].

Per chiarire il concetto di espressione genica, di seguito verranno descritti la struttura delle biomolecole coinvolte nel processo, gli acidi nucleici, e i principali meccanismi cellulari che regolano il processo.

## 1.1 Acidi nucleici

Esistono due tipi di acidi nucleici: **DNA** e **RNA**. Le unità strutturali degli acidi nucleici sono i **nucleotidi**, i quali sono caratterizzati da tre componenti: una *base azotata*, contenente azoto e a struttura eterociclica; uno *zucchero pentoso*, ovvero a cinque atomi di carbonio e infine un *gruppo fosfato*. Quest'ultimo è un gruppo funzionale chimico costituito dall'anione fosfato ( $PO_4^{3-}$ ) legato a una molecola organica tramite uno dei suoi quattro ossigeni. La molecola formata da zucchero pentoso e base azotata prende il nome di nucleoside.

I nucleotidi sono legati tra loro tramite "ponti" covalenti, in cui il gruppo fosfato legato al carbonio dello zucchero in posizione 5' di un'unità nucleotidica reagisce con il gruppo ossidrilico (-OH) legato al carbonio in posizione 3' del nucleotide successivo, formando così un **legame fosfodiesterico**. Lo scheletro degli acidi nucleici è costituito da un'alternanza di gruppi fosforici e di residui di zucchero pentoso, mentre le basi azotate possono essere considerate come gruppi laterali uniti allo scheletro a intervalli regolari. I gruppi ossidrilici degli zuccheri formano legami idrogeno con l'acqua e per questo lo scheletro di DNA e RNA risulta essere idrofilico. I gruppi fosforici dello scheletro polare sono completamente ionizzati e carichi negativamente a pH 7.

Tutti i legami fosfodiesterici delle catene hanno lo stesso orientamento lungo la catena, determinando per ogni acido nucleico una specifica polarità, ed estremità 3' e 5' distinte.

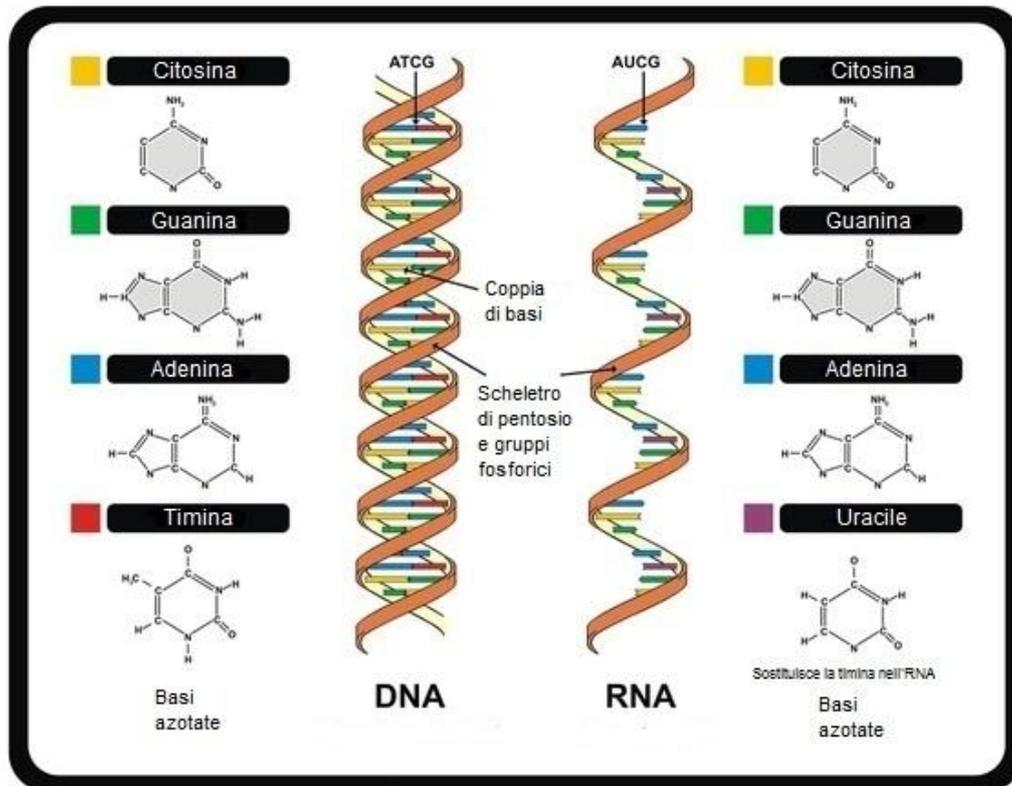


Figura 1.1 Composizione e struttura di DNA e RNA.

### 1.1.1 DNA

L'acido desossiribonucleico o deossiribonucleico (**DNA**) è un acido nucleico che ha la funzione di codificare e trasmettere l'informazione genetica. Infatti contiene le informazioni indispensabili per la biosintesi delle proteine.

Chimicamente il DNA è un polimero di nucleotidi legati tra loro. L'ordine nella disposizione sequenziale dei nucleotidi costituisce l'informazione genetica (gene), la quale è tradotta nelle macromolecole che realizzano la specifica funzione codificata, cioè le proteine. I nucleotidi del DNA sono caratterizzati da uno zucchero pentoso, il 2'*deossi-ribosio*, chiamato così per la mancanza di un atomo di ossigeno in posizione 2', unito a una delle quattro basi azotate tipiche del DNA e a un gruppo fosfato. Le basi azotate possono essere suddivise in *basi puriniche* e *basi pirimidiniche*, a seconda della struttura biciclica o monociclica che posseggono. Infatti le purine sono basi con struttura ciclica a due anelli, mentre le pirimidine hanno una struttura ciclica con un solo anello. Sono basi puriniche l'*adenina* e la *guanina*, indicate rispettivamente con (A) e (G), mentre la *timina* (T) e la *citosina* (C) rappresentano le basi pirimidiniche del DNA.

Un importante passo in avanti per comprendere la struttura del DNA derivò dal lavoro di *Erwin Chargaff*, biochimico austriaco che stabilì che le molecole di DNA hanno composizioni in basi distinte. I risultati ottenuti portarono Chargaff a concludere che in tutte le molecole di DNA il numero di residui di adenina è pari al numero di residui di timina ( $A = T$ ); e analogamente il numero dei residui di guanina e citosina coincidono ( $G = C$ ). Ciò porta a concludere che la somma dei residui purinici è pari a quella dei residui pirimidinici: ( $A + G = T + C$ ).

Due chimici inglesi, *Rosalind Franklin* e *Maurice Wilkins*, analizzarono i cristalli di DNA attraverso il metodo della diffrazione a raggi X. Attraverso questo metodo dimostrarono che il DNA produce una diffrazione caratteristica. I ricercatori notarono che i polimeri di DNA erano elicoidali con due periodicità lungo il loro asse, una primaria di 0,34 nm (3,4 Å) e una secondaria di 3,4 nm (34 Å).

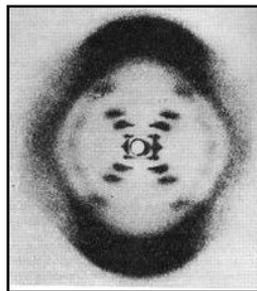


Figura 1.2 Immagine ai raggi X del DNA.

Inoltre fu grazie a queste immagini che *James Watson* e *Francis Crick* intuirono il **modello a doppia elica** della molecola di DNA: nel 1953 proposero un modello tridimensionale del DNA che poteva spiegare tutte le proprietà fino ad allora note [J.D. Watson et al., 2009].

Dal modello si deduce che il DNA è costituito da due catene elicoidali avvolte attorno allo stesso asse per formare una doppia elica destrorsa. Lo scheletro covalente idrofilico, costituito da un'alternanza di deossiribosio e gruppi fosforici è all'esterno della doppia elica, mentre le basi puriniche e pirimidiniche sono impilate all'interno della doppia elica, e le loro strutture planari ad anello sono poste in posizione perpendicolare al lungo asse longitudinale della molecola. Si crea una relazione spaziale tra le due catene che genera una scanalatura maggiore e una scanalatura minore sulla superficie della doppia elica. La scanalatura maggiore è larga 22Å, e quella minore 12Å. La differente ampiezza

dei due solchi determina una differente accessibilità delle basi. I fattori proteici che regolano il processo della trascrizione genica (vedi paragrafo 1.2.2) solitamente prendono contatto con le basi presenti nel solco maggiore [D.L. Nelson et al, 2002]. I due filamenti costituenti la doppia elica hanno un'estensione antiparallela, ovvero il senso di un filamento è opposto a quello dell'altro filamento, ed entrambi sono orientati in direzione 5'-3'.

Le due catene polinucleotidiche, antiparallele, sono tra loro complementari. Le basi puriniche formano legami idrogeno con le basi pirimidiniche: A può legare solo T e G può legare solo C, con due o tre legami idrogeno rispettivamente. Questi legami idrogeno che si formano implicano la complementarità tra le catene di DNA. L'accoppiamento di due basi viene comunemente chiamato "*paio di basi*"; esso costituisce l'unità di misura maggiormente utilizzata per definire la lunghezza di una molecola di DNA.

Un'altra forza che mantiene unite le due catene è dovuta all'interazione tra le basi impilate. Queste interazioni non sono specifiche ma sono coinvolte nella stabilità della doppia elica. Dato che i legami idrogeno sono legami deboli, essi possono essere rotti e riuniti in modo relativamente semplice. I due filamenti possono essere separati tra loro (processo di denaturazione del DNA) sia utilizzando alte temperatura che sfruttando un'azione meccanica. La temperatura alla quale il 50% del DNA è denaturato è detta "**temperatura di melting**" ( $T_m$ ), ed è funzione della sequenza di nucleosidi; infatti se sono presenti molte coppie (C)-(G) la  $T_m$  risulterà più alta, rispetto a una sequenza con più coppie (A)-(T), perché nel primo caso è necessario rompere tre legami a idrogeno, rispetto ai due del secondo. Indicativamente la  $T_m$  di una molecola di DNA è circa uguale a 90°C, con l'abbassamento di temperatura le basi si riappaiano.

### **1.1.2 RNA**

L'acido ribonucleico (**RNA**) è un polimero organico, risultante dalla polimerizzazione dei ribonucleotidi. Chimicamente l'RNA e il DNA sono simili, infatti entrambi sono formati da una successione di nucleotidi. Esistono però alcune differenze tra i due acidi nucleici: lo zucchero pentoso caratteristico dei nucleotidi di RNA è il *ribosio*, il quale ha un ossigeno in posizione 2', mancante nello zucchero pentoso del DNA. Inoltre una delle basi, la *timina*, è sostituita

dall'*uracile* (base pirimidinica), che si lega all'adenina tramite due legami idrogeno, analogamente alla timina. La struttura dell'RNA è solitamente costituita da un *unico filamento*, a differenza della struttura a doppia elica del DNA.

Un'altra differenza tra i due acidi nucleici riguarda gli aspetti funzionali. Il DNA svolge essenzialmente la funzione di conservare e trasmettere l'informazione, mentre esistono diversi tipi di RNA che svolgono diverse funzioni. Le principali categorie di RNA sono:

- **RNA messaggero (mRNA):** l'mRNA è una copia di una sequenza di DNA codificante per un prodotto biologicamente attivo, definito gene. Esso trasporta nella sua sequenza codificante l'informazione operativa contenuta nel DNA dal nucleo al citoplasma per la sintesi delle proteine. Diversamente dal DNA, l'mRNA è una molecola "instabile", a causa dell'ossidrilico in posizione 2'; questa caratteristica è associata al suo utilizzo temporaneo. Le dinamiche della sua sintesi e degradazione determinano la presenza/assenza nella cellula del prodotto attivo di ogni specifico gene. Il numero di copie di mRNA trascritte (e l'efficienza della loro successiva traduzione) influenza il quantitativo della proteina. Le due estremità del RNA messaggero sono diverse per rendere riconoscibile il corretto senso di lettura.
- **RNA di trasferimento (tRNA):** decifra la codifica e traduce le informazioni dell'mRNA in amminoacidi, le unità costitutive delle proteine. Il tRNA associa l'amminoacido specifico per la tripletta di nucleotidi del mRNA, chiamata codone, tramite un anticodone complementare.
- **RNA ribosomiale (rRNA):** forma insieme a specifiche proteine i ribosomi (organelli cellulari che operano la sintesi delle proteine nelle cellule). I ribosomi sono formati da due subunità, una maggiore detta 60s e una minore 40s.

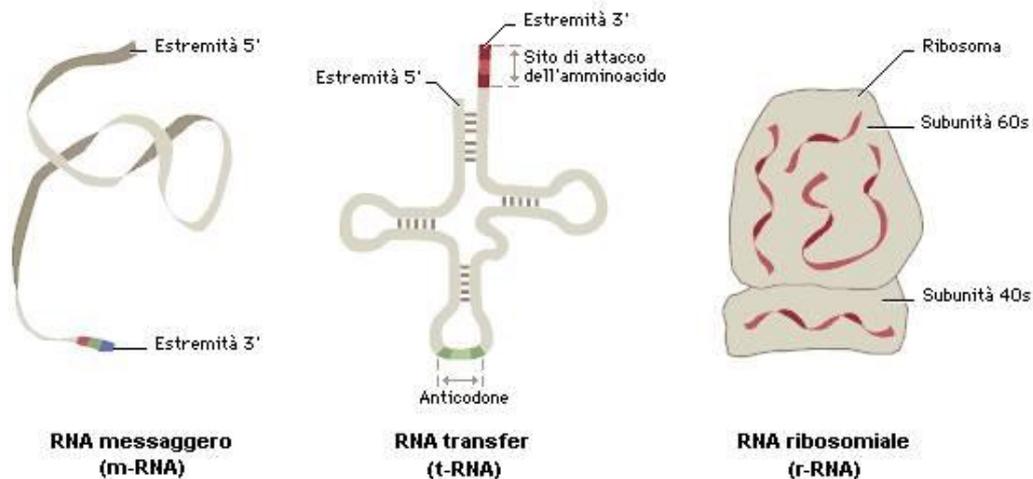


Figura 1.3 Descrizione delle strutture dei diversi tipi di RNA.

## 1.2 Espressione genica

La molecola di DNA all'interno della cellula è organizzata in strutture chiamate **cromosomi**. Un singolo cromosoma solitamente contiene migliaia di geni. L'insieme di DNA genico e intergenico (DNA che si localizza tra i diversi geni) è chiamato *genoma cellulare*.

La conoscenza dei geni è aumentata notevolmente nel corso dell'ultimo secolo. Inizialmente il gene era definito come la porzione di un cromosoma che influenza un singolo fenotipo (caratteristica visibile dell'organismo). Successivamente, dopo diversi esperimenti di *Beadle* e *Tatum* con i raggi X, venne proposta l'ipotesi di "un gene-una proteina" ovvero si definì il gene come quel segmento di materiale genetico che codifica una proteina.

La definizione biochimica moderna di gene è ancora più precisa e tiene conto del fatto che molte proteine sono formate da catene polipeptidiche multiple. In alcune proteine a catena multiple, tutte le catene sono identiche e vengono codificate da uno stesso gene. Altre proteine sono composte da più tipi diversi di catene polipeptidiche, ognuna con una sequenza amminoacidica specifica. La relazione tra gene e proteina allora viene identificata più precisamente con il concetto "un gene-un polipeptide". Oltre ai geni, il DNA contiene anche sequenze regolatrici

che forniscono segnali di inizio e fine dei geni o ne influenzano la trascrizione. [D.L. Nelson et al, 2002]

Con il termine **espressione genica** si definisce il processo attraverso cui l'informazione contenuta in un gene (costituito di DNA) viene convertita in una macromolecola funzionale (tipicamente una proteina), attraverso i meccanismi di trascrizione del DNA in mRNA e poi di traduzione dell'mRNA in proteina. Non tutti i geni sono sempre necessari alla vita di una cellula: alcuni geni, chiamati costitutivi, sono espressi sempre; altri invece solo quando si presenta l'esigenza. L'espressione dei geni è quindi regolata dalle necessità della cellula in base alle funzioni richieste e alle condizioni ambientali.

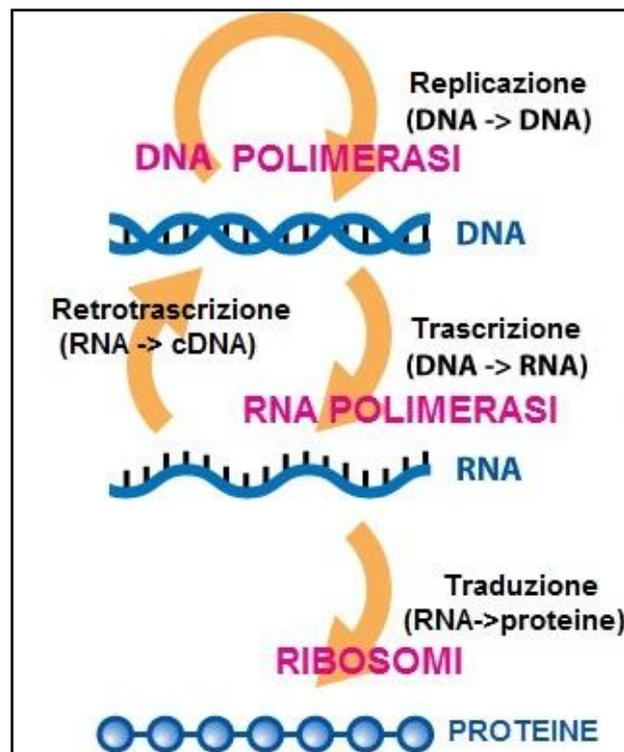


Figura 1.4 Descrizione sintetica dei processi di utilizzo dell'informazione genetica: "Dogma centrale della biologia molecolare".

### 1.2.1 *Replicazione del DNA*

Dopo le scoperte di Watson e Crick sulla struttura del DNA si capì come esso potesse fungere da stampo per la replicazione e la trasmissione dell'informazione genetica. Le rigorose regole dell'appaiamento delle basi spiegano come si possa utilizzare uno dei due filamenti di DNA come stampo per produrre un'altra catena di DNA con una sequenza complementare alla prima.

La replicazione del DNA viene perciò definita *semiconservativa*: le due molecole di DNA prodotte conterranno un filamento parentale, che ha svolto la funzione di stampo, e un filamento ad esso complementare neosintetizzato. Questa ipotesi, proposta da Watson e Crick, fu poi dimostrata con diversi esperimenti dagli studiosi *Meselson* e *Stahl* nel 1957. La replicazione è un processo coordinato, in cui le catene di DNA parentali vengono svolte, separate e replicate simultaneamente, a partire da siti specifici chiamati *forcelle di replicazione*. La sintesi di una nuova catena di DNA procede sempre in direzione  $5' \rightarrow 3'$ . Poiché le due eliche di DNA sono antiparallele l'elica usata come stampo viene letta dall'estremità 3' verso quella 5'.

Nel 1960 il ricercatore giapponese *Reiji Okazaki* scoprì che una delle due nuove catene viene sintetizzata sotto forma di brevi frammenti, che presero il nome di "*frammenti di Okazaki*". Quindi si intuì che una delle due catene viene sintetizzata in modo continuo mentre l'altra in modo discontinuo. La catena continua, chiamata anche catena veloce, è quella in cui la sintesi procede nella stessa direzione del movimento della forcella di replicazione, seguendo anche il senso  $5' \rightarrow 3'$ . L'altra catena, chiamata anche catena lenta, è quella in cui la sintesi procede in senso opposto rispetto al movimento della forcella.

Il DNA viene sintetizzato grazie a un enzima chiamato DNA polimerasi, scoperto nel 1955 da *Arthur Kornberg*, premio Nobel per la medicina. La **DNA polimerasi** catalizza la polimerizzazione dei nucleotidi, uno alla volta, in corrispondenza del carbonio in posizione 3' del nucleotide della nuova molecola. Gli studi sulla DNA polimerasi portarono a comprendere la necessità di avere delle molecole d'innescò per il processo di sintesi, queste molecole sono chiamate **primer**. Un primer è un piccolo segmento di acidi nucleici, complementare alla catena stampo, con un gruppo ossidrilico 3' libero a cui si possano legare i nucleotidi. I primer sono spesso oligonucleotidi, composti da RNA, sintetizzati da enzimi particolari [D.L. Nelson et al, 2002]. Dopo l'aggiunta di un nucleotide alla catena di

DNA neosintetizzata, la DNA polimerasi si muove lungo il filamento stampo e aggiunge un ulteriore nucleotide. Il numero medio di nucleotidi aggiunti prima che la polimerasi si dissocia dal DNA determina un parametro chiamato processività. Esistono in natura diverse DNA polimerasi, ognuna caratterizzata da una propria processività.

La replicazione procede per fasi successive. Possiamo suddividere la sintesi di una molecola di DNA in tre fasi: *inizio*, *allungamento* e *terminazione*. Queste tre fasi sono distinguibili sia per le reazioni che per gli enzimi che intervengono.

Inizialmente si ha l'apertura della doppia elica a causa della rottura dei legami idrogeno operata da enzimi specifici. La denaturazione dei filamenti agisce specialmente nelle regioni ricche di coppie (A)-(T). Le due catene quindi si separano lasciando disaccoppiate le basi azotate di ciascuna di esse. Nella fase di allungamento si ha il processo di sintesi della catena lenta e della catena veloce. Infine si ha il processo di terminazione in cui le forcelle di replicazione si incontrano in una regione terminale, dove è presente una sequenza di basi particolare chiamata *Ter*.

Il processo di replicazione è molto accurato e gli errori sono minimi; infatti anche se un nucleotide sbagliato potrebbe riuscire a formare legami idrogeno con un'altra base nello stampo, in realtà non riesce ad entrare nel sito attivo della DNA polimerasi, quindi i nucleotidi sbagliati vengono solitamente eliminati prima che si crei il legame fosfodiesterico.

Nonostante i minimi errori (un nucleotide sbagliato su 100 milioni appaiati correttamente) possono capitare degli accoppiamenti sbagliati che provocano un cambiamento nella sequenza delle basi azotate. Se il cambiamento è replicato, e trasmesso alle cellule future, si parla di *mutazioni* e la modifica risulta permanente.

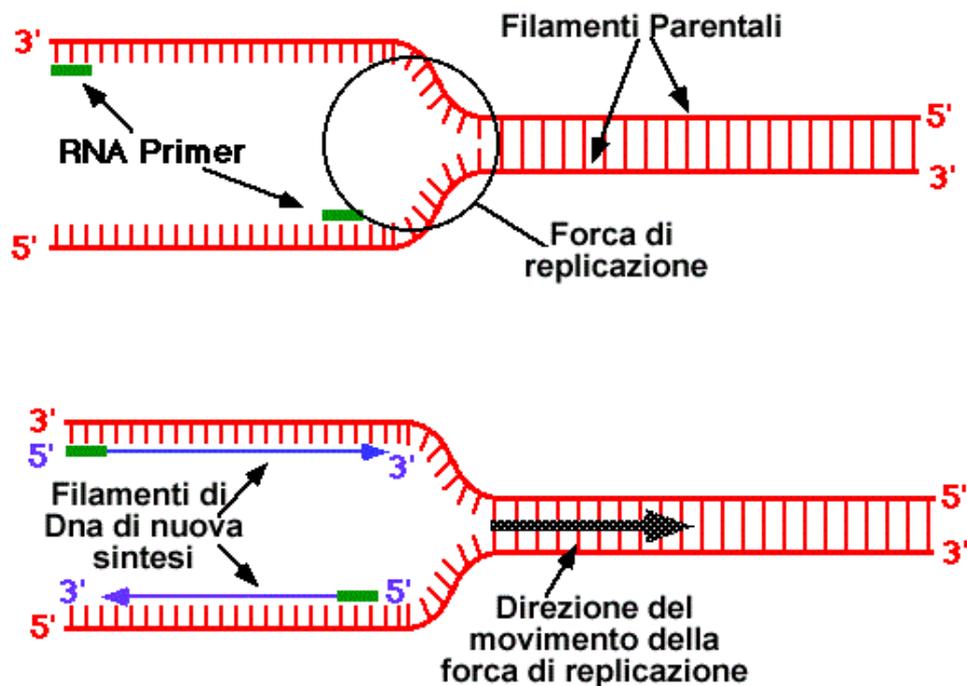


Figura 1.5 Schematizzazione del processo di replicazione del DNA

### 1.2.2 Trascrizione

La **trascrizione** è il processo di copia, in forma di RNA, di specifiche parti dell'informazione genetica contenuta nel DNA. A differenza del processo di replicazione del DNA, la trascrizione è molto *selettiva*: infatti, a seconda della necessità, vengono trascritti soltanto particolari sequenze di DNA.

La trascrizione è simile alla replicazione per quanto riguarda il meccanismo chimico e l'utilizzo di un filamento di DNA come stampo. Inoltre anche per la trascrizione la direzione di sintesi procede in direzione  $5' \rightarrow 3'$ . A differenza della replicazione, la trascrizione, non necessita di punti d'innescio (*primer*) e, come già detto, interessa solo brevi porzioni di DNA. La realizzazione del processo di trascrizione è operata da un enzima: la **RNA polimerasi**. La doppia elica di DNA viene svolta e separata per opera di specifici enzimi. L'RNA polimerasi utilizza il filamento di DNA come stampo per produrre un filamento complementare di RNA, che conterrà invece che basi Timina, Uracile. Il posizionamento corretto dell'RNA polimerasi all'inizio della sequenza codificante da trascrivere, il gene, è

determinato da regioni regolatrici chiamate **promotore**, poste a monte del tratto codificante. La trascrizione di ciascun gene è accuratamente regolata, in particolar modo nella fase iniziale. L'RNA polimerasi è guidata da fattori cooperativi detti "*fattori di trascrizione*", i quali interagiscono con tipiche sequenze di consenso, posizionate nei pressi del punto di inizio della trascrizione. A seguito dell'interazione tra un fattore di trascrizione (TF) e la sua sequenza di consenso, il TF determina una distorsione della topologia del duplex che lo renderà accessibile al complesso trascrizionale. Negli eucarioti l'inizio della trascrizione necessita di ulteriori *proteine attivatrici*.

Dopo che l'RNA polimerasi ha creato una copia della sequenza contenente il gene da trascrivere, essa si stacca e l'RNA viene rilasciato. Una molecola di RNA appena sintetizzata viene denominata **trascritto primario**. La sua maturazione genera una molecola funzionale di mRNA che codifica per una singola proteina. Il trascritto è composto da sequenze codificanti, chiamate *esoni*, interposte a tratti non codificanti, chiamati *introni*. Gli mRNA eucariotici, durante la maturazione, vengono modificati alle estremità: viene aggiunta all'estremità 5' una struttura, chiamata "cappuccio", che protegge l'RNA dalla degradazione enzimatica.

L'estremità 3' viene tagliata e ad essa si aggiunge una serie di 80-250 residui di (A). Questa serie di residui prende il nome di *coda di poli(A)* e porta a influenzare l'emivita della molecola. Durante la fase finale della maturazione dell'mRNA, in un processo denominato *splicing*, gli introni vengono rimossi dal trascritto primario e gli esoni vengono uniti formando così una sequenza continua. Le fasi della maturazione dell'mRNA, negli eucarioti, sono contemporanee alla sua esportazione dal nucleo al citoplasma, dove sono localizzati i ribosomi, dove verrà tradotto in una sequenza di amminoacidi [D.L. Nelson et al,2011].

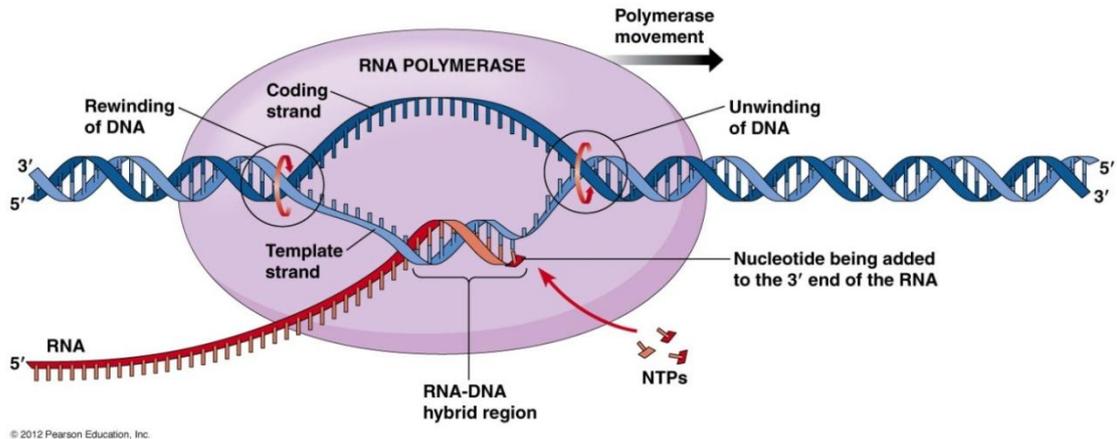


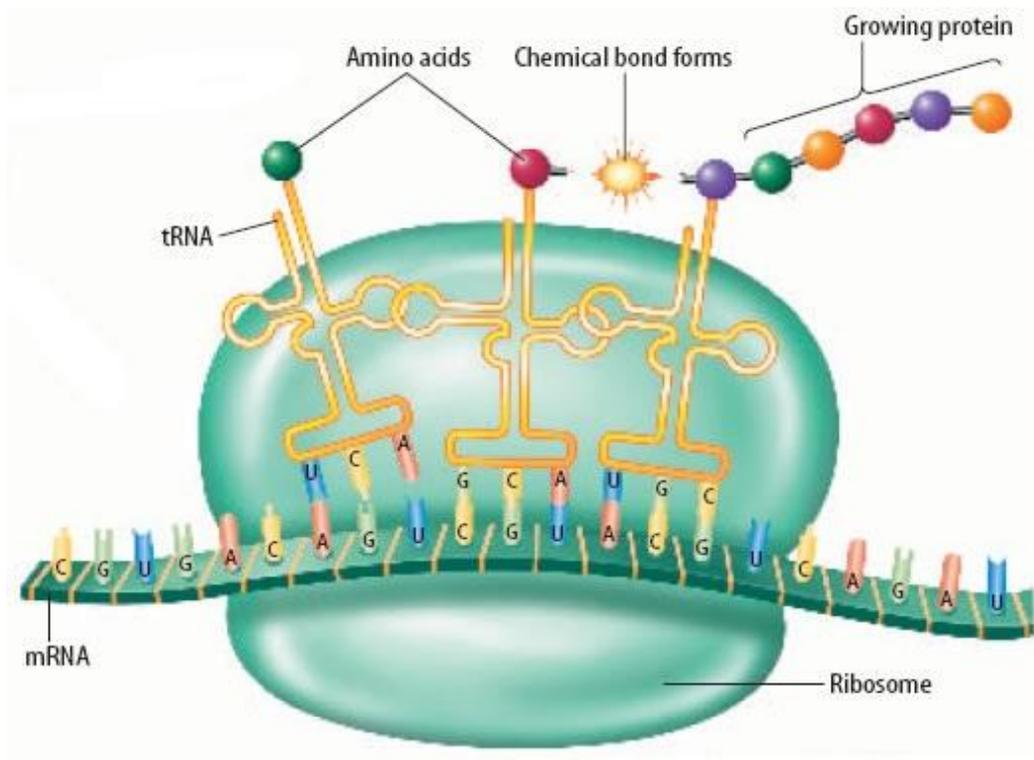
Figura 1.6 Processo di trascrizione.

### 1.2.3 Sintesi proteica

La **sintesi proteica**, chiamata anche **traduzione**, è il processo biochimico attraverso il quale l'informazione genetica contenuta nel mRNA viene convertita in una sequenza di amminoacidi, costituenti una proteina funzionale specifica. Il processo di traduzione avviene nei *ribosomi*, localizzati nel citoplasma. Il processo inizia quando l'mRNA si lega alla subunità minore del ribosoma, il quale decifra l'informazione genetica leggendo una sequenza di tre nucleotidi (*tripletta*) per volta. Ogni tripletta di nucleotidi, definita **codone**, corrisponde a uno specifico amminoacido. Gli amminoacidi sono trasportati da molecole di tRNA, le quali hanno un anticodone complementare alla tripletta dell'mRNA. I tRNA posizionano gli amminoacidi in sequenza per essere uniti con un legame peptidico. La catena polipeptidica nascente si allunga per unione di unità amminoacidiche successive, ciascuna trasportata verso il ribosoma e correttamente posizionata dal rispettivo tRNA. L'allungamento viene coordinato da proteine citosoliche chiamate *fattori di allungamento*. Il completamento della catena è segnalato da un codone di terminazione dell'mRNA. A questo punto la catena viene rilasciata dal ribosoma, con l'aiuto di specifici fattori di rilascio.

Per acquisire la sua forma biologicamente attiva, il polipeptide deve ripiegarsi nella sua conformazione tridimensionale e subire delle modifiche post-traduzionali. Questi cambiamenti consistono in modificazioni enzimatiche come

la rimozione di amminoacidi alle estremità o l'aggiunta di gruppi chimici funzionali. Questa fase finale avviene in compartimenti subcellulari specializzati come il reticolo endoplasmatico o l'apparato di Golgi. Al termine della loro maturazione, le proteine, vengono veicolate nella loro sede operativa tramite delle vescicole create dall'apparato di Golgi.



**Figura 1.7** Schema del processo di sintesi proteica.



## 2. OPERAZIONI PRELIMINARI ALL'ANALISI DELL'ESPRESSIONE GENICA

Per poter effettuare un'analisi di espressione genica in laboratorio tramite la tecnica real-time RT-PCR, è necessario prima di tutto isolare l'RNA cellulare con un protocollo di estrazione dalla popolazione cellulare di interesse. Si procede poi con la valutazione della sua qualità e purezza e si determina la concentrazione dell'estratto tramite spettrofotometro. Infine, dopo aver retro-trascritto l'RNA in cDNA (DNA complementare), una molecola più stabile, si può procedere all'analisi di espressione genica tramite real-time PCR.

### 2.1 Colture cellulari

La popolazione cellulare, da cui verrà estratto l'RNA, può essere una coltura cellulare costituita da un insieme omogeneo di cellule provenienti da tessuti animali. Nei laboratori solitamente si utilizzano delle linee cellulari commerciali immortalizzate.

Queste sono cellule derivate da tumori oppure cellule che hanno subito una modificazione genetica che le ha rese capaci di crescere indefinitamente (immortalizzazione) e possono costituire un modello sperimentale di cui si vuole studiare e modificare il comportamento. Negli organismi viventi sono presenti diversi meccanismi di regolazione che controllano le funzioni vitali e mantengono in omeostasi temperatura e pH. Per ottenere un modello cellulare conforme è importante conservare le stesse condizioni fisico-chimiche nella coltura cellulare in vitro.

Le colture vengono mantenute in un terreno liquido che fornisce gli elementi essenziali per il metabolismo cellulare. Esistono diversi tipi di terreni di coltura ma ognuno è sostanzialmente formato da una soluzione isotonica e tamponata contenente *sali inorganici*, i quali riforniscono le cellule di ioni come  $Na^+$ ,  $K^+$ ,  $Mg^{2+}$ ,  $Ca^+$ ,  $Cl^-$ , per mantenere il bilancio osmotico. Nella soluzione sono presenti gli *aminoacidi essenziali*, indispensabili per la sintesi proteica, e il *glucosio*, che interviene nelle funzioni metaboliche come fonte energetica. Anche

il *siero* (fonte di proteine e fattori di crescita), o un suo analogo sintetico, è un costituente importante del terreno.

Il pH del terreno di coltura è mantenuto al valore di 7.4 grazie al tampone bicarbonato di sodio/acido carbonico. Di solito i terreni sono addizionati di un colorante il Rosso Fenolo, che funge da indicatore di pH del terreno, perché il suo colore vira dal rosso, indice di pH neutro, al giallo, quando il pH diventa acido, o al fucsia, in condizioni di pH basico.

Le cellule vengono fatte crescere in contenitori di plastica monouso sterili, la cui superficie è stata opportunamente trattata per favorire l'adesione delle cellule. Le piastre più utilizzate sono le piastre Petri, solitamente rotonde di varie dimensioni, o le piastre multi pozzetto, cioè piastre rettangolari suddivise in diversi pozzetti (da 4 a 96) di volume variabile a seconda del numero di pozzetti.

Una condizione necessaria per il mantenimento delle colture cellulari è la sterilità. Per evitare di contaminare le colture cellulari sono necessari diversi accorgimenti: solitamente nei terreni di coltura vengono addizionati degli antibiotici o antimicotici; i materiali che vengono a contatto con le cellule sono sterili e usualmente monouso; inoltre le manipolazioni delle cellule vengono svolte in cappe a flusso laminare dotate di filtri, che consentono di lavorare in condizioni di sterilità.

Le colture cellulari vengono mantenute in un incubatore, il quale mantiene controllate le condizioni di temperatura e atmosfera. La temperatura viene mantenuta costante a 37°C, grazie a un termostato. La concentrazione di  $CO_2$  è settata al 5%, per tamponare il pH del terreno di coltura sfruttando il sistema tampone bicarbonato di sodio ( $NaHCO_3$ )/acido carbonico presente nel terreno.

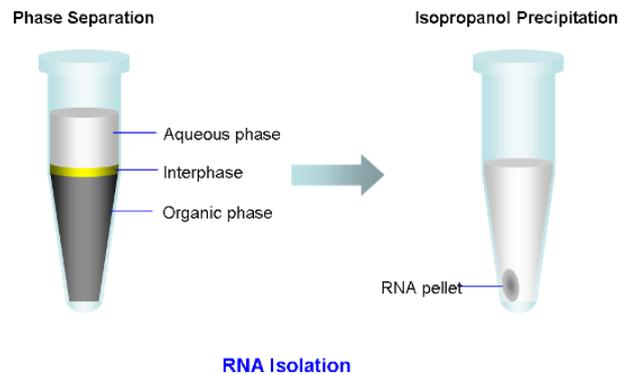
## 2.2 Estrazione RNA

Il protocollo di laboratorio di estrazione di RNA si può suddividere in quattro fasi principali:

1. **Preparazione dei campioni:** per estrarre l'RNA da cellule adese su piastra, è necessario distaccarle utilizzando uno scraper. Poi si aspira la soluzione e si centrifuga in modo da separare il terreno dalle cellule. Si

risospinde il pellet, composto dalle sole cellule, in Trizol (una sostanza a base di fenolo).

2. **Separazione di fase**: attraverso una centrifugazione ad alta velocità la miscela si separa in una fase organica apolare composta da fenolo-cloroformio, un'interfase contenente il DNA e una fase acquosa superiore. L'RNA si trova esclusivamente nella fase acquosa.
3. **Precipitazione e lavaggio**: Si trasferisce quindi la fase acquosa in una nuova provetta; per evitare il rischio di raccogliere anche componenti dell'interfase, si aspira solo il 60% della soluzione contenente RNA. Si esegue quindi la precipitazione dell'RNA aggiungendo isopropanolo. Dopo aver incubato e centrifugato la soluzione si otterrà un pellet di RNA. Dopo aver rimosso il surnatante si effettua un lavaggio con etanolo in modo da rimuovere i sali.



**Figura 2.1 Separazione di fase e precipitazione dell'RNA con isopropanolo**

4. **Solubilizzazione**: Nell'ultima fase si porta a secco il preparato rimuovendo la parte liquida. Infine si discioglie il pellet di RNA in acqua.

## 2.3 Dosaggio allo spettrofotometro

Il dosaggio dell'estratto di RNA allo spettrofotometro è eseguito per ricavare la concentrazione del campione e valutarne la purezza. Gli acidi nucleici infatti assorbono la luce in uno spettro che varia da 230nm a 280nm con un picco di assorbimento in corrispondenza di 260nm. Tale assorbimento è determinato dalle basi azotate e per ciò è caratteristico sia del DNA che dell'RNA.

Per calcolare l'assorbanza a una data lunghezza d'onda  $\lambda$  si utilizza la **legge di Lambert-Beer**, espressa come:

$$A_{\lambda} = \log\left(\frac{I_0}{I}\right) = \varepsilon c l$$

Dove  $I_0$  e  $I$  rappresentano rispettivamente l'intensità della luce incidente e l'intensità di luce trasmessa;  $\varepsilon$  è il coefficiente di estinzione molare;  $c$  rappresenta la concentrazione della specie che assorbe la luce e  $l$  il cammino ottico. La legge di Lambert-Beer dimostra che la frazione di luce incidente, assorbita da una soluzione a una data lunghezza d'onda, è proporzionale allo spessore della soluzione (*cammino ottico*) e alla concentrazione della specie chimica che assorbe la luce.

Dalla misura di assorbanza si ricava la concentrazione incognita  $C$  del campione; considerando un cammino ottico pari a 1 cm (dato dalla dimensione della cuvetta utilizzata per la misura del campione) il coefficiente di estinzione molare medio del RNA  $\varepsilon_{RNA} = 1/40 (\mu\text{g}/\text{mL})^{-1} \text{ cm}^{-1}$  e il fattore di diluizione del campione  $D$ , si ricava:

$$C [\mu\text{g}/\text{mL}] = A_{260} D / l \varepsilon_{RNA}$$

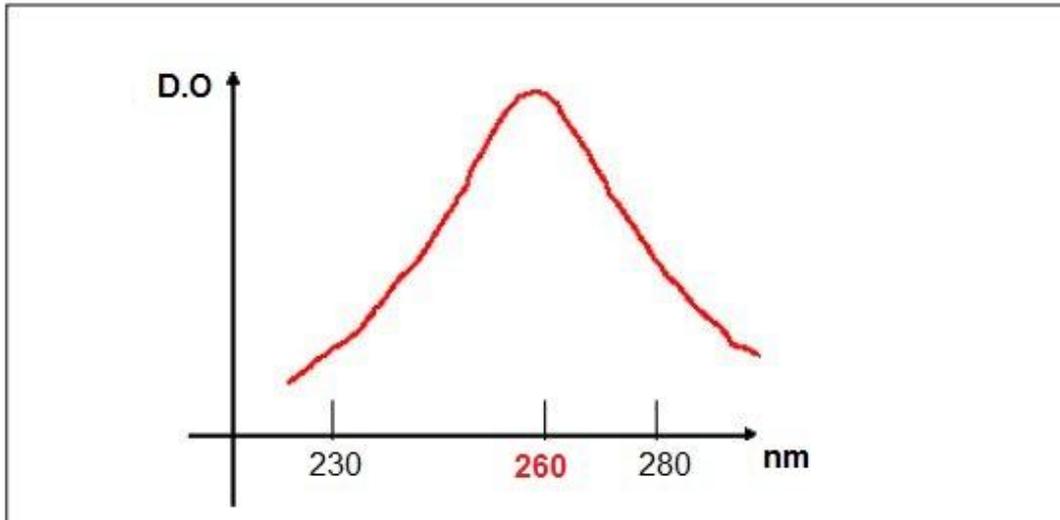


Figura 2.2 Curva dell'assorbanza, misurata in densità ottica, per gli acidi nucleici.

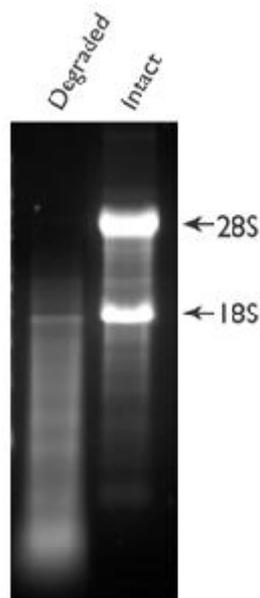
Per determinare la purezza del campione si misurano due rapporti che indicano rispettivamente la purezza da fenoli ( $\frac{A_{260}}{A_{230}}$ ) e quella da proteine ( $\frac{A_{260}}{A_{280}}$ ); campioni con buoni indici di purezza devono essere caratterizzati da un valore del rapporto  $\frac{A_{260}}{A_{230}}$  maggiore di 1.8, e  $\frac{A_{260}}{A_{280}}$  compreso tra 1.8 e 2.

## 2.4 Elettroforesi su gel di agarosio

Dopo aver verificato la purezza dell'RNA si controlla la sua integrità tramite **elettroforesi su gel di agarosio**. Questa tecnica permette la separazione di molecole cariche sottoposte ad un campo elettrico. Infatti le molecole migrano in modo differente attraverso le maglie del gel in base al loro rapporto carica/massa. L'attrezzatura per elettroforesi è composta da una cella elettroforetica contenente il gel di agarosio collegata a un alimentatore che, esercita una differenza di potenziale tra i due elettrodi, e determina quindi un flusso di corrente nella cella. Il gel è composto di agarosio sciolto nella soluzione salina usata anche per riempire la cella e permettere la trasmissione del campo elettrico. L'agarosio è uno zucchero solubile in acqua alla temperatura di ebollizione; raffreddandosi solidifica, formando così una matrice, attraverso legami a idrogeno tra le catene lineari. Gli acidi nucleici, carichi negativamente a causa della presenza dei gruppi fosfato, se sottoposti a un campo elettrico migreranno verso il polo positivo. Si può aggiungere al gel un colorante come ad esempio l'Etidio Bromuro, il quale

intercalandosi tra le basi degli acidi nucleici permette la loro visualizzazione ai raggi UV al termine della corsa. Solitamente a fianco del campione di interesse si aggiunge anche un marcatore di peso molecolare, formato da un mix di acidi nucleici di peso molecolare noto, che serve come termine di paragone per confrontare le bande dei campioni ignoti.

Per determinare l'integrità dell'RNA estratto, si sottopone quindi 1µg circa di RNA ad elettroforesi su gel. Il risultato della corsa elettroforetica, se il materiale è in buono stato, deve permettere l'identificazione di due bande ben definite corrispondenti alle subunità 28s e 18s dell'RNA ribosomiale, caratterizzate da un'intensità luminosa della prima doppia rispetto alla seconda. Se l'RNA fosse invece degradato non si distinguerebbero bande definite ma si visualizzerebbe una strisciata uniforme, detta "smear".



**Figura 2.3** Immagine di un gel d'agarosio sottoposto ai raggi UV, con una campione di RNA degradato e uno in buono stato.

## 2.5 Retro-Trascrizione

La retro-trascrizione consiste nella sintesi di una molecola di DNA (chiamata cDNA, *DNA complementare*) a partire da RNA. L'enzima che catalizza questa particolare reazione viene chiamato *trascrittasi inversa*, e consiste in una particolare DNA polimerasi che utilizza come filamento stampo una molecola di RNA [D.L.Nelson et al.,2011].

L'enzima *trascrittasi inversa* per catalizzare la sintesi di una catena di cDNA complementare all'RNA, oltre all'RNA stampo necessita di deossiribonucleotidi trifosfati (dNTPs), primer, una soluzione tampone e  $Mg^{2+}$ . I primer, che costituiscono gli inneschi per l'enzima, possono essere di due tipi: *Random Examer Primer*, cioè sequenze oligonucleotidiche di 6 basi, che legano in maniera casuale le molecole dell'RNA, o *Oligonucleotide Primer Di Poli-T*, cioè sequenze oligonucleotidiche costituite da basi timina, che legano la coda poli-A dell'RNA.

Questa tecnica è molto utilizzata in biologia molecolare perché permette di convertire l'mRNA in cDNA, molecola molto più stabile e meno sensibile ai processi di degradazione. Il cDNA specifico per un gene si diversifica dal corrispettivo DNA perché privo di introni.

Durante la reazione di RT i primer si legano alle specifiche sequenze complementari di RNA per venir poi estesi dell'enzima, formando un composto ibrido dato da un filamento di cDNA e uno di RNA. Il composto può essere utilizzato direttamente nella reazione di real-time PCR, per determinare la quantità relativa di specifici trascritti.

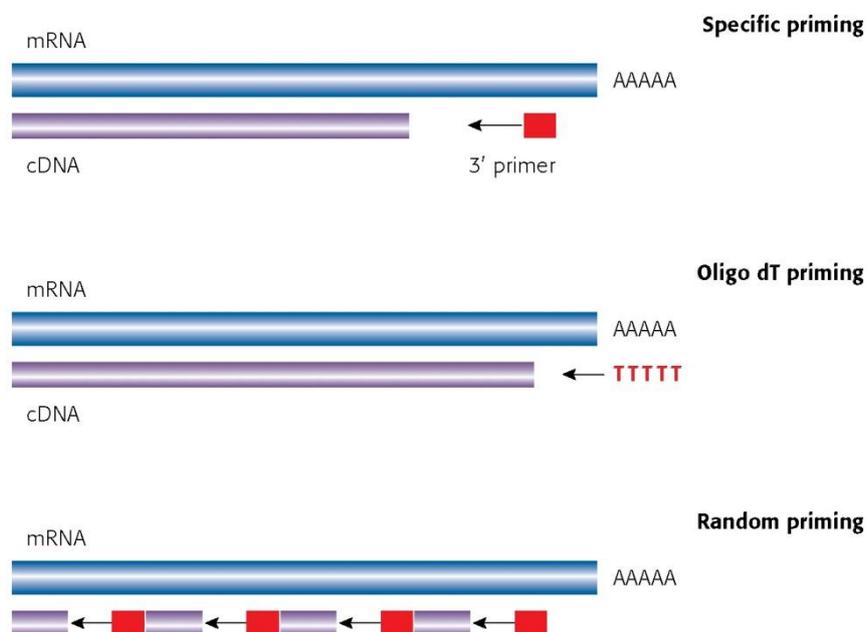


Figura 2.4 Reazione di Retro-Trascrizione che porta a creare un composto ibrido cDNA-RNA.



## 3. REAL-TIME PCR

### 3.1 Reazione a catena della polimerasi

La *Reazione a Catena della Polimerasi* (Polymerase Chain Reaction, PCR) è un metodo *in vitro* di amplificazione degli acidi nucleici. La PCR permette una sintesi esponenziale di un segmento di DNA a partire da un filamento di DNA chiamato stampo. Si utilizza un enzima la DNA-polimerasi che lega le sequenze oligonucleotidiche costruite in modo tale da legarsi all'estremità della sequenza di interesse. La miscela di amplificazione deve perciò contenere:

- **Sequenza target:** consiste nel materiale genetico, contenente le regioni che si desidera amplificare;
- **DNA polimerasi** termo-resistente e termo-stabile. Una possibile DNA polimerasi termoresistente è quella che viene chiamata "*Taq pol*", la quale è una polimerasi isolata dal batterio termofilo "*Thermus aquaticus*" capace di resistere ad elevate temperature, quali quelle necessarie per denaturare il DNA;
- **Due Primer:** sono degli oligonucleotidi a singolo filamento complementari rispettivamente all'estremità 3' e 5' della sequenza target, chiamati perciò senso e antisense. Costituiscono gli elementi di innesco per l'attività della DNA polimerasi.
- **Soluzione buffer:** una soluzione salina acquosa che mantiene stabile il pH.
- **Deossiribonucleotidtrifosfati** (dNTPs): i singoli nucleotidi che costituiscono i mattoncini necessari per la sintesi di nuove eliche.

Nella soluzione si trova anche cloruro di magnesio ( $MgCl_2$ ) che svolge un ruolo di cofattore per la DNA polimerasi e influenza l'appaiamento del primer allo stampo. Un parametro importante da valutare è la sua concentrazione perché influenza la specificità dell'appaiamento dei primer e l'attività dell'enzima[A. Scialpi, A. Mengoni, 2008].

La PCR fu sviluppata nel 1983 dal *Dott. Kary Banks Mullis*, vincitore del premio Nobel per la chimica nel 1993. La reazione prevede il susseguirsi di cicli di amplificazione realizzati grazie all'alternanza di specifici cicli di temperature. Le

fasi che si possono distinguere, per ogni ciclo, sono tre: la *denaturazione*, l'*appaiamento* o *annealing* e infine l'*estensione*.

Durante la denaturazione si raggiungono alte temperature di circa 95°C. In questa fase avviene la separazione della doppia elica del DNA stampo in due filamenti singoli (*denaturazione termica*). Ciò avviene poiché l'alta temperatura rende i legami idrogeno, tra le catene, instabili, permettendone la rottura. La denaturazione viene favorita dalla presenza di concentrazioni saline relativamente alte. Siccome l'emivita della "*Taq pol*" è di circa 30 minuti in ambiente a 95°C, l'attività dell'enzima si mantiene stabile per la realizzazione dei 30-40 cicli necessari ad ottenere milioni di copie dello stesso sequenza bersaglio.

Nella seconda fase, l'appaiamento o annealing, la temperatura viene abbassata (raggiungendo temperature comprese tra i 55 e i 65°C): a queste temperature si verifica il legame dei primer alle sequenze di DNA a singola elica ad essi complementari localizzati alle estremità del frammento di interesse. La temperatura di annealing del primer ( $T_a$ ) è condizionata dal loro contenuto di basi (G) e (C) e dalla loro lunghezza. Se consideriamo dei primer di media lunghezza (circa 20 basi) si utilizza spesso per il calcolo della temperatura di melting  $T_m$  (temperatura alla quale si verifica il distacco del primer) la formula seguente :

$$T_m = [4(G + C) + 2(A + T)] \text{ } ^\circ C$$

Dove G,C,A e T indicano il numero di nucleotidi contenenti le basi azotate corrispondenti. Nel caso che i due primer abbiano delle  $T_m$  differenti, si utilizza in genere la  $T_m$  più bassa.

La temperatura di annealing (temperatura alla quale si verifica l'appaiamento del primer) spesso viene stimata, sulla base di quella di melting, secondo la seguente relazione:

$$T_a = T_m - 5^\circ C$$

Nell'ultima fase, quella di estensione, la temperatura viene leggermente alzata per massimizzare l'azione dell'enzima (la temperatura di lavoro dell'enzima è solitamente compresa tra i 68 e 72°C). In questa fase i primer oligonucleotidici

vengono estesi dalla DNA polimerasi che incorpora i singoli deossiribonucleotiditriposfati (dNTPs) complementari al DNA stampo.

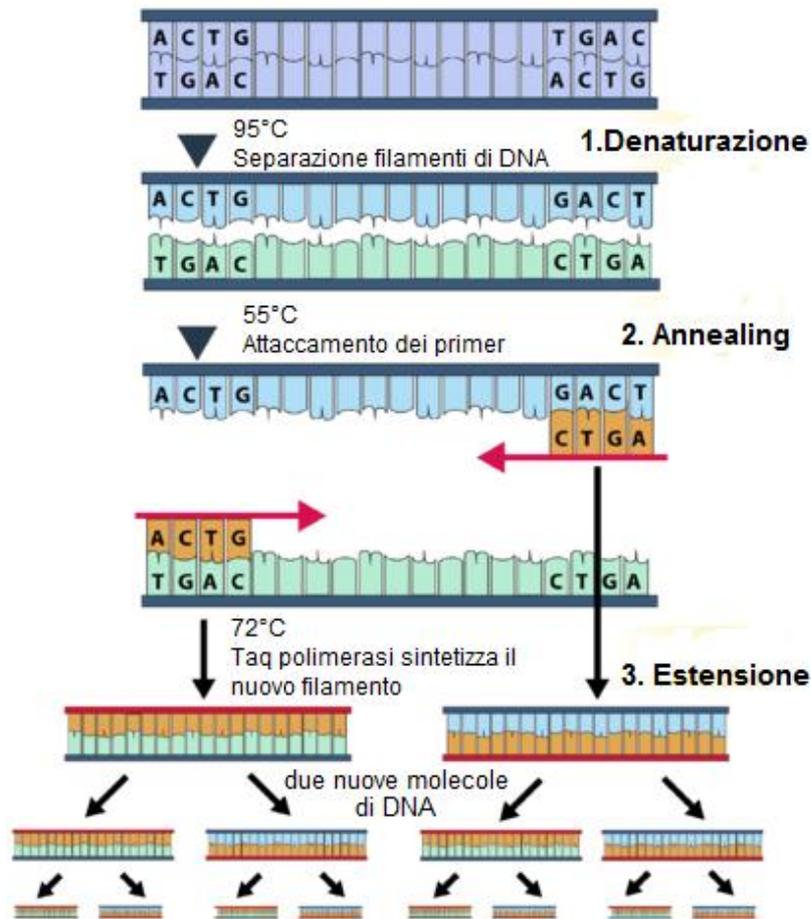


Figura 3.1 Schematizzazione delle fasi di amplificazione della PCR (denaturazione, annealing, estensione).

Il ciclo di amplificazione, costituito dalle tre fasi appena descritte, viene ripetuto 30-40 volte per ottenere un'amplificazione esponenziale della sequenza target. Negli ultimi cicli di amplificazione la reazione rallenta, a causa del consumo di reagenti e della riduzione di attività della polimerasi, fino al raggiungimento di un plateau, nel quale non si ha più amplificazione a causa dell'esaurimento dei reagenti e dell'accumulo di prodotto.

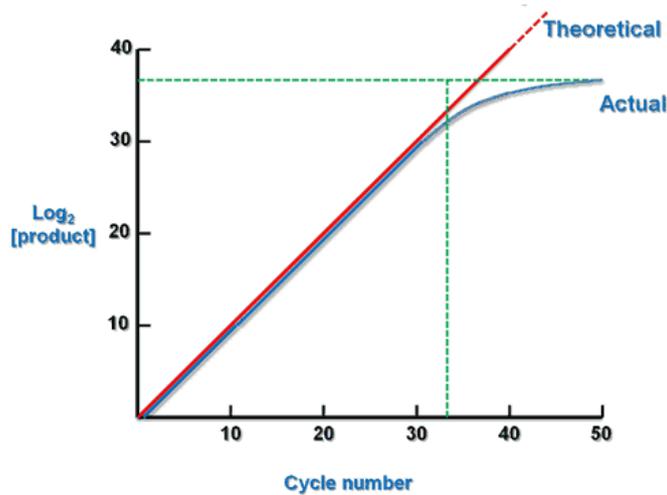


Figura 3.2 Differenza di amplificazione ideale e reale.

### 3.2 La fluorescenza

Per poter comprendere l'evoluzione della tecnica di PCR in real-time PCR è necessario introdurre il fenomeno della fluorescenza.

La **fluorescenza** è la proprietà di alcune sostanze capaci di emettere luce nel visibile, dopo essere state eccitate da radiazioni elettromagnetiche a specifiche lunghezze d'onda.

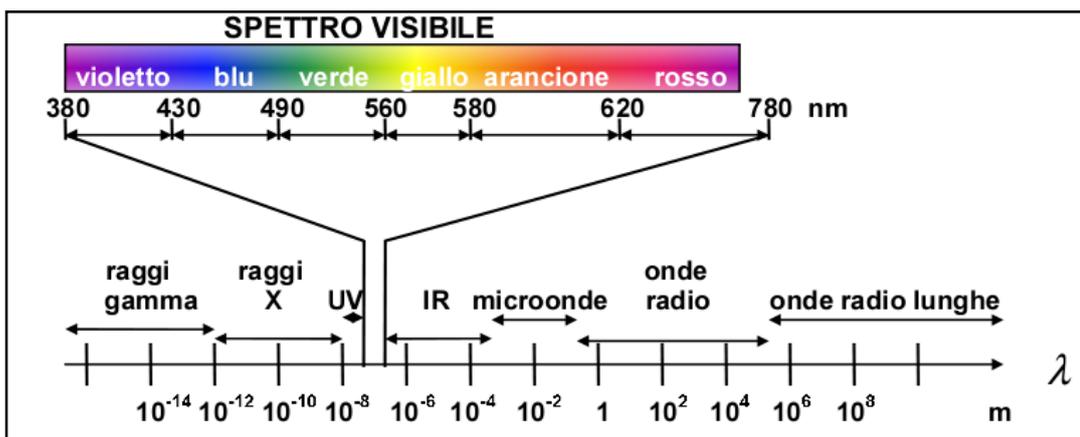


Figura 3.3 Spettro elettromagnetico: indica l'insieme delle diverse lunghezze d'onda relative alle radiazioni elettromagnetiche.

Le radiazioni elettromagnetiche consistono in un fenomeno ondulatorio caratterizzato da un trasferimento di energia. Due parametri importanti che le descrivono sono rappresentati dalla *lunghezza d'onda* ( $\lambda$ ) e dalla *frequenza* ( $\nu$ ).

All'inizio del novecento, a seguito di studi sull'effetto fotoelettrico, emerse l'ipotesi che gli scambi energetici tra luce e materia si fondassero su un'energia quantizzata. *Max Planck*, fisico tedesco, formalizzò questo concetto affermando che l'energia associata a una radiazione elettromagnetica è trasmessa sotto forma di fotoni, ovvero in quanti di energia. L'energia di un singolo fotone  $E$ , è data dalla **legge di Planck**:

$$E = h\nu$$

dove  $\nu$  rappresenta la frequenza e  $h$  è la *costante di Planck* (valore sperimentale della costante  $h = 6,626 \times 10^{-34} \text{ Js}$ ). Risulta quindi che un raggio di luce si comporta come un flusso di particelle, la cui energia è proporzionale alla frequenza [D. Halliday et al. 1992].

Il fotone possiede sia le proprietà di una particella sia quelle di un'onda. Questa particolarità viene definita "*dualismo onda-particella*". Questo dualismo si verifica poiché i fotoni, o in generale le radiazioni elettromagnetiche, si comportano come onde nel loro movimento nello spazio ma nel momento in cui interagiscono con altre particelle elementari manifestano la loro natura quantistica. Dal punto di vista ondulatorio, il prodotto della frequenza e della lunghezza d'onda del fotone è pari alla velocità di propagazione dell'onda. Considerando una situazione di radiazione nel vuoto assume la forma:

$$\lambda\nu = c$$

dove  $c$  è la velocità della luce.

Da queste osservazioni segue che frequenza e lunghezza d'onda sono inversamente proporzionali, e all'aumentare dell'una diminuisce l'altra.

Il principio di fluorescenza si origina quando una radiazione incidente eccita gli atomi della sostanza fluorescente. Gli elettroni vengono promossi ad un orbitale più esterno caratterizzato da un livello energetico maggiore. Entro poche decine di nanosecondi l'elettrone eccitato ritorna al livello energetico di partenza, passando per stati eccitati a energia intermedia: questo processo è chiamato decadimento. I passaggi per gli stati intermedi non determinano l'emissione di

una radiazione, tranne l'ultimo, che genera luce a una lunghezza d'onda maggiore rispetto a quella della radiazione incidente e quindi dotato di energia minore. Questa luce viene chiamata *fluorescenza*. Le variazioni degli stati energetici sono rappresentate dal *diagramma Jablonski* in figura 3.4, che prende il nome dal fisico polacco *Aleksander Jablonski* [Lakowicz et al. 1999].

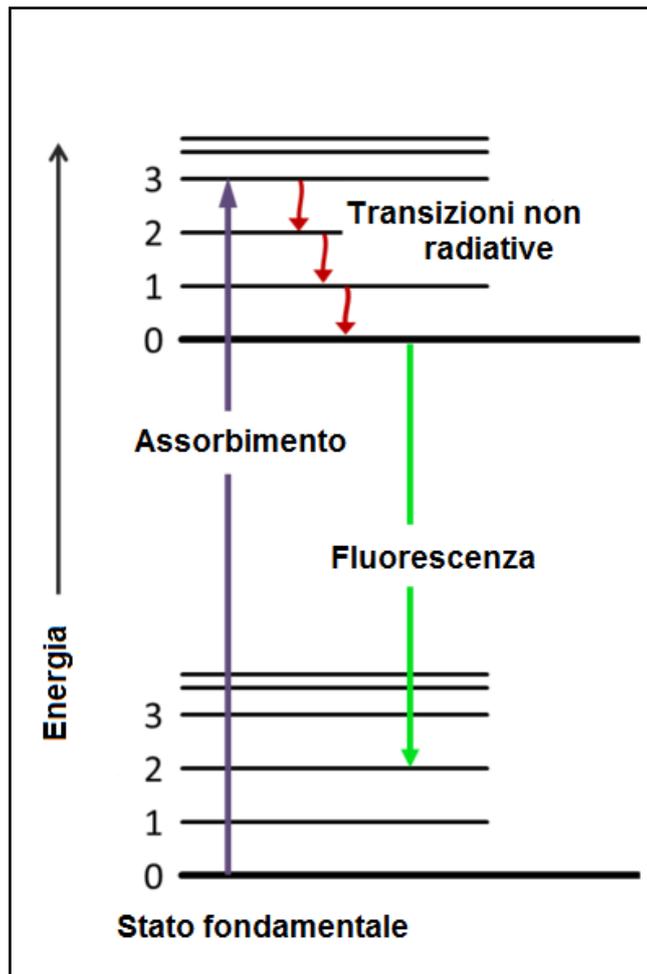


Figura 3.4 Diagramma Jablonski.

La fluorescenza è un fenomeno di emissione spontanea; se consideriamo una sorgente luminosa in uno stato eccitato con energia  $E_2$ , essa può decadere spontaneamente verso lo stato fondamentale con energia  $E_1$ . La differenza di energia tra i due stati viene rilasciata, sotto forma di fotone:

$$E_2 - E_1 = h\nu$$

L'energia della radiazione emessa è sempre inferiore a quella assorbita, rispetto a quest'ultima avrà quindi una frequenza inferiore e una lunghezza d'onda superiore: questo fenomeno è chiamato Stokes Shift.

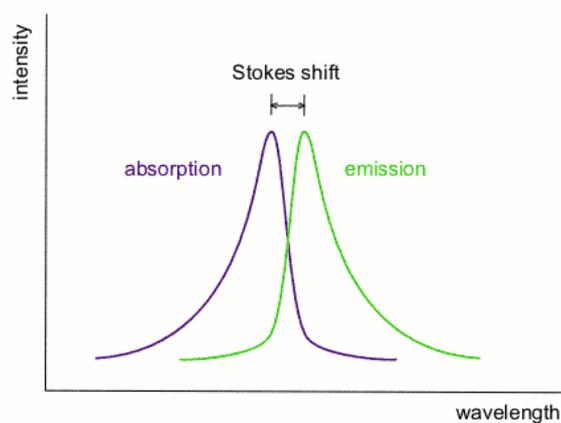


Figura 3.5 Stokes shift.

A differenza di altri fenomeni di luminescenza, la fluorescenza, ha un decadimento istantaneo, ovvero l'emissione di radiazione termina nel momento in cui si smette di eccitare il campione.

Un parametro che si può valutare nel fenomeno di fluorescenza è la **resa quantica**, essa ci fornisce una stima dell'efficienza del processo di fluorescenza. E' definita come il rapporto tra il numero di fotoni emessi e il numero di quelli assorbiti:

$$\Phi = \frac{\text{Numero fotoni emessi}}{\text{Numero fotoni assorbiti}}$$

La massima resa quantica si ha per  $\Phi$  pari a 1 (100%) ovvero quando il numero di fotoni emessi e assorbiti risulta uguale.

### 3.3 Real-Time PCR

La tecnica di PCR convenzionale è una tecnica qualitativa; solo assumendo particolari accorgimenti può arrivare a essere semi-quantitativa. Infatti non è possibile correlare la quantità di prodotto finale con la quantità di DNA stampo presente inizialmente, perché nella fase finale l'efficienza della reazione può essere variabile. Uno sviluppo della tecnica di PCR consiste nella **real-time PCR**.

La **real-time PCR**, o *qPCR (PCR quantitativa)*, consente di quantificare la sintesi del prodotto di PCR ad ogni ciclo di amplificazione in tempo reale. Questo permette di effettuare un'analisi quantitativa della quantità di DNA stampo iniziale. Il segnale che viene quantificato è rappresentato dalla fluorescenza emessa da fluorofori, cioè coloranti fluorescenti in grado di legarsi alle molecole di DNA prodotte ad ogni ciclo di amplificazione. I fluorofori possono intercalarsi al DNA in maniera aspecifica, oppure fungere da marcatori di sonde oligonucleotidiche complementari a specifiche sequenze.

Le informazioni che si ottengono con questa tecnica sono quindi maggiori rispetto a quelle di una classica PCR. La real-time PCR è spesso utilizzata in combinazione con la reazione di retro-trascrizione (RT), per quantificare i livelli di espressione di specifici geni di interesse.

#### 3.3.1 SYBR Green

I fluorofori usati in real-time PCR sono caratterizzati dallo sviluppare fluorescenza solo quando vengono incorporati nei filamenti di DNA neosintetizzati. La fluorescenza emessa aumenta in rapporto al numero di siti di legame ai doppi filamenti di DNA prodotti. La quantità di prodotto amplificato può essere quindi determinata ad ogni ciclo di amplificazione, rilevando al termine della fase di estensione, la radiazione emessa del fluoroforo.

Uno dei fluorofori più utilizzati è il **SYBR® Green**, una molecola fluorescente che durante le reazioni di PCR si intercala nel doppio filamento del DNA. Come si nota dalla figura 3.6 il complesso DNA-SYBR® Green assorbe luce blu ad una lunghezza d'onda  $\lambda_{max} = 488 \text{ nm}$  ed emette luce verde a una lunghezza d'onda  $\lambda_{max} = 522 \text{ nm}$ . Altri picchi di assorbimento, molto più deboli, li troviamo nella regione dell'ultravioletto. Il SYBR® Green si intercala nel DNA a doppio filamento,

motivo per cui non è adatto per quantificare RNA a singolo filamento [Ian M. Mackay, 2007].

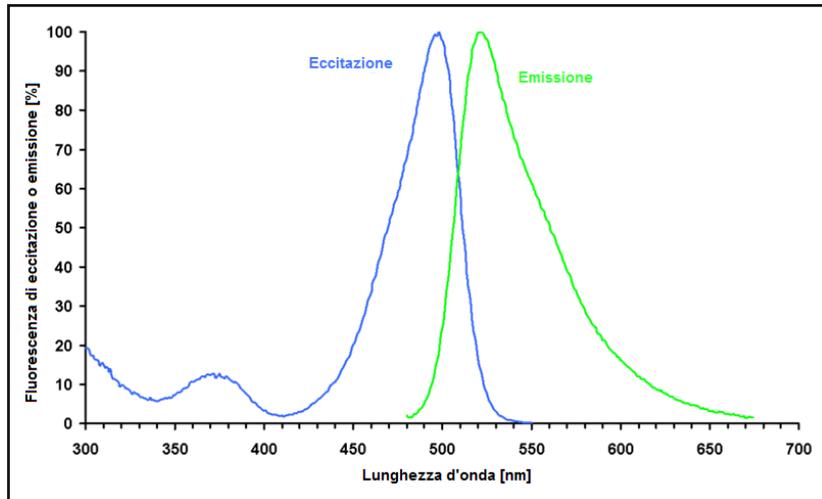


Figura 3.6 Spettro di eccitazione e di emissione del SYBR® Green.

Durante la fase di denaturazione il SYBR green si trova libero nella miscela di reazione, successivamente nella fase di annealing esso si posiziona in maniera aspecifica nel solco minore del DNA. Nella fase di elongazione il colorante si intercala nella molecola di DNA e quando eccitato emette fluorescenza in modo proporzionale al numero di copie di DNA prodotte (figura 3.7).

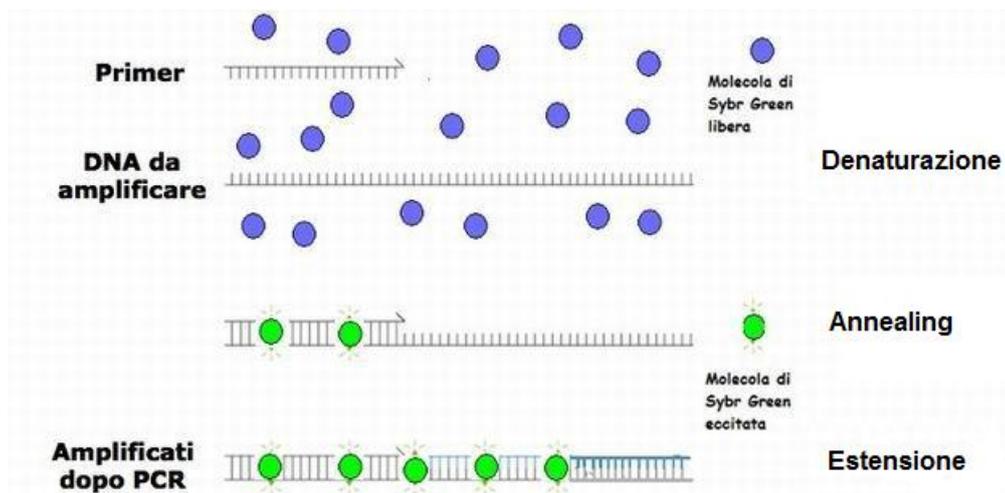


Figura 3.7 Incorporazione del SYBR green nelle tre fasi di PCR.

### 3.3.2 Sonda

Nelle analisi basate sull'utilizzo di sonde specifiche, il segnale fluorescente viene rilevato solo in conseguenza dell'appaiamento della sonda alla sequenza bersaglio. Le sonde tipicamente usate possono essere di due tipi: *sonde idrolitiche* o di *ibridazione*.

Le **sonde idrolitiche** presentano rispettivamente alle due estremità un fluoroforo ad alta energia (*reporter*) e un inibitore del fluoroforo (*quencher*). Quando la sonda è appaiata alla sequenza bersaglio, l'inibitore è sufficientemente vicino al fluoroforo da bloccarne l'emissione del segnale fluorescente. Durante l'allungamento, in ogni ciclo di amplificazione, la polimerasi idrolizza la sonda. In questo modo il fluoroforo viene liberato nella miscela di reazione e si allontana dall'azione dell'inibitore, con conseguente emissione del segnale fluorescente. Quindi nelle analisi che utilizzano sonde idrolitiche l'emissione del segnale fluorescente dipende dall'attività della *Taq DNA polimerasi*. Le sonde idrolitiche vengono anche chiamate *sonde TaqMan*.

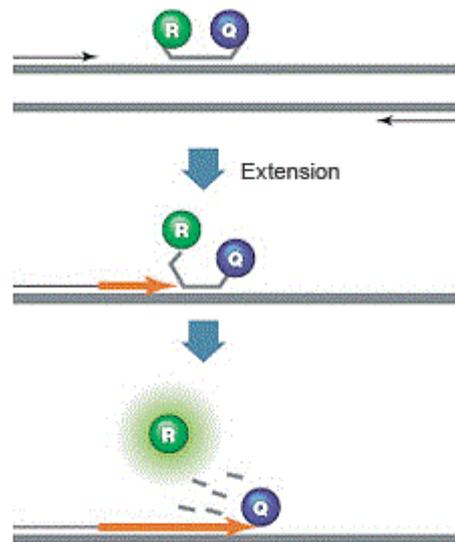
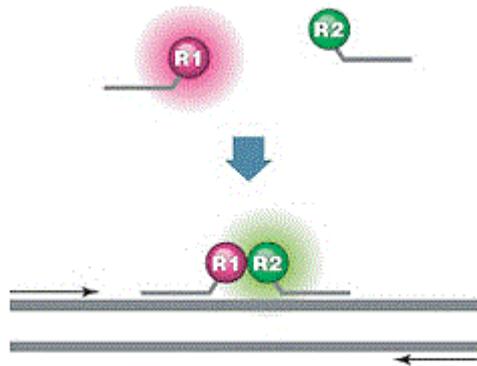


Figura 3.8 Sonda idrolitica TaqMan. R=reporter, Q=Quencher

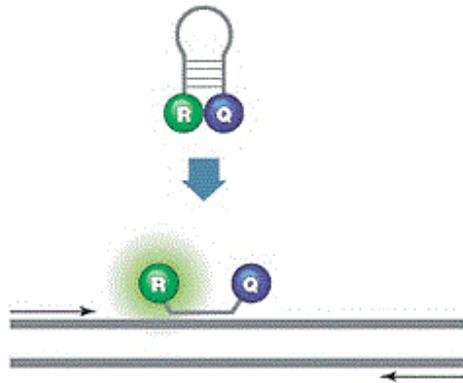
Le **sonde di ibridazione**, invece, permettono di rilevare il segnale nel momento in cui si legano alla sequenza bersaglio. Esistono diversi modelli di sonde di ibridazione, tra queste troviamo un tipo di sonda che sfrutta il trasferimento di energia per risonanza fluorescente (FRET). Durante il processo di FRET,

l'eccitazione di un fluoroforo donatore provoca il trasferimento di energia ad un fluoroforo accettore. Le *sonde FRET* sono formate da una coppia di oligonucleotidi, progettati in modo tale da ibridarsi sulla sequenza bersaglio. Il segnale che si rileva è proporzionale alla quantità di sonda ibridata e, di conseguenza, alla quantità di prodotto amplificato.



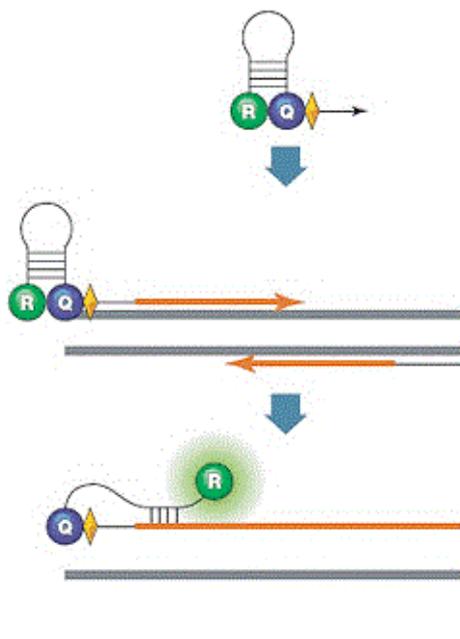
**Figura 3.9 Sonda FRET. R1 rappresenta il fluoroforo donatore, mentre R2 è il fluoroforo accettore.**

Un'altra variante delle sonde di ibridazione consiste nei "*molecular beacon*" (*sonde beacons*). Queste sonde sono progettate in modo da avere le estremità auto-complementari, cosicché la molecola assume una conformazione a forcina. Alle estremità sono legati, rispettivamente, un fluoroforo e un suo inibitore. Il segnale viene emesso nel momento in cui la sonda si lega alla sequenza bersaglio. Infatti, con la distensione della forcina, il fluoroforo si allontana dall'inibitore e risulta quindi capace di emettere una radiazione luminosa.



**Figura 3.10** Sonda “Molecular Beacon”.

Un altro modello di sonde di ibridazione è costituito dalle “sonde *Scorpion*”. Queste sono singole molecole con una configurazione simile a quelle delle sonde beacons. La particolarità consiste in una sequenza specifica complementare a una porzione della sequenza target, legato all'estremità 3'. Le sonde *Scorpion* contengono anche un “*PCR blocker*” per impedire la lettura durante l'estensione del filamento opposto. Quando avviene la reazione di PCR, la sonda si accoppia alla sequenza complementare che si trova all'interno della molecola di DNA, l'ibridazione apre l'ansa e allontana il reporter fluorescente dal suo inibitore, con conseguente emissione di fluorescenza misurabile [A. Scialpi, A.Mengoni, 2008].



**Figura 3.11** Sonda *Scorpion*

### 3.4 Applicazioni

In un breve periodo di tempo, dall'inizio della sua introduzione a metà anni Novanta, la macchina di real-time PCR divenne ampiamente disponibile; inizialmente le sue applicazioni principali erano quelle della PCR convenzionale. L'aumento di affidabilità e il miglioramento della precisione quantitativa, uniti a una maggiore rapidità del processo, hanno contribuito all'adozione della tecnica di real-time PCR in campi molto diversi [J.M.J.Logan et al. 2009].

Le applicazioni per le quali si utilizza la tecnica di real-time PCR sono le più svariate. Tra queste troviamo, naturalmente, molte applicazioni nel campo della *ricerca* e dell'*ingegneria genetica*. In combinazione con la PCR a trascrizione inversa (RT-PCR), i saggi di real-time PCR possono essere utilizzati per quantificare precisamente i cambiamenti nell'espressione genica, come un aumento o una diminuzione dell'espressione di geni target in risposta a diverse condizioni ambientali o trattamenti farmacologici.

Inoltre questa tecnica trova ampia applicazione in *campo alimentare*, per la rilevazione di organismi geneticamente modificati negli alimenti, e in *campo medico*, come ad esempio nella valutazione di possibili mutazioni genetiche correlate a diverse patologie genetiche e non o nei test di paternità.

Il vantaggio di test basati sulla PCR è la sensibilità, cioè la possibilità di effettuare analisi accurate partendo da piccole quantità di campioni.

Altri campi in cui è utilizzato il test di real-time PCR sono quelli di *medicina legale* e di *investigazione forense*.

### 3.5 Hardware

Tutti gli strumenti di real-time PCR combinano un termociclatore e un sistema ottico, uniti a un computer e a un software specializzato per l'elaborazione dei dati.

Il termociclatore è uno strumento che riesce a realizzare automaticamente le variazioni cicliche di temperatura richieste dai diversi protocolli. E' composto da un blocco termico dotato di alloggi per i campioni e un sistema in grado di variare la temperatura sul blocco. Alcuni modelli sono in grado di creare dei gradienti

termici lungo il blocco, questa possibilità è utile nel caso in cui si vogliono utilizzare temperature di annealing diverse per coppie di primer. La maggioranza dei blocchi di riscaldamento sono basati sull'*effetto Peltier*. L'*effetto Peltier* è il fenomeno termoelettrico per cui una corrente che scorre tra due metalli o semiconduttori diversi posti a contatto produce un trasferimento di calore.

Il calore assorbito è pari a:

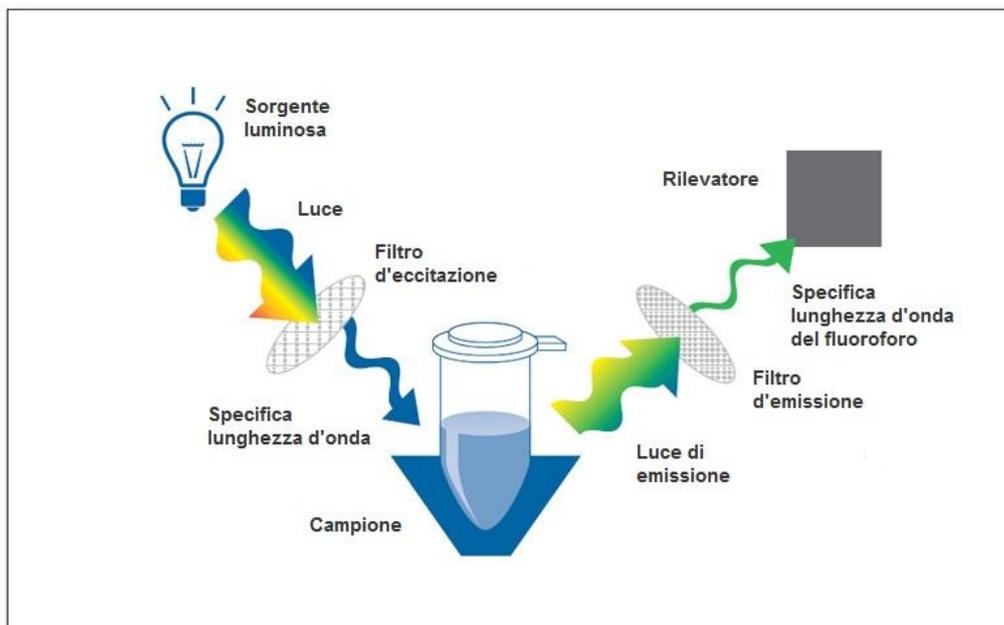
$$Q = (\Pi_B - \Pi_A) It$$

dove  $\Pi$  rappresenta il coefficiente di Peltier dei singoli materiali,  $I$  è la corrente che scorre e  $t$  rappresenta il tempo.

Un vantaggio delle tecnologie basate sull'effetto Peltier, consiste nella loro possibile combinazione con tecniche resistive o convettive per migliorare il controllo della temperatura e le performance lungo il blocco.

Solitamente i blocchi hanno una massa significativa e di conseguenza un certo grado di inerzia termica. Recentemente è stata introdotta la tecnologia brevettata "*Therma-Base*" il cui principio si basa su l'evaporazione e la condensazione di un fluido di lavoro, contenuto in una piccola cavità sotto vuoto posta sotto gli elementi peltier: questa tecnologia permette di controllare accuratamente la variazione di temperatura in ogni pozzetto.

La precisione nella realizzazione dei cicli di temperatura dipende dalla regolazione della temperatura nella piastra multipozzetto dove vengono alloggiati le provette contenenti i campioni e dalla velocità con cui le temperature possono essere raggiunte [J.M.J.Logan et al. 2009].



**Figura 3.12 Schema del sistema ottico di eccitazione/rilevazione.**

Il sistema ottico ha la funzione di eccitare i fluorofori e catturare in seguito, la fluorescenza emessa corrispondente all'accumulo dei prodotti. E' composto, come si nota in figura 3.12, da una sorgente luminosa, da filtri di eccitazione e di emissione e da un rivelatore. La luce emessa dalla sorgente luminosa viene filtrata in modo da selezionare le lunghezze d'onda specifiche per l'eccitazione dei fluorofori. La luce d'emissione prodotta dal campione viene filtrata, in modo da selezionare il picco di emissione del fluoroforo che si sta utilizzando, prima di raggiungere il rivelatore [J.M.J Logan et al. 2009].

La sorgente luminosa può essere, a seconda dello strumento, una lampada alogena o allo Xenon, un laser oppure un insieme di diodi emettitori di luce (LED). I filtri, che selezionano l'appropriata lunghezza d'onda d'eccitazione o d'emissione, possono essere principalmente di due tipi: filtri a intervallo ristretto oppure filtri ad ampia gamma. Il rivelatore, che converte la radiazione luminosa rivelata in un segnale elettrico può essere realizzato da: fotodiodi, telecamere con dispositivo ad accoppiamento di carica (CCD) o tubo fotomoltiplicatore. Un fotodiodo trasforma la luce incidente in un segnale elettrico mentre un dispositivo CCD trasforma la luce in dati digitali, la qualità dell'immagine catturata viene determinata dalla risoluzione in megapixel. Solitamente i rivelatori utilizzati in real-time sono impostati per misurare strette bande dello spettro.

Lo strumento di real-time PCR presente in laboratorio di Ingegneria Cellulare e Molecolare “S. Cavalcanti” è il modello CFX Connect™ di Bio-Rad. Il metodo di riscaldamento e raffreddamento del termo blocco è basato sull'effetto Peltier e il range di temperature è compreso tra 0-100°C con una precisione di  $\pm 0.2^\circ\text{C}$ . Lo strumento permette anche di creare un gradiente termico lungo il blocco.



**Figura 3.13 CFX Connect™, apparecchio per real-time PCR di Bio-Rad.**

Il sistema ottico è composto da tre LED accoppiati a tre filtri di eccitazione e tre fotodiodi accoppiati ad altrettanti filtri di emissione per la rilevazione. Il range delle lunghezze d'onda di eccitazione/emissione è compreso tra 450-580nm. Il tempo di scansione della piastra è pari a 3 secondi utilizzando un unico canale, corrispondente a un fluoroforo.

Lo strumento ha una capacità di analisi di 96 pozzetti, con un volume di reazione del campione compreso tra 1 e 50 $\mu\text{l}$  (il manuale tecnico consiglia l'utilizzo di almeno 10-25  $\mu\text{l}$  di campione).

### **3.6 Software**

Molte aziende biotecnologiche offrono pacchetti software per l'analisi dei dati di real-time PCR.

In laboratorio di Ingegneria Cellulare e Molecolare “S. Cavalcanti” a Cesena si utilizza il software Bio-Rad CFX Manager™. CFX Manager™ è un software in

lingua inglese di facile utilizzo, e offre sia gli strumenti per la gestione e la programmazione dell'esperimento sia le diverse modalità di analisi dei dati.

Il software consente la programmazione di diversi protocolli sperimentali. La configurazione del protocollo permette di settare tempi e temperature per le diverse fasi del ciclo di PCR e specificare il momento in cui raccogliere il segnale fluorescente.

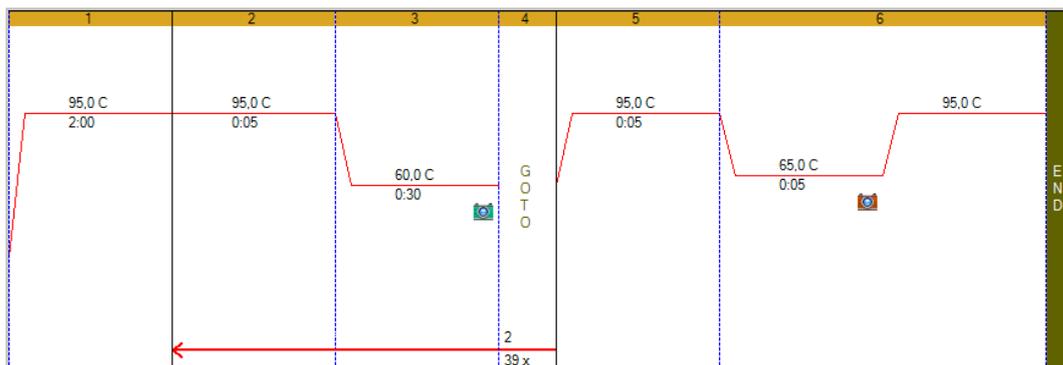


Figura 3.14 Protocollo dei cicli di PCR.

Il software permette anche di identificare la descrizione del contenuto della piastra da 96 campioni, specificando se i campioni sono incogniti o controlli, come per esempio il *no template control* (NTC) o il calibratore interno (IC, *internal calibrator*). Si identificano i geni target o reference che si sta amplificando e il nome dei vari campioni, indicando la presenza di replicati tecnici o biologici. Si specifica infine il tipo di fluoroforo utilizzato. L'impostazione della piastra può essere completata prima, durante o a corsa conclusa.

Durante i cicli di PCR avviene la raccolta del segnale fluorescente: l'intensità della fluorescenza della miscela di reazione in ciascun pozzetto viene misurata e tracciata rispetto al ciclo di reazione. Se è presente un target la sua amplificazione può essere monitorata in tempo reale.

Al termine di una corsa viene automaticamente generato un file dati, che può essere utilizzato nelle successive fasi di analisi. Il software permette la visualizzazione delle curve di amplificazione e relativi valori numerici di singoli campioni o di gruppi di essi grazie alla possibilità di creare dei "*well-groups*", che verranno analizzati indipendentemente tra loro.

E' possibile poi esportare rapidamente sia dati, grafici e tabelle in file processabili da altri programmi per ulteriori analisi.

## **3.7 Progettazione e ottimizzazione della reazione di PCR**

Le performance della PCR e la validità dei risultati ottenuti dipendono da alcuni fattori chiave di cui bisogna tenere conto. Per ottimizzare un saggio di real-time PCR si possono perfezionare i protocolli termici e il disegno dei primer. Inoltre è importante controllare la qualità del campione di partenza. Un esperimento di PCR ottimizzato deve risultare sensibile, specifico e riproducibile.

### **3.7.1 Qualità dell'RNA**

La validità e la precisione dell'espressione genica stimata tramite real-time PCR è profondamente influenzata dalla qualità, dalla purezza e dall'integrità del materiale genetico di partenza. Per essere considerato "*high quality*" l'RNA da utilizzare nella retro-trascrizione deve essere privo di RNasi, di proteine e di DNA genomico[S. Kennedy et al. 2001]. Occorre quindi porre un'attenzione particolare durante i processi di estrazione e conservazione dell'RNA, così come nelle fasi in preparazione della reazione di PCR, infatti è proprio in questi momenti che si verifica la maggior parte delle imprecisioni che possono pregiudicare il risultato finale.

### **3.7.2 Protocollo termico di PCR**

Nell'impostazione del protocollo termico è importante stabilire i tempi e le temperature ottimali in ogni fase del ciclo di PCR. Infatti esiste un'alta correlazione tra la qualità dei risultati ottenuti e i parametri impostati per le fasi di annealing e di estensione[S. Kennedy et al. 2001].

Nel programmare la temperatura di annealing  $T_a$  si possono seguire due possibilità:

- $T_a$  costante durante i cicli;
- $T_a$  che decresce al seguirsi dei cicli (*touch-down*).

La strategia più comune è quella che prevede di mantenere la  $T_a$  costante, senza effettuare variazioni in base al numero dei cicli. L'utilizzo del secondo metodo invece ha il vantaggio di rendere i primi cicli di PCR molto specifici, infatti in questo modo viene promossa solo l'amplificazione di determinati frammenti, rendendo instabili eventuali accoppiamenti con sequenze non perfettamente complementari. Infatti valori troppo bassi della  $T_a$  potrebbero causare appaiamenti dei primer a sequenze di DNA non esattamente complementari, e di conseguenza la loro amplificazione comprometterebbe la specificità della reazione. La difficoltà opposta sta nel fatto che  $T_a$  troppo elevate possono ridurre il rendimento, in quanto solo una frazione delle molecole del primer è in grado di innescare la polimerizzazione, data l'elevata instabilità del loro accoppiamento con il DNA stampo [A. Scialpi, A. Mengoni, 2008]. Per aumentare la specificità occorre considerare anche che la durata della fase di annealing non deve essere troppo prolungata per sfavorire appaiamenti con bassa complementarietà. Solitamente si utilizzano tempi inferiori o pari a circa 30 secondi.

Il tempo di estensione solitamente viene calibrato in funzione della lunghezza del target da amplificare. In alcuni casi può essere utile aumentare il tempo di estensione al susseguirsi dei cicli per prevenire, almeno in parte, la riduzione di concentrazione della DNA polimerasi attiva.

### **3.7.3 Disegno dei Primer**

In commercio si trovano primer e sonde ottimizzate per i protocolli di analisi dell'espressione genica in real-time PCR. Esistono tuttavia diversi software specifici per il disegno e la validazione dei primer. Questi software permettono di verificare la possibilità che si creino i cosiddetti dimeri di primer, costituiti dall'appaiamento di primer tra loro, a causa di una loro parziale complementarietà. La produzione di dimeri di primer costituisce un problema in real-time PCR in quanto riduce il numero di primer liberi disponibili per la reazione di PCR e determina un segnale aspecifico. I parametri da tenere in

considerazione nella fase di progettazione di una coppia di primer sono la lunghezza, la temperatura di melting e la composizione in basi delle sequenze[T.Weissensteiner et al, 2004]. La presenza di dimeri di primer può essere valutata controllando i picchi delle curve di melting: se nella curva è presente un solo picco relativo alla temperatura in cui si verifica il disappaiamento del prodotto di PCR allora si può escludere la presenza di dimeri di primer o di prodotti di amplificazione aspecifici.

#### **3.7.4 Importanza dei controlli**

Un saggio deve essere progettato includendo l'inserimento di controlli selezionati. Le forme di controllo più semplici sono costituite dai controlli negativi e dai controlli positivi. I controlli negativi identificano la presenza di prodotti aspecifici o la contaminazione dei campioni. Un controllo negativo è il controllo NTC, “*no template control*”, il quale è costituito da tutti i componenti della miscela di reazione eccetto il materiale genetico da amplificare. Questo è importante sia per valutare la presenza di una contaminazione dei reagenti con cui si prepara la miscela di reazione[S.Kennedy et al. 2001].

I controlli positivi sono utilizzati per confermare che le condizioni di reazione e i reagenti utilizzati siano idonei ad ottenere un amplificato di PCR. A questo scopo si impiegano come controlli positivi dei campioni validati di cui si conosce l'esito dell'amplificazione.

Si può inserire anche un campione in cui si amplifica l'RNA di partenza non retroscritto per verificare un'eventuale contaminazione da DNA genomico.

Solitamente si utilizzano più replicati tecnici, cioè la stessa reazione di PCR viene eseguita su più campioni uguali per migliorare la precisione ed escludere errori tecnici imputabili all'operatore nella fase di dispensamento dei campioni nella piastra. Si possono quindi escludere dall'analisi i replicati anomali e mediare i risultati dei replicati simili.

#### **3.7.5 Riproducibilità, Sensibilità e Specificità**

Un saggio PCR ottimale deve essere riproducibile, ci deve essere cioè uniformità tra i risultati dei replicati tecnici relativi ad uno stesso campione, e dei replicati biologici, corrispondenti a ripetizioni indipendenti di uno stesso esperimento.

La *sensibilità analitica*, della real-time PCR, è definita come il minimo numero di copie nel campione che possono essere rilevate, utilizzando un saggio e condizioni sperimentali definite. Spesso viene indicata come “*limite di rilevamento*” (limit of detection, **LOD**) e viene espresso come una probabilità di rilevazione. Per esempio una probabilità di rilevazione del 95% esprime il 5% di possibilità del fallimento della rilevazione della sequenza target. La sensibilità può essere riportata anche come percentuale di campione identificato correttamente, ad esempio una sensibilità del 98% indica che 98% del campione è stato riconosciuto correttamente[S.Kennedy et al. 2001].

La *specificità analitica* si riferisce alla capacità del saggio di identificare il target desiderato. La procedura di real-time PCR risulta specifica quando l'amplificato è costituito unicamente dal DNA target. La specificità è influenzata dai primer scelti e si può valutare tramite l'analisi delle curve di melting[J.P.Clewley,1995]. La specificità deve essere particolarmente alta nel caso in cui la PCR sia utilizzata nella diagnostica clinica.



## 4. ANALISI DEL SEGNALE

L'analisi del segnale di real-time PCR viene spesso eseguita con il software fornito con lo strumento, ma esistono diversi programmi di analisi che differiscono più che altro nelle interfacce grafiche ma che implementano solitamente gli stessi metodi di analisi. In ogni caso l'analisi delle curve di amplificazione e delle curve di melting è di fondamentale importanza per procedere ad un'accurata analisi quantitativa.

### 4.1 Curva di amplificazione

La curva di amplificazione descrive l'aumento di fluorescenza in funzione del numero di cicli di PCR. L'intensità della fluorescenza emessa durante la real-time PCR è correlata con la quantità di prodotto formato. Il segnale fluorescente emesso ad ogni ciclo di amplificazione viene rilevato dal sistema ottico e memorizzato da un computer. Dopo i primi cicli di amplificazione, la fluorescenza supera un livello di soglia arbitrario e inizia ad aumentare in modo esponenziale.

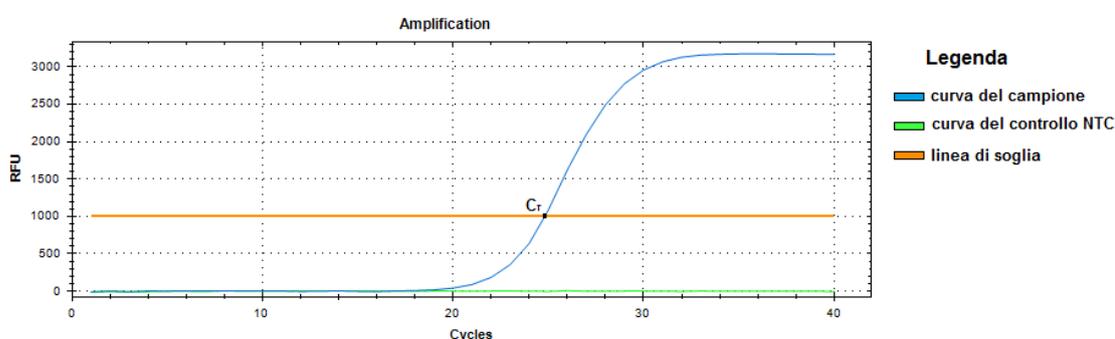


Figura 4.1 Esempio di curva di amplificazione di un campione e di un controllo NTC

Come si nota dalla figura 4.1, la curva di amplificazione ha un andamento tipicamente sigmoideale con una fase esponenziale, una fase lineare e una fase di plateau in cui termina l'amplificazione, e la fluorescenza si stabilizza a un valore di regime. Il segnale relativo al controllo NTC mostra che non sono presenti contaminazioni o amplificati aspecifici, in quanto il suo valore di fluorescenza rimane costante e prossimo allo zero.

La curva di amplificazione si può caratterizzare attraverso un *modello sigmoidale*:

$$f(x) = y_0 + \frac{a}{1 + e^{-\left(\frac{x-x_0}{b}\right)}}$$

Questo modello è caratterizzato dai seguenti parametri:  $x$  è il numero di cicli;  $f(x)$  è la fluorescenza calcolata in funzione del numero di cicli  $x$ ;  $y_0$  è la fluorescenza di fondo;  $a$  è la differenza tra fluorescenza massima raggiunta nella fase di plateau e la fluorescenza di fondo;  $x_0$  è la coordinata del punto di flesso della curva; infine  $b$  descrive la pendenza nella fase log-lineare [M.T.Dorak, 2006]. La curva di amplificazione può essere visualizzata anche in un grafico semi-logaritmico per identificare meglio la fase lineare.

La pendenza (*slope*) della fase log-lineare della reazione di amplificazione ci fornisce una misura dell'efficienza di amplificazione. L'efficienza può essere determinata dalla seguente equazione:

$$E = 10^{(-1/slope)}$$

Per ottenere dei risultati accurati e riproducibili le reazioni dovrebbero avere un'efficienza vicina al 100%, che equivale a una pendenza pari a -3.32. Un'efficienza del 100% corrisponde al caso in cui ad ogni ciclo il numero di copie della sequenza target è raddoppiato, ed equivale ad un'efficienza di reazione pari a 2.

#### **4.1.1 Linea Threshold**

La linea di soglia della reazione di PCR (*threshold line*) rappresenta il livello di segnale che raggiunge un aumento statisticamente significativo rispetto al segnale di background. Viene impostata automaticamente dal software, ma può essere poi riposizionata manualmente dall'operatore in modo che essa intersechi tutti i campioni dell'esperimento nella loro fase esponenziale. Solitamente si posiziona a circa metà della fase di amplificazione lineare, meglio identificabile visualizzando le curve su scala logaritmica.

### 4.1.2 Ciclo Soglia

Il *ciclo soglia*  $C_T$ , chiamato anche *ciclo di quantificazione*  $C_Q$ , rappresenta il numero di cicli necessari in cui il segnale di amplificazione raggiunge la linea di soglia impostata. Graficamente è il punto di intersezione tra la *linea di soglia* (*threshold line*) e la curva di amplificazione. Su questo parametro si basano i diversi metodi per la quantificazione del segnale.

Il  $C_T$  di ogni campione è inversamente proporzionale alla quantità iniziale: infatti minore è la quantità di materiale di partenza, più cicli saranno necessari per raggiungere il livello di soglia. Ad esempio se si confrontano due campioni contenenti una quantità doppia di sequenza target rispetto all'altro, ipotizzando che l'efficienza per entrambe le reazioni di PCR sia del 100%, la curva di amplificazione del campione più concentrato intersecherà la linea di soglia un ciclo prima rispetto alla curva del campione meno concentrato.

## 4.2 Curve di Melting

Le curve di melting, o curve di fusione, costruite al termine della reazione di PCR, descrivono la variazione della fluorescenza in funzione della temperatura. E' possibile ottenere le curve di melting se si utilizzano particolari saggi, come ad esempio il test basato sull'intercalante aspecifico SYBR® green. Ogni amplificato di PCR ha la sua curva di melting caratteristica, essa si ottiene alla fine della reazione dei cicli di amplificazione incrementando la temperatura da 65 a 95°C con step di 0.5°C ed effettuando la lettura della fluorescenza ad ogni step. All'aumentare della temperatura, fino al raggiungimento della temperatura di melting, il doppio filamento target amplificato verrà denaturato con conseguente diminuzione della fluorescenza rilevata.

L'analisi delle curve di melting è un pratico e utile controllo della specificità di amplificazione della real-time PCR. Il controllo dinamico dell'intera curva di melting al variare della temperatura definisce completamente la dinamica di fusione[S. Kennedy et al 2001]. Esistono due modalità di visualizzazione delle curve di melting: la prima consiste nella curva di fluorescenza in funzione della temperatura (figura 4.2); la seconda, rappresenta la derivata negativa della

fluorescenza rispetto alla temperatura  $-\frac{dF}{dT}$  (figura 4.3). L'unità di misura della fluorescenza rilevata è la RFU "*relative fluorescence units*".

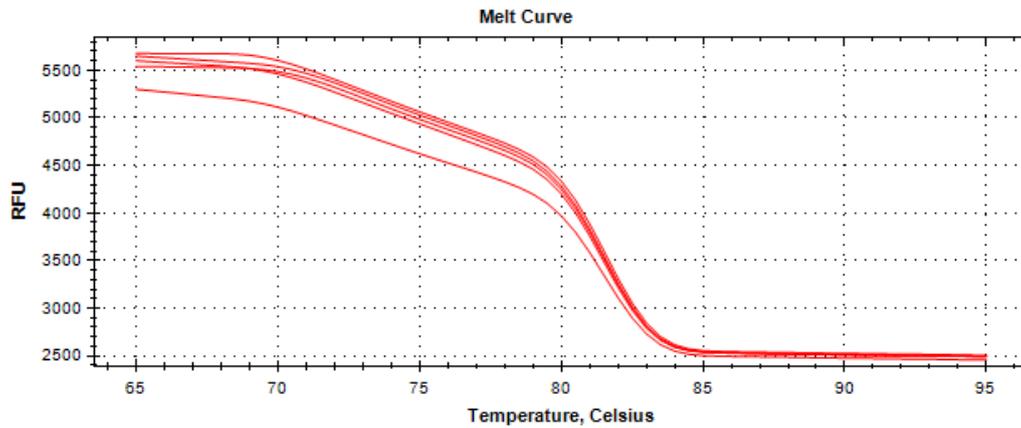


Figura 4.2 Curve di Melting di un insieme di campioni valutati per lo stesso target in un esperimento di real-time PCR.

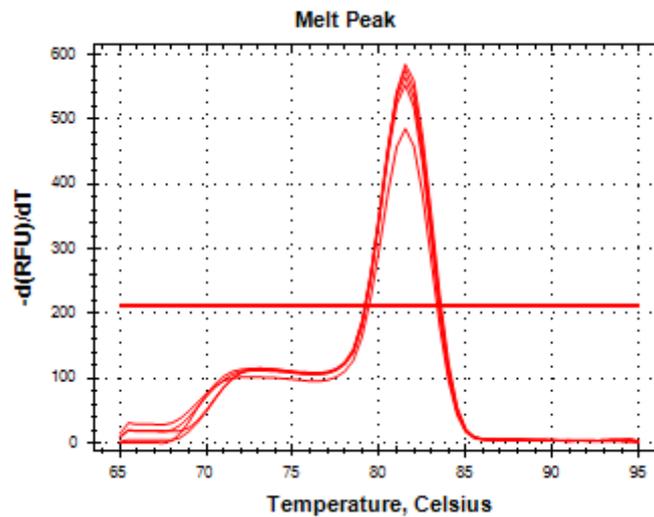
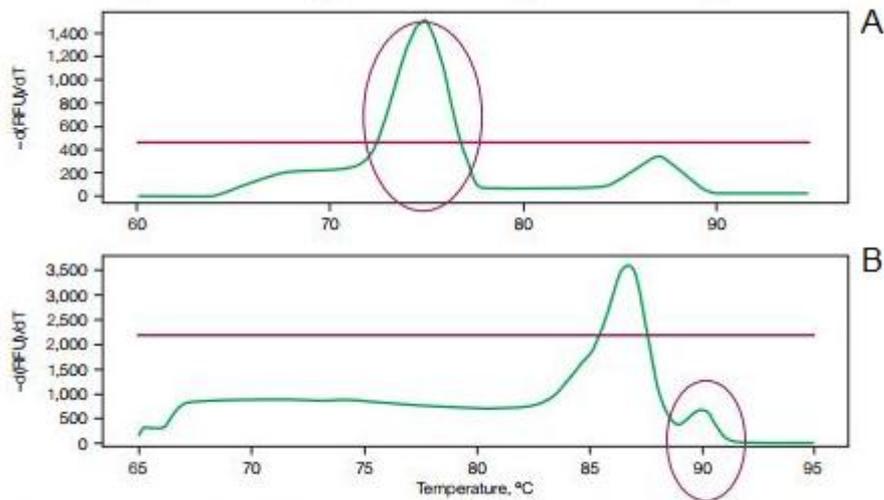


Figura 4.3 Rappresentazione della curva di melting in forma derivata di un insieme campioni valutati per lo stesso target in un esperimento di real-time PCR .

La figura 4.3, ci fornisce una visione completa sulle dinamiche di fusione. Il secondo punto di flesso della figura 4.2, corrisponde al picco dell'immagine 4.3.



**Figura 4.4** Profili di curve di melting con segnale aspecifico. **A:** presenza di dimeri di primer,  $T_m < 80^\circ\text{C}$  ; **B:** prodotti con bassa specificità.

I profili delle curve di melting ci forniscono uno strumento per valutare la specificità dell'amplificato. Un amplificato di PCR specifico, di solito di lunghezza intorno al centinaio di paia di basi, ha una temperatura di melting  $T_m > 80^\circ\text{C}$ ; il picco della curva si troverà quindi in corrispondenza di questo range di valori. I dimeri di primer, prodotti aspecifici dovuti al legame tra primer parzialmente complementari, dando amplificati più corti, si denaturano a una temperatura  $T_m < 80^\circ\text{C}$ ; la loro presenza è quindi identificata nella curva da un picco a temperature inferiori di  $80^\circ\text{C}$ . Altri prodotti di amplificazione non specifici hanno profili di melting diversi rispetto al campione amplificato, solitamente si notano più picchi a temperature diverse.

### 4.3 Analisi quantitativa

La quantificazione del segnale di fluorescenza di real-time PCR viene effettuata seguendo due strategie diverse. I livelli di espressione genica possono essere riportati tramite *quantificazione assoluta* o *relativa*. La prima rappresenta una misura del numero di copie di DNA target di partenza e viene valutata utilizzando una *curva di calibrazione*. La quantificazione relativa, invece, rappresenta una misura relativa dei livelli di espressione rispetto a una specifica condizione sperimentale[M. Filion, 2012].

Per ottenere dati biologicamente significativi, con entrambe le metodologie i livelli di espressione dei geni target devono essere normalizzati rispetto ai livelli di espressione di geni di riferimento, normalmente costanti nelle varie condizioni sperimentali.

## 4.4 Quantificazione assoluta

Il metodo della *quantificazione assoluta* prevede di comparare il valore soglia  $C_T$  dei campioni con una curva standard, detta curva di calibrazione ottenuta amplificando diluizioni scalari di quantità note di template. Una volta validate le curve di calibrazione sono altamente riproducibili e permettono la creazione di dati altamente specifici e sensibili[S.Bustin 2004]. In questo tipo di quantificazione per standardizzare la quantità di amplificato ottenuta rispetto alla quota di campione si utilizzano dei normalizzatori generalmente identificati dal numero delle cellule o i  $\mu\text{g}$  di acidi nucleici caricati nella reazione di PCR.

### 4.4.1 Curve Standard

Per costruire una curva standard si utilizza un campione di concentrazione nota del quale si esegue una serie di diluizioni. Le concentrazioni finali scelte per la curva standard devono comprendere valori nell'intervallo ipotizzato per il campione incognito. Ai campioni diluiti, amplificati in una reazione di real-time PCR, saranno associati diversi valori di  $C_T$ , che rappresentati in funzione della concentrazione del campione, identificheranno la curva di calibrazione. Spesso la curva di calibrazione è rappresentata in un diagramma  $C_T / \log(\text{concentrazione})$ . La curva standard, costruita dal modello di campione diluito, può essere quindi utilizzata per determinare la concentrazione del target nel campione incognito, interpolando il valore di  $C_T$ , del campione incognito sulla retta.

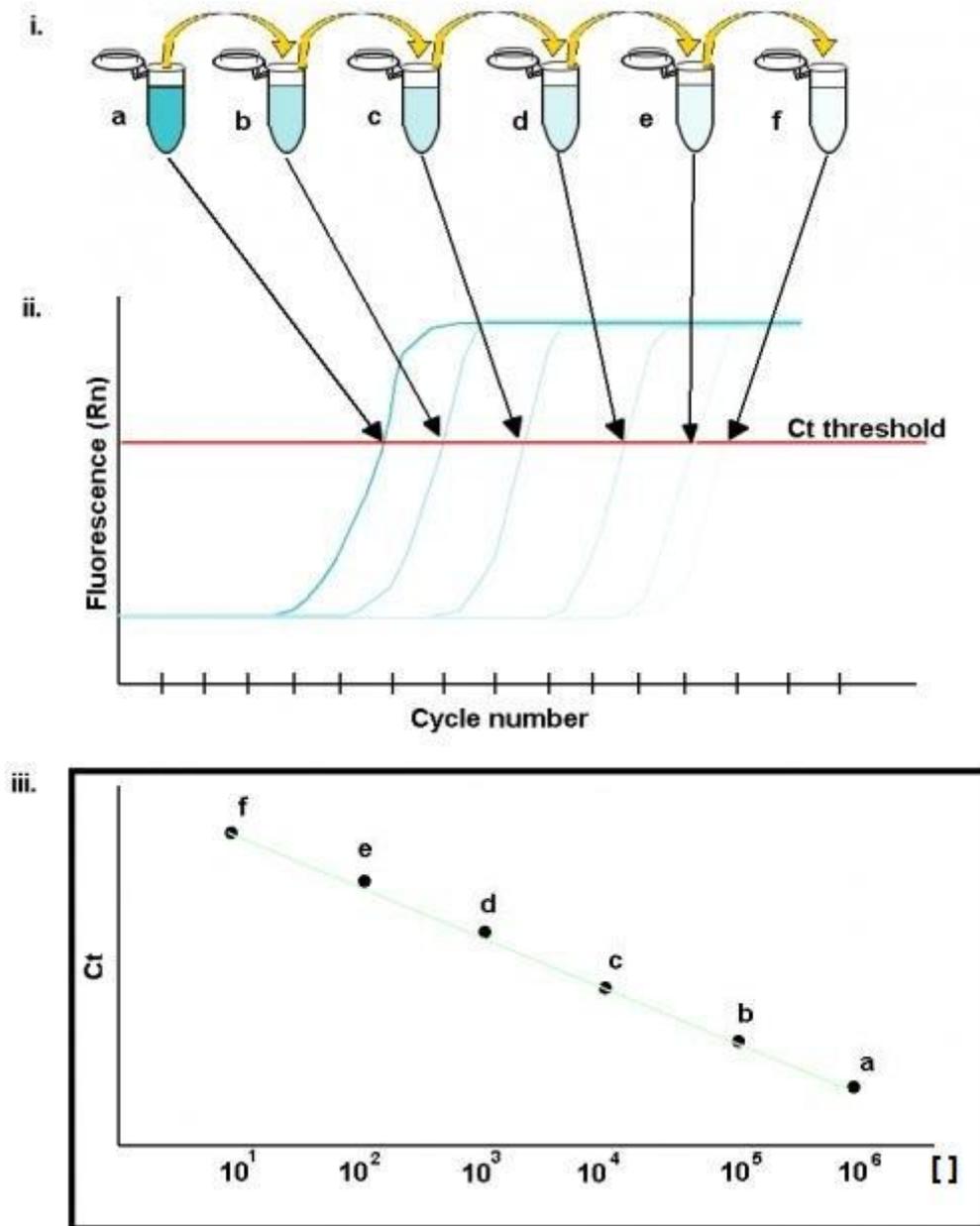


Figura 4.5 i) diluizioni seriali del campione noto; ii) curve di amplificazioni ottenute dai campioni e rispettivi cicli soglia; iii) curva standard di calibrazione.

Nell'analisi di quantificazione assoluta l'accuratezza della retta di calibrazione è verificata valutando il valore del coefficiente di correlazione  $R^2$  della retta. Questo valore fornisce un parametro di misura della precisione con cui la retta approssima i dati sperimentali; rappresenta quindi una valutazione della correttezza delle diluizioni effettuate. Quanto più  $R^2$  è prossimo al valore 1 tanto più la retta riproduce l'andamento dei dati sperimentali: buone rette di calibrazione hanno  $R^2$  pari a 0,99.

Un altro parametro che si può valutare è l'intercetta con l'asse y, nel grafico semilogaritmico della curva standard, esso corrisponde al numero di cicli necessari per rilevare una copia del DNA target (limite teorico di rilevazione della reazione). Questo valore può essere utile per confrontare diversi sistemi e target di amplificazione.

Il metodo di quantificazione assoluta, basato sulle diluizioni scalari, ha il vantaggio di essere altamente riproducibile. Occorre però tener presente che curve di calibrazione costruite con DNA o cDNA possono dare risultati di amplificazioni non strettamente confrontabili coi campioni in esame, in quanto nel caso di utilizzo di cDNA bisogna tener conto anche della resa della reazione di retro-trascrizione che può differire da campione a campione in base alla qualità dell'RNA di partenza. [M. Filion, 2012]

## 4.5 Quantificazione relativa

La quantificazione relativa determina i cambiamenti di livello di espressione di un gene target in più campioni rispetto a un controllo interno, ovvero un gene di riferimento che può essere co-amplificato insieme al target di interesse. Un gene di riferimento (*reference gene*) è un gene il cui livello di espressione rimane costante in tutti i campioni del test, e la cui espressione non è influenzata dal trattamento sperimentale di cui si vuole studiare gli effetti. L'impiego di un gene di riferimento è vantaggioso nei casi dove non è possibile una precisa quantificazione del materiale di partenza e quando è disponibile solo una piccola quantità di template iniziale.

Per ogni esperimento di quantificazione relativa sono necessari:

- Target: la sequenza di cDNA da analizzare;
- Calibratore: il campione da usare come riferimento per l'analisi comparativa;
- Controllo endogeno: un gene espresso costitutivamente in tutti i campioni (*gene di riferimento*), necessario per normalizzare i dati rispetto alla quantità di cDNA caricato e a possibili variazioni di efficienza della reazione.

Il calibratore si impiega nel confronto tra campioni multipli; l'espressione del gene target in tutti i campioni è rappresentata da un incremento o decremento relativo rispetto al calibratore. Abitualmente viene scelto come calibratore il campione in condizione di assenza di trattamento, chiamata anche condizione di controllo.

Il vantaggio del metodo basato sul normalizzatore con geni di riferimento, invece che le unità di massa, consiste nell'eludere la necessità di una quantificazione assoluta e di una precisione molto fine della quantità di template caricato per la reazione. E' molto conveniente quando, eseguendo esperimenti di espressione genica, il materiale di partenza è limitato. L'inconveniente è che il procedimento necessita di geni di riferimento che abbiano un livello di espressione costante in tutti i campioni, cioè che non cambi a causa del trattamento farmacologico in esame. La scelta dei geni di riferimenti non è banale, e va valutata caso per caso. Sono disponibili diversi software gratuiti come *geNorm*, *BestKeeper* o *NormFinder* per la scelta e validazione dei geni di riferimento[M.Filion, 2012]. Tra i *reference gene* utilizzati troviamo spesso il *GAPDH*, la beta 2-microglobulina (*B2M*) e il gene *HPRT1*. Recentemente è stato proposto l'impiego di più geni di riferimento contemporaneamente per realizzare una quantificazione più accurata.

Per calcolare l'espressione di un gene target in relazione a un gene di riferimento sono stati proposti vari modelli matematici. I calcoli si basano sul confronto di distinti cicli, ad esempio i valori soglia ( $C_T$ , cycle threshold) corrispondenti ad un livello costante della fluorescenza o l'acquisizione di  $C_T$  rispettando un determinato algoritmo matematico. In questo tipo di analisi il risultato che si ottiene è un rapporto (*ratio*). Ad oggi sono stati sviluppati diversi modelli matematici che determinano l'espressione relativa[S.Bustin 2004].

I metodi successivamente descritti, derivano dall'equazione che esprime l'amplificazione esponenziale della PCR:

$$X_n = X_0(1 + E_X)^n$$

dove  $X_n$  è il numero di molecole del target al ciclo  $n$  della reazione,  $X_0$  è il numero iniziale di molecole del target,  $E_X$  è l'efficienza di amplificazione del target e  $n$  è il numero di cicli. Il ciclo soglia  $C_T$  indica il numero di ciclo in cui la curva di amplificazione del gene target interseca un valore di soglia fissata. Si può quindi esprimere l'amplificazione esponenziale raggiunta al ciclo soglia in questo modo:

$$X_T = X_0(1 + E_X)^{C_{T,X}} = K_X$$

dove  $X_T$  rappresenta il numero di soglia delle molecole target,  $C_{T,X}$  è il ciclo di soglia per l'amplificazione del target e  $K_X$  è una costante.

Riformulando l'espressione precedente otteniamo:

$$X_0 = K_X(1 + E_X)^{-C_{T,X}}$$

#### 4.5.1 Metodo del $\Delta C_T$

Il metodo del  $\Delta C_T$  si basa sulla differenza di ciclo soglia tra il calibratore e il campione di test.

$$\Delta C_T = C_{T(test)} - C_{T(calibrator)}$$

$$Ratio = E^{-\Delta C_T}$$

dove E è l'efficienza della reazione. Se assumiamo l'ipotesi di un'efficienza d'amplificazione ideale, l'equazione precedente diventa:

$$Ratio = 2^{-\Delta C_T}$$

La derivazione di questo metodo si basa sulle equazioni che esprimono l'amplificazione esponenziale di PCR. Infatti partendo da queste equazioni, e facendo il rapporto tra  $X_0$  di un qualsiasi campione  $q$  e  $X_0$  del campione di controllo o calibratore ( $cal$ ) risulta:

$$\frac{X_{0,q}}{X_{0,cal}} = \frac{K_X(1 + E_X)^{-C_{T,q}}}{K_X(1 + E_X)^{-C_{T,cal}}} = (1 + E_X)^{-\Delta C_T}$$

dove  $\Delta C_T = C_{T,q} - C_{T,cal}$ .

Assumendo un'amplificazione ottimizzata, l'efficienza risulta essere del 100% e sostituendo tale valore nell'equazione precedente, si ottiene l'espressione del ratio nel caso di amplificazione ideale  $2^{-\Delta C_T}$  [K.J. Livak et al, 2001].

#### 4.5.2 Metodo Livak o del $\Delta\Delta C_T$

Il metodo del  $\Delta\Delta C_T$  è ampiamente utilizzato e facilmente eseguibile. Esso assume l'ipotesi che sia i geni target che quelli di riferimento siano amplificati con un'efficienza che si avvicina al 100%, con uno scarto accettabile del 5%. Prima di utilizzare questo metodo è necessario verificare la validità di questa ipotesi, controllando le efficienze di amplificazione sia dei geni target che dei geni di riferimento. Il valore di efficienza per i primer assay commerciali validati da Bio-rad è indicato nel "Assay Validation Report", che consiste in un documento che contiene per ogni coppia di primer disegnati per l'amplificazione di uno specifico gene, la descrizione completa delle loro caratteristiche.

Nel caso di primer disegnati o trovati in letteratura, il valore di efficienza deve essere calcolato amplificando diluizioni scalari di una miscela di template calibratore e di test e costruendo una retta che interpola i fattori di diluizioni del template in scala logarimica in funzione dei valori di ciclo soglia ottenuti. L'efficienza si calcola come:

$$E = 10^{(-1/slope)} - 1$$

Una volta stabilito che l'ipotesi iniziale è corretta si può procedere normalizzando il  $C_T$  del gene target rispetto a quello del gene di riferimento (*ref*), sia per il campione trattato (*test*) sia per il campione di controllo (*cal*):

$$\Delta C_{T(test)} = C_{T(target, test)} - C_{T(ref, test)} \quad \Delta C_{T(cal)} = C_{T(target, cal)} - C_{T(ref, cal)}$$

In secondo luogo si normalizza il  $\Delta C_T$  del campione di test con il  $\Delta C_T$  del calibratore:

$$\Delta\Delta C_T = \Delta C_{T(test)} - \Delta C_{T(cal)}$$

E infine si calcola l'espressione del rapporto (*ratio*):

$$Ratio = 2^{-\Delta\Delta C_T}$$

Il risultato ottenuto quantifica l'incremento (o decremento) del gene target nel campione di test relativo al campione calibratore e normalizzato rispetto all'espressione del gene di riferimento [K.J. Livak et al, 2001].

Questa normalizzazione compensa qualsiasi possibile differenza della quota di materiale caricato nelle varie reazioni di PCR. Se l'efficienza non è ideale, si modifica l'ultima equazione sostituendo a 2 il valore dell'efficienza reale.

Supponendo un valore di  $\Delta\Delta C_T > 0$ , che corrisponde alla condizione in cui  $\Delta C_{T(test)} > \Delta C_{T(cal)}$ , e sostituendo il suo valore nell'espressione del rapporto si otterrà un **ratio**  $< 1$ , indice di sottoespressione del gene di interesse nel campione trattato (test), rispetto al campione di controllo (calibratore).

Al contrario un valore di  $\Delta\Delta C_T < 0$ ; ovvero  $\Delta C_{T(test)} < \Delta C_{T(cal)}$ , darà un **ratio**  $> 1$ , che corrisponde ad un incremento di espressione del gene nel campione trattato rispetto al controllo.

Attraverso queste valutazioni si identificano i livelli di espressione dei geni target nei vari campioni.

Questo metodo, come quello del  $\Delta C_T$ , deriva dalle equazioni di amplificazione esponenziale. Infatti considerando queste equazioni sia per il campione target:

$$X_T = X_0(1 + E_X)^{C_{T,X}} = K_X ,$$

sia per il gene di riferimento endogeno:

$$R_T = R_0(1 + E_R)^{C_{T,R}} = K_R ;$$

e facendo il rapporto tra l'espressione di  $X_T$  e l'espressione di  $R_T$ :

$$\frac{X_T}{R_T} = \frac{X_0(1 + E_X)^{C_{T,X}}}{R_0(1 + E_R)^{C_{T,R}}} = \frac{K_X}{K_R} = K$$

Assumendo le efficienze uguali sia per il target che per il reference:  $E_X = E_R = E$ :

$$\frac{X_0}{R_0}(1 + E)^{C_{T,X} - C_{T,R}} = K$$

Oppure:

$$X_n (1 + E)^{\Delta C_T} = K ,$$

dove  $X_n$  è pari alla quantità normalizzata del target  $\frac{X_0}{R_0}$ , e  $\Delta C_T$  è pari alla differenza di cicli soglia tra il target e il reference. Modificando quest'ultima equazione possiamo ottenere:

$$X_n = K(1 + E)^{-\Delta C_T}$$

La dimostrazione finale si ottiene facendo il rapporto tra l'espressione  $X_n$  di un qualsiasi campione  $q$ , e  $X_n$  del calibratore (*cal*):

$$\frac{X_{N,q}}{X_{N,cal}} = \frac{K(1 + E)^{-\Delta C_{T,q}}}{K(1 + E)^{-\Delta C_{T,cal}}} = (1 + E)^{-\Delta \Delta C_T}$$

Qui  $-\Delta \Delta C_T = -(\Delta C_{T,q} - \Delta C_{T,cal})$ .

Per coppie di primer progettate in modo da dare amplificati di 150bp (*paia di basi*) e con concentrazioni di primer e  $Mg^{2+}$  propriamente ottimizzate si può considerare l'efficienza pari al 100%. Quindi la quantità di target, normalizzata con il gene di riferimento e con la condizione di controllo risulta essere  $2^{-\Delta \Delta C_T}$ , come ricavato inizialmente [K.J. Livak et al, 2001].

Esiste anche una variante del metodo di Livak, che consiste nell'utilizzare la differenza tra i valori di  $C_T$  del riferimento (*reference*) e del target per ogni campione dell'esperimento.

$$Ratio_{(reference / target)} = 2^{C_T(reference) - C_T(target)}$$

Le assunzioni matematiche per questo approccio sono le stesse del metodo del  $\Delta \Delta C_T$ .

#### **4.5.3 Metodo Pfaffl**

Il metodo Pfaffl è un modello per la quantificazione relativa che utilizza una *correzione d'efficienza*. Viene utilizzato quando l'efficienza di due amplificati non è la stessa. Nel software per l'analisi si indicano i geni scelti come riferimento, e si precisa il loro valore d'efficienza. La correzione è applicata separatamente per il gene target ( $E_{target}$ ) e il gene di riferimento ( $E_{ref}$ ).

$$ratio = \frac{(E_{target})^{\Delta C_{t,target} (calibrator - test)}}{(E_{ref})^{\Delta C_{t,ref} (calibrator - test)}}$$

Questa equazione assume la stessa efficienza di amplificazione nel campione di test e nel calibratore, sia per il gene target che per quello di riferimento. Questa efficienza non è però necessariamente uguale tra gene target e reference.

$$ratio = \frac{(E_{ref})^{C_{t, test}}}{(E_{target})^{C_{t, test}}} \div \frac{(E_{ref})^{C_{t, calibrator}}}{(E_{target})^{C_{t, calibrator}}}$$

Del metodo Pfaffl esistono alcune varianti basate su campioni multipli e/o multipli geni di riferimento. Il modello di calcolo basato su campioni multipli è il seguente:

$$ratio = \frac{(E_{target})^{\Delta C_{t, target} (MEAN calibrator - MEAN test)}}{(E_{ref})^{\Delta C_{t, ref} (MEAN calibrator - MEAN test)}}$$

Mentre il modello di calcolo, con correzione d'efficienza, basato su campioni multipli e multipli geni di riferimento (chiamati REF index) assume la forma:

$$ratio = \frac{(E_{target})^{\Delta C_{t, target} (MEAN control - MEAN sample)}}{(E_{REF index})^{\Delta C_{t, ref index} (MEAN control - MEAN sample)}}$$

In questi modelli l'espressione del gene target viene normalizzata con più geni di riferimento, per ottenere un'analisi più accurata.

Il metodo del  $\Delta\Delta C_T$  è un caso particolare del metodo Pfaffl. Infatti se sostuiamo nel metodo Pfaffl  $E_{target} = E_{ref} = 2$  semplificando si ottengono le espressioni relative al metodo del  $\Delta\Delta C_T$ :

$$ratio = \frac{2^{\Delta C_{t,target} (calibrator - test)}}{2^{\Delta C_{t,ref} (calibrator - test)}}$$

$$= 2^{-[C_{t,target} (test) - C_{t,target} (calibrator)] - [C_{t,ref} (test) - C_{t,ref} (calibrator)]}$$

$$= 2^{-\Delta\Delta C_T}$$

## 4.6 MIQE

Il MIQE (*Minimum Information for publication of Quantitative real-time PCR Experiments*) consiste nelle linee guida da seguire per la pubblicazione di esperimenti di real-time PCR. L'utilizzo di queste linee guida garantisce l'inclusione di dati e informazioni chiave nella relazione dei risultati sperimentali, in modo tale da poter valutare criticamente la qualità dei risultati presentati o ripetere gli esperimenti di altri ricercatori. Questo è risultato da un'iniziativa dell'"*International Real-time PCR Data Markup Language*" (RDML), un'associazione internazionale finalizzata a creare uno standard universale per permettere lo scambio di dati ottenuti tramite real-time PCR. Le informazioni minime necessarie per valutare un esperimento di real-time PCR comprendono una lista di tutte le condizioni sperimentali rilevanti e un'informazione completa su reagenti, sequenze e metodi di analisi utilizzati. L'utilizzo di queste linee guida è fondamentale per uniformare i dati presentati nelle letterature scientifica e permettere un'univoca interpretazione dei risultati [Stephen A. Bustin et al, 2009]. L'RDML inoltre si adopera per creare un formato standard per i dati di real-time PCR e uniformare la terminologia tecnica.

## 4.7 Significatività statistica

L'obiettivo di un'analisi di espressione genica è definire conclusioni biologicamente significative basate sui risultati ottenuti. Per ottenere conclusioni biologicamente significative è importante valutare la significatività statistica dei risultati. Infatti senza delle corrette analisi statistiche, l'interpretazione dei dati di real-time PCR potrebbe portare il ricercatore a conclusioni errate come i cosiddetti "*falsi positivi*", cioè campioni classificati erroneamente come positivi per la manifestazione di una particolare caratteristica, con potenziale danno per il paziente nelle applicazioni cliniche.

Esistono diversi software per fare analisi statistica, come ad esempio GraphPad Prism®. Questo programma fornisce un aiuto all'utente per decidere quale è la corretta analisi statistica da effettuare [M.T. Dorak, 2006]. Nel software è possibile delineare le caratteristiche dell'analisi, scegliendo tra test:

- *Parametrici o non parametrici*

- *One-tailed o two-tailed*
- *Paired o unpaired*

I test parametrici si basano sull'ipotesi che i dati seguano una distribuzione gaussiana, tra questi troviamo il t-test e l'analisi della varianza. I test non parametrici vengono utilizzati quando non sono possibili ipotesi a priori sulla distribuzione. I paired test sono effettuati quando le misurazioni sono accoppiate naturalmente e le variabili misurate sono dipendenti. Sono utilizzati ad esempio in esperimenti in cui vengono somministrati trattamenti diversi a cellule della stessa cultura o in saggi in cui un parametro è misurato prima e dopo il trattamento [Ramon Goni et al, 2009]. Dal test effettuato, dopo aver calcolato media e deviazione standard, si ottiene un parametro definito p-value. Questo è un parametro importante, indice del livello di significatività osservato. Solitamente una differenza tra due serie di dati risulta statisticamente significativa se caratterizzata da  $p\text{-value} < 0.05$ ; valori minori ( $p\text{-value} < 0.01$  o  $p\text{-value} < 0.001$ ) indicano una più alta significatività. L'interpretazione di un p-value è collegata agli obiettivi sperimentali che ci si pone [J.S Yuan et al, 2006].

In un'analisi di espressione genica per definire conclusioni affidabili e significative occorre che il dato ottenuto sia ripetibile. A questo scopo si utilizzano replicati biologici indipendenti e replicati tecnici. Il replicato tecnico consiste nella ripetizione della stessa tecnica analitica di misurazione sul medesimo campione biologico: serve semplicemente per confermare il dato campionario ed individuare, nella deviazione, la variabilità tecnica legata all'operatore o alla misurazione. Il replicato biologico consiste invece nella ripetizione della stessa tecnica analitica su una nuova campionatura biologica, e serve per individuare la variabilità biologica del fenomeno biologico. Il numero minimo di replicati dipende dal test statistico e dal potere statistico desiderati.

# 5. CASO DI STUDIO: ANALISI DI ESPRESSIONE GENICA TRAMITE REAL-TIME PCR

## 5.1 Scopo e metodi

Durante l'esperienza di tirocinio presso il laboratorio di Ingegneria Cellulare e Molecolare (ICM) "S. Cavalcanti" dell'Università di Bologna, ho assistito a diversi esperimenti di real-time PCR. Lo scopo dello studio era la valutazione dell'espressione di due geni target, *E-caderina (CDH1)* e *vimentina (VIM)*, in cellule tumorali trattate con una molecola sintetizzata da collaboratori del dipartimento di Farmacia e BioTecnologie (FaBiT) dell'Università di Bologna, per poter verificare il potenziale antitumorale del farmaco.

Per verificare gli effetti antitumorali della molecola, denominata **LF26**, sono stati scelti due geni, indici di un fenotipo epiteliale meno aggressivo (*CDH1*) e di un fenotipo mesenchimale, più aggressivo che correla con la tendenza a metastatizzazione (*VIM*). Come modello sperimentale sono state utilizzate due linee cellulari tumorali, **HCT116** (carcinoma del colon) e **A549** (adenocarcinoma del tessuto polmonare). Le linee cellulari sono state coltivate in terreno RPM1-1640 integrato con il 10% di siero fetale bovino, 100U/mL di Penicillina, 100 µg/mL di Streptomicina e 2 mM di L-Glutamina; e mantenute a 37°C in un incubatore con atmosfera umidificata e 5% di CO<sub>2</sub>. La molecola LF26 è stata disciolta in una soluzione acquosa di dimetilsolfossido (DMSO) in proporzione 1:1 e conservata a 4°C.

Le cellule HCT116 e A549 sono state seminate 24 ore prima del trattamento farmacologico alla densità di 17000 cellule/cm<sup>2</sup> in duplicato, prevedendo una condizione di controllo, cioè assenza di farmaco e una condizione di trattamento con LF26. La molecola è stata aggiunta al terreno alla concentrazione di 5µM e lasciata per un tempo di trattamento pari a 48 ore.

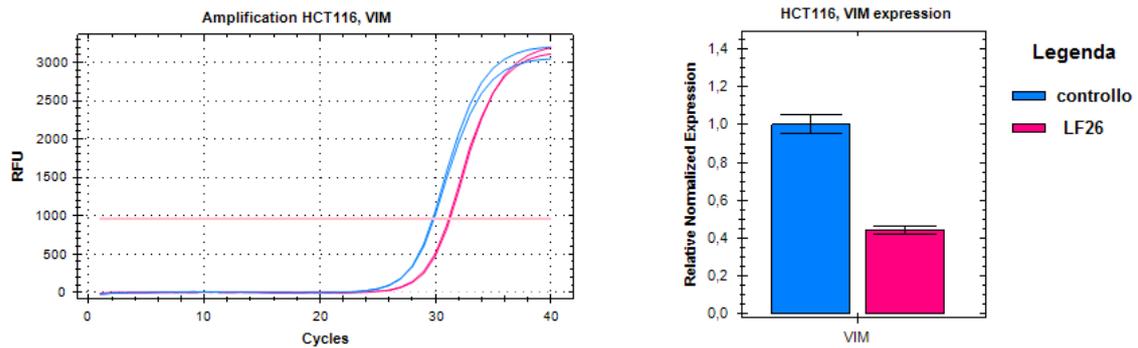
L'RNA è stato isolato usando un protocollo standard con PureZOL™ (732-6890, Bio-Rad) e trattato con DNAsi (AM1906, Ambion) secondo le istruzioni del fabbricante. 1µg dell'RNA totale è stato retro-trascritto in cDNA con l' iScript™ cDNA Synthesis Kit (170-8891, Bio-Rad). Il cDNA è stato diluito 1:10, e 5µL della

soluzione sono stati amplificati in real-time PCR utilizzando l'SsoAdvanced™ SYBR® Green Supermix (172-5261, Bio-rad). I primer validati, specifici per i geni target (*CDH1* e *VIM*) e per i geni di riferimento (*GAPDH*, *B2M*, *HPRT1*) sono stati acquistati dall'azienda Bio-rad (PrimePCR™ Assay, qHsaCID0015365, qHsaCID0012604, qHsaCED0038674, qHsaCID0015347 e qHsaCID0016375, rispettivamente).

Gli esperimenti di real-time PCR sono stati eseguiti con duplicati tecnici con cDNA ottenuto da quattro esperimenti indipendenti, corrispondenti quindi a quattro replicati biologici. Lo strumento di PCR utilizzato in laboratorio è il CFX Connect™ Real-Time PCR Detection System di Bio-rad. Il protocollo termico eseguito è il seguente: 95°C per 2 minuti necessari per l'attivazione dell'enzima, seguito da 40 cicli composti da due step: 95°C per 5 secondi (denaturazione) e 60°C per 30 secondi (annealing ed estensione); il protocollo include anche la costruzione della curva di melting prevedendo un incremento di temperatura tra i 65 e 95°C con step di 0.5°C.

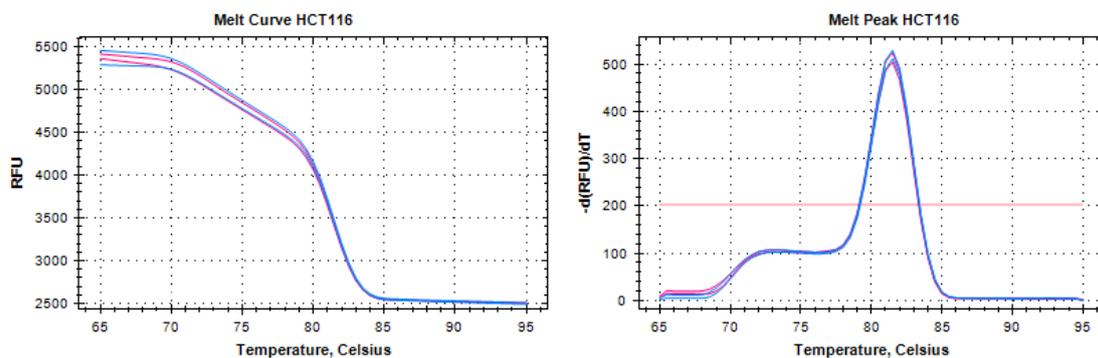
## 5.2 Analisi dell'espressione genica

L'analisi dei dati di espressione genica è stata effettuata con il CFX Manager™ Software di Bio-rad. Sono state valutate le curve di amplificazione di ogni campione e dopo aver settato la linea di soglia, sono stati valutati i rispettivi cicli soglia, cioè i punti di intersezione delle curve di amplificazione con la linea di soglia. I dati sono stati normalizzati rispetto all'espressione dei geni di riferimento utilizzati (*GAPDH*, *B2M*, *HPRT1*). Nella figura 5.1 è riportato il risultato di un esperimento rappresentativo dove si nota la correlazione tra il valore del ciclo soglia  $C_T$  e il livello di espressione genica corrispondente riportato in istogramma. Il livello di espressione genica che si ottiene rappresenta il valore medio di espressione nella popolazione cellulare.



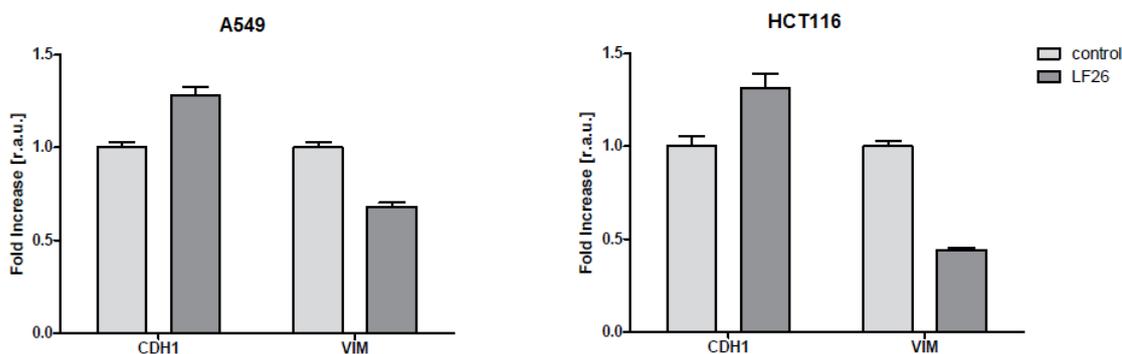
**Figura 5.1** Curva di amplificazione per VIM nelle cellule HCT116 trattate con 5 $\mu$ M di LF26 e relativo istogramma di espressione genica.

E' stata effettuata anche l'analisi delle curve di melting di ogni amplificato specifico per ogni gene. La figura 5.2 riporta un esempio delle curve di melting di VIM relative alle cellule HCT116.



**Figura 5.2** Curve di melting e picchi di fusione dell'amplificato specifico per VIM relativi alle cellule HCT116. L'andamento delle curve mostra un'ottima specificità dell'amplificato.

Per effettuare la normalizzazione tra i quattro esperimenti di real-time PCR, il software permette di analizzare i risultati delle quattro corse insieme costruendo un cosiddetto "gene study". Questo metodo prevede di normalizzare differenze tra le corse grazie all'utilizzo di un calibratore interno, cioè lo stessa miscela di cDNA templato amplificata con la stessa coppia di primer in ogni sessione di real-time PCR. Il risultato del "gene study" è costituito da valor medio e deviazione standard relativa alle quattro sessioni di analisi effettuate ed è riportato nella figura 5.3.



**Figura 5.3** Analisi di espressione genica dei livelli di *CDH1* e *VIM*, nelle linee cellulari A549 e HCT116 trattate per 48 ore con 5 $\mu$ M di LF26.

L'analisi di espressione genica ha evidenziato potenzialità terapeutica della molecola LF26 come agente di riprogrammazione verso un fenotipo meno aggressivo. Infatti LF26 si è dimostrato efficace nell'indurre una modulazione dell'espressione genica, determinando un incremento del livello di espressione di CDH1, e un decremento del livello di espressione di VIM, in entrambi i modelli cellulari tumorali.

I dati sono stati validati attraverso un'analisi statistica effettuata tramite il software GraphPad Prism<sup>®</sup>. E' stato effettuato un *t-test unpaired two-tailed* sui dati ottenuti dagli esperimenti, ottenendo alti valori di significatività statistica (p-value < 0.01): i dati risultano biologicamente significativi. Quindi la diversità di espressione tra i campioni relativi alla condizione di controllo o di trattamento con LF26 è significativa.

# Conclusioni

Questo lavoro di tesi evidenzia come la real-time PCR sia un metodo innovativo di PCR quantitativa, con enormi potenzialità rispetto alla PCR convenzionale. Considerando che la rilevazione del prodotto di amplificazione avviene durante i cicli di PCR, il processo d'analisi risulta facilmente automatizzabile. I vantaggi di questa tecnica sono l'elevata sensibilità, la rapidità e la possibilità di effettuare analisi con limitate quantità del campione di partenza. Inoltre l'intero processo di amplificazione e analisi viene eseguito nella stessa provetta di reazione, riducendo la possibilità di contaminazione del prodotto ed eliminando la necessità di manipolazioni post-PCR, come ad esempio l'elettroforesi su gel d'agarosio, indispensabile per la verifica del risultato di amplificazione della PCR convenzionale. Queste prerogative hanno fatto sì che la real-time PCR venisse adottata in diversi campi applicativi, come la *ricerca*, l'*ingegneria genetica*, e la *diagnostica clinica*. La grande diffusione di questa tecnologia è dimostrata sia dal numero crescente di aziende che commercializzano la strumentazione e i reagenti necessari, sia dalla numerosità di pubblicazioni scientifiche che presentano dati ottenuti con questa tecnica.

La precisione e la sensibilità della tecnica sono tali da permettere di valutare anche minime variazioni del livello di espressione di geni d'interesse. I possibili sviluppi nel campo della ricerca clinica per l'analisi di geni coinvolti in particolari patologie e l'elaborazione di possibili trattamenti farmacologici sono notevoli. Come prospettiva futura ci si aspetta di ridurre i costi, in modo da permettere l'accesso globale a questa tecnologia, e di aumentare la velocità e la portabilità delle macchine di real-time PCR, riducendo i tempi di processo[M. A. Valasek et al, 2005].

Concludendo, la real-time PCR ci permette di analizzare l'espressione genica, la cui comprensione restringe la distanza nella conoscenza tra le "istruzioni" (il genoma) e le "funzioni" (prodotti genici) in biologia.

# BIBLIOGRAFIA

A. Scialpi, A. Mengoni: *“La PCR e le sue varianti: quaderno di laboratorio.”* Firenze 2008, Firenze University Press

Campbell, Reece, Simon: *“L'essenziale di biologia”* terza edizione, Pearson Benjamin Cummings 2008

D. L. Nelson, M. M. Cox: *“Introduzione alla biochimica di Lehninger”*, quarta edizione, Zanichelli 2011

D. L. Nelson, M. M. Cox: *“I principi di biochimica di Lehninger”*, terza edizione, Zanichelli 2002

D. Halliday, R. Resnick, K.S. Krane: *“Fisica 2”*, quarta edizione, casa editrice Ambrosiana, Milano 1992

Ian M. Mackay: *“Real-Time PCR in Microbiology . From Diagnosis to Characterization.”* Norfolk 2007, Caister Academic Press

J. D. Watson, M. Levine, A. Gann, R. Losick, T. A. Baker, S. P. Bell: *“Biologia molecolare del gene”* Zanichelli sesta edizione, 2009

J. M. J. Logan, K. J. Edwards, N. A. Saunders: *“Real-time PCR: Current Technology and Applications”* Norfolk 2009 Caister Academic Press

J. P. Clewley: *“The Polymerase Chain Reaction (PCR) for Human Viral Diagnosis”* 1995 CRC Press Inc.

J. S Yuan, Ann Reed, Feng Chen e C Neal Stewart: *“Statistical analysis of real-time PCR data”* BMC Bioinformatics 2006

K. J. Livak, T. D. Schmittgen: *“Analysis of Relative Gene Expression Data Using RealTime Quantitative PCR and the  $2^{-\Delta\Delta Ct}$  Method”* 2001

Lakowicz *“Principles of Fluorescence Spectroscopy”* Kluwer Academic / Plenum Publishers 1999

M. A. Valasek, J. J. Repa: *“The power of real-time PCR”* Advances in Physiology Education Published 1 September 2005

*M. Filion: "Quantitative Real-time PCR in Applied Microbiology" Caster Academic Press, Norfolk 2012*

*M.T. Dorak: "Real time PCR", 2006 Bios Advanced Methods*

*Ramon Goni, Patricia García e Sylvain Foissac: "The qPCR data statistical analysis" Integromics White Paper - September 2009*

*S. Bustin: "A-Z of Quantitative PCR" \_ International University Line, 2004*

*S. Kennedy, N. Oswald: "PCR Troubleshooting and Optimization: The Essential Guide", Caister Academic Press Norfolk 2011*

*Stephen A. Bustin, Vladimir Benes, Jeremy A. Garson, , Jan Hellemans, Jim Huggett, Mikael Kubista, Reinhold Mueller, Tania Nolan, Michael W. Pfaffl, Gregory L. Shipley, Jo Vandesompele, and Carl T. Wittwer: "The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments" 2009*

*T. Weissensteiner, T. Nolan, S. A. Bustin, H. G. Griffin, A. Griffin: "PCR Technology: Current Innovations", Second Edition CRC Press 2004*

# SITOGRAFIA

1. <https://mygenomix.wordpress.com/tag/dogma-centrale-della-biologia-molecolare/>
2. <http://www.federica.unina.it/scienze-biotecnologiche/biotecnologie-cellulari-e-molecolari/vitalita-e-proliferazione-di-cellule-in-coltura/>
3. [http://xfiles.farmacia.uniba.it/farmol/didattica\\_web/40/argomenti/Analisi%20dei%20Farmaci%20IIa%20parteCTF.pdf](http://xfiles.farmacia.uniba.it/farmol/didattica_web/40/argomenti/Analisi%20dei%20Farmaci%20IIa%20parteCTF.pdf)
4. <http://www.farm.unisi.it/didactics/donnini/Saggi-e-dosaggi-culture-cellulari.pdf>
5. [http://biomedicalgenomics.org/RNA\\_quality\\_control\\_Bioanalyzer.html](http://biomedicalgenomics.org/RNA_quality_control_Bioanalyzer.html)
6. <http://www.thermoscientificbio.com/applications/pcr-and-qpcr/introduction-to-qpcr/>
7. [http://www.genedia.it/italiano/quaderno/Opuscoli%20da%20pubb/RealTimeOPFO\\_tec1.pdf](http://www.genedia.it/italiano/quaderno/Opuscoli%20da%20pubb/RealTimeOPFO_tec1.pdf)
8. [http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin\\_5279.pdf](http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_5279.pdf)
9. <http://www.bio-rad.com/en-us/product/cfx-connect-real-time-pcr-detection-system>
10. <http://www.rdml.org/>