

Biologia molecolare della cellula 2

Giacomo Fantoni

Telegram: @GiacomoFantoni

Github: <https://github.com/giacThePhantom/BioMolCellula2>

4 novembre 2020

Indice

1	Struttura e funzione dei cromosomi	2
1.1	Organizzazione dei cromosomi	2
1.1.1	Ploidia	2
1.1.2	Ulteriore DNA presente nelle cellule	3
1.2	Impacchettamento del DNA cromosomale	3
1.2.1	Il nucleotide	3
1.2.2	DNA eucariotico	4
1.2.3	Il ciclo cellulare e la dinamicità dei cromosomi	4
1.3	Impacchettamento del DNA cromosomale degli eucarioti	4
1.3.1	Istoni	4
1.3.2	Livelli di compattazione	5
1.4	Modifiche covalenti degli istoni	6
1.4.1	Epigenetica	6
1.4.2	Acetilazione	6
1.4.3	Metilazione	7
1.4.4	Fosforilazione	7
1.4.5	Ubiquitinazione e sumoilazione	7
1.4.6	Codice istonico	7
1.5	Complessi rimodellatori dei nucleosomi	8
1.5.1	Ruoli dei rimodellatori della cromatina	8
1.5.2	Sottofamiglie	8
1.6	Variazione nella struttura cromatinica	9
1.6.1	Eucromatina	9
1.6.2	Eterocromatina	9
1.6.3	Effetti della cromatina	9
1.6.4	Nucleolo	9
1.7	Metilazione del DNA	10
1.7.1	DNA metilasi	10
1.7.2	Effetti della metilazione	10
1.7.3	La disattivazione del cromosoma X è un esempio di silenziamento epigenetico della metilazione di cromatina nei mammiferi	11
1.7.4	Imprinting genetico	11
1.8	La separazione dei domini cromatinici da elementi barriera	11
1.8.1	Variegazione da effetto di posizione	11
1.8.2	Elementi di barriera	12
1.9	Elementi richiesti per la funzione dei cromosomi	12

1.9.1	Origine di replicazione	12
1.9.2	Centromeri	12
1.9.3	Telomeri	13
2	Replicazione	15
2.1	Replicazione del DNA semi-conservativa	15
2.1.1	Modelli di replicazione	15
2.2	Il modello dei repliconi	16
2.2.1	Scoperta del modello	16
2.2.2	Origini di replicazione	17
2.3	Identificazione delle origini di replicazione	17
2.3.1	Esperimento	17
2.3.2	Origini di replicazione negli eucarioti	17
2.4	Panoramica della replicazione del DNA	18
2.4.1	Le fasi della replicazione del DNA	18
2.5	Iniziazione	19
2.5.1	Svolgimento dell' <i>Ori</i> nei procarioti - E. coli	19
2.5.2	Svolgimento dell' <i>Ori</i> negli eucarioti	20
2.5.3	DNA elicasi	21
2.5.4	Sintesi di primer a RNA o RNA-DNA dalla DNA primasi	22
2.6	Allungamento	22
2.6.1	Pinza scorrevole	22
2.6.2	Sintesi del DNA	23
2.6.3	DNA polimerasi	23
2.6.4	Fedeltà della polimerizzazione del DNA	23
2.6.5	Sintesi del DNA discontinua	25
2.6.6	Attività del replisoma alla forcella di replicazione	25
2.7	Terminazione	26
2.7.1	Terminazione nei batteri	26
2.7.2	Terminazione negli eucarioti	26
2.8	Replicazione dei telomeri	27
2.8.1	Telomerasi	27
2.8.2	Mantenimento della lunghezza	27
2.8.3	Problemi nel mantenimento della lunghezza del telomero	27
2.8.4	Il limite di Haflick	27
2.9	Correzione degli errori post-replicativa	28
2.10	Mantenimento delle modifiche istoniche	28
2.11	DNA polimerasi specializzate	28
2.11.1	DNA polimerasi batteriche	28
2.11.2	DNA polimerasi specializzate	29
3	Trascrizione	30
3.1	Panoramica della trascrizione	30
3.1.1	Il processo di trascrizione	30
3.1.2	Nomenclatura dei geni	31
3.1.3	Regolazione della trascrizione	31
3.2	L'enzima centrale della RNA polimerasi	31
3.2.1	Struttura	32

3.3	Riconoscimento dei promotori	32
3.3.1	Batteri	33
3.3.2	Eucarioti	33
3.3.3	Formazione del complesso di pre-iniziazione	34
3.3.4	Differenze tra le RNA polimerasi eucariotiche	34
3.4	Iniziazione della trascrizione e transizione a un complesso di allungamento	34
3.4.1	Modello di inizio abortivo	35
3.4.2	Sintesi del RNA	35
3.4.3	Promoter clearance	35
3.5	Allungamento della trascrizione	35
3.5.1	Pause e arresti nella trascrizione	35
3.5.2	Processamento del mRNA	36
3.5.3	Backtrack	36
3.5.4	Problemi dell'allungamento	36
3.6	Terminazione della trascrizione	37
3.6.1	Batteri	37
3.6.2	Eucarioti	37
3.7	Principi della regolazione della trascrizione	38
3.7.1	Meccanismi di regolazione	38
3.7.2	Enhancer	39
3.7.3	Silencer	39
3.7.4	Insulator	39
3.7.5	Struttura delle proteine regolatrici	39
3.7.6	Eventi con effetto sulla trascrizione	39
3.7.7	Confronto con replicazione	40
3.8	Domini leganti il DNA in proteine che regolano la trascrizione	40
3.8.1	Helix-turn-helix	40
3.8.2	Zinc finger	41
3.8.3	Leucine zipper	41
3.8.4	Helix-loop-helix	41
3.8.5	Ribbon-helix-helix	41
3.8.6	Interazioni con gli enhancer	41
3.9	Meccanismi per regolare l'iniziazione della trascrizione nei batteri	42
3.9.1	Operone	42
3.9.2	Regolazione della trascrizione	42
3.10	L'operone <i>lac</i> in <i>E. coli</i>	43
3.10.1	Genetica e analisi funzionale	43
3.10.2	Il repressore <i>LacI</i>	43
3.10.3	L'operone <i>lac</i> e induzione positiva	44
3.11	L'operone triptofano <i>trp</i> in <i>E. coli</i>	44
3.11.1	Regolazione all'iniziazione della trascrizione	44
3.11.2	Regolazione della trascrizione attraverso attenuazione	45
3.12	Regolazione della trascrizione da parte di riboswitches del trascritto	45
3.12.1	Riboswitch di adenina di <i>B. subtilis</i>	45
3.13	Regolazione dell'espressione genica del batteriofago λ in <i>E. coli</i>	46
3.13.1	Il path lisogenico	46
3.13.2	I due cicli vitali del fago λ	46
3.13.3	Il genoma del fago λ e sue interazioni	46

3.13.4	Il ciclo litico del fago λ	47
3.13.5	Passaggio al ciclo lisogenico	49
3.13.6	Determinare il destino dell'infezione	50
3.13.7	Attivazione del profago e ciclo litico nei batteri lisogeni	51
3.14	Regolazione della trascrizione da sistemi di trasduzione del segnale a due componenti	52
3.15	Regolazione dell'iniziazione della trascrizione ed allungamento negli eucarioti	52
3.15.1	Regolazione dell'iniziazione della trascrizione	52
3.15.2	Regolazione dell'allungamento della trascrizione	53
3.16	Il ruolo delle cascate di segnalazione nella regolazione della trascrizione	53
3.16.1	Esempio di risposta immunitaria	54
3.17	Silenziamento genico attraverso imprinting genomico	54
3.17.1	DNA metilato	54
3.17.2	Silenziamento trascrizionale ai mating loci fissione del lievito	54
4	Processamento dell'RNA	56
4.1	Panoramica del processamento del RNA	56
4.1.1	Modifiche al RNA	56
4.2	Processamento di rRNA e di tRNA	56
4.2.1	Procarioti	56
4.2.2	Eucarioti	58
4.3	Modifiche dei nucleotidi di rRNA e di tRNA	58
4.3.1	Modifiche chimiche dei tRNA procarioti ed eucarioti	59
4.3.2	Modifiche chimiche degli rRNA	59
4.4	Capping e poliadenilazione di mRNA	59
4.4.1	Capping dei pre-mRNA eucarioti al 5'	59
4.4.2	Poli-adenilazione 3' dei pre-mRNA eucarioti	60
4.4.3	Legame con altre attività di polimerizzazione del RNA	61
4.5	RNA splicing	61
4.5.1	Scoperta degli introni	62
4.5.2	Tipi di splicing del RNA	62
4.6	Definizione degli esoni e splicing alternativo	66
4.6.1	Splicing alternativo: diversità e complessità delle specie	66
4.6.2	Elementi di sequenze di RNA addizionali	67
4.6.3	Exon shuffling	67
4.7	<i>miRNA</i> e <i>siRNA</i>	67
4.7.1	<i>miRNA</i>	67
4.7.2	<i>siRNA</i>	68
4.8	Ribozimi auto-catalitici	68
4.8.1	Ribozimi hammerhead	68
4.9	RNA editing	68
4.9.1	Editing del trascritto attraverso inserzione o delezione di nucleotidi	69
4.10	Traslocazione nucleo-citoplasmatica di RNA processati	69
4.11	Degradazione di RNA endogeni	70
4.11.1	Stabilità dei mRNA	70
4.11.2	Degradazione del RNA nei procarioti	70
4.11.3	Degradazione del mRNA negli eucarioti	71
4.12	Degradazione di RNA esogeni <i>siRNA</i> <i>CRISPR</i>	71
4.12.1	Interferenza <i>CRISPR</i> nei batteri	71

INDICE

5	RNA regolatori	73
6	Traduzione	74
7	Modifica e targeting delle proteine	75
8	DNA mobile	76
9	Strumenti e tecniche della biologia molecolare	77

Capitolo 1

Struttura e funzione dei cromosomi

1.1 Organizzazione dei cromosomi

L'informazione genetica è impacchettata in almeno una molecola di DNA molto lunga, un cromosoma. Ogni cromosoma contiene una molecola di DNA a doppio filamento con molti geni e regioni di DNA non codificante. Si dicono intergeniche le regioni tra i geni. Se batteri ed archea possiedono cromosomi circolari gli eucarioti ne possiedono di lineari. La distribuzione dei geni varia tra gli organismi: quelli meno complessi tendono ad avere geni ordinati più densamente. La densità genica può variare anche sui diversi cromosomi degli organismi. Il numero dei cromosomi è caratteristico per una specie. Si possono fare incroci tra specie con numero di cromosomi diverso: in questo caso sono incapaci di accoppiarsi durante la prima parte della meiosi e l'incrocio risulta sterile.

1.1.1 Ploidia

Con ploidia si intende quanti cromosomi identici possiede un organismo:

- Aploidia: 1 cromosoma come nel lievito.
- Diploidia: 2 cromosomi come negli umani.
- Poliploidia: più di 2 cromosomi come nelle piante.
- Aneuploidia: un numero anormale di cromosomi, può avvenire in caso di sindromi genetiche o cancro.

La poliploidia viene sfruttata nei prodotti ortofrutticoli per aumentarne le dimensioni.

Aneuploidia e aborti

L'aneuploidia può essere sopportata in un certo numero dagli organismi: si nota per la trisomia del cromosoma 21 (sindrome di Down) e le poliploidie, ma può essere mortale e causare un aborto spontaneo.

Gametogenesi femminile

Si nota come con l'aumentare dell'età della donna aumenta il rischio di aneuploidia per i figli. Questo avviene in quanto ogni donna nasce con tutte le uova diploidi già presenti anche se immature. Queste maturano una alla volta dopo la pubertà una volta al mese. La continuazione della meiosi bloccata comincia il giorno prima dell'ovulazione a causa della gonadotropina. Un oocita primario può causare più errori durante la segregazione cromosomica nelle due fasi della meiosi rispetto a un uovo più giovane risultando in un uovo aploide con più o meno cromosomi.

1.1.2 Ulteriore DNA presente nelle cellule

Cellule eucariote

Le cellule eucariote possono avere DNA addizionale oltre il DNA cromosomale, in particolare in:

- Mitochondri: forniscono le cellule con ATP e sono organelli racchiusi da membrana con il proprio, solitamente circolare, cromosoma singolo.
- Cloroplasti: derivano l'energia dalla luce solare nelle piante, possiedono un proprio cromosoma.

Si pensa che questi organelli derivino da un batterio ancestrale assorbito e mantenuto da un altro organismo unicellulare.

Cellule batteriche

Le cellule batteriche possiedono DNA addizionale nelle proprie cellule: piccolo DNA circolare detto plasmide. Questi tipicamente codificano poche proteine che conferiscono un vantaggio selettivo come una resistenza ad un antibiotico.

Virus

I virus sono agenti infettivi che trasportano informazioni genetiche come piccoli cromosomi a DNA o RNA. I cromosoma virale può essere lineare o circolare, a doppio o singolo filamento.

1.2 Impacchettamento del DNA cromosomale

Si nota come per potersi adattare alle dimensioni del nucleo, delle cellule o di organelli intracellulari il DNA deve essere compattato. Compattare il genoma svolge anche una funzione di protezione, rendendolo meno accessibile da agenti esterni.

1.2.1 Il nucleotide

Nei procarioti, in assenza di nucleo il DNA si organizza in un nucleotide. È composto per l'80% di DNA e per il restante 20 di proteine di compattamento e RNA. Il cromosoma è pertanto composto da un grande complesso DNA proteine detto cromatina. Il nucleotide appare come una regione che esclude cromosomi, occupa $\frac{1}{3}$ del volume della cellula ed è ancorato all'origine di replicazione nella membrana cellulare. Forma "loops" o domini di circa 40kb grazie alla proteina *HIF* (integration host factor), una piccola proteina carica positivamente per bilanciare le cariche negative sul backbone. Il DNA è successivamente superavvolto da altre proteine che piegano il DNA. L'IHF è costituito da un dimero su cui si forma il loop. Il superavvolgimento è controllato da altri fattori come fattori di

trascrizione e l'attività di DNA ed RNA polimerasi che creano due superavvolgimenti con polarità opposta ai lati della bolla.

1.2.2 DNA eucariotico

Nel nucleo degli eucarioti i cromosomi subiscono cambi visibili durante il ciclo di divisione cellulare. Nelle cellule umane diploidi il DNA deve essere compattato 300 000-400 000 volte.

1.2.3 Il ciclo cellulare e la dinamicità dei cromosomi

Durante la fase G_2 del ciclo cellulare i cromosomi replicati si trovano in uno stato poco avvolto e si forma il centromero. Nella profase compaiono le fibre del fuso e i cromosomi si condensano. Nella prometafase le fibre si attaccano ai cromosomi che continuano a condensarsi. Nella metafase i cromosomi si allineano. Nell'anafase i centromeri si dividono e i cromatidi fratelli si muovono ai poli opposti. Durante la telofase si riforma la membrana nucleare, i cromosomi si decondensano e scompaiono le fibre del fuso.

1.3 Impacchettamento del DNA cromosomale degli eucarioti

1.3.1 Istoni

Negli eucarioti gli istoni sono proteine leganti il DNA. Si trovano quattro istoni del nucleo: nascono molto presto nell'evoluzione ed essendo cruciali per la sopravvivenza sono altamente conservati. Gli istoni sono basici in quanto ricchi in lisina e arginina cariche positivamente grazie all'gruppo ammino $+NH_3$ che stabilizzano le interazioni tra il DNA e gli istoni. 146bp si arrotolano 1.75 volte intorno a un complesso istonico in maniera sinistrorsa per formare un nucleosoma. Il complesso istonico prende il nome di ottamero istonico. Il superavvolgimento negativo facilita la separazione più facile, necessaria per la replicazione e la trascrizione. L'ottamero istonico ha due di ognuno dei quattro istoni del nucleo: $H2A$, $H2B$, $H3$, $H4$. Inizialmente due dimeri $H3-H4$ si associano con il DNA e reclutano poi due dimeri $H2A-H2B$ per la formazione dell'ottamero. Nonostante tutto il DNA eucariote sia impacchettato dagli istoni i nucleosomi si formano preferenzialmente a sequenze di DNA. Il DNA è generalmente piegato dolcemente intorno agli istoni ma presenta curve più acute ???????? La scanalatura minore deve diventare più stretta durante il piegamento, cosa più favorevole in regioni ricche di AT. Gli istoni fanno 13 interazioni con gli istoni del DNA nucleosomale: i due dimeri $H3-H4$ legano il centro e le terminazioni del DNA mentre $2(H2A-H2B)$ legano 30bp su un lato del nucleosoma. Il core istonico è composto dai domini di histone-fold composti da tre α -eliche.

Code istoniche

Il nucleo di una proteina istonica è legato a una lunga coda N-terminale che si estende verso l'esterno. Sono lunghe tra i 20 e i 39 amminoacidi e non hanno strutture. Interagiscono con altri nucleosomi per aiutare un ulteriore compattamento del DNA e strutture cromatiniche di livello superiore. La coda può essere modificata chimicamente in modo da modificare la struttura della cromatina e la sua funzione promuovendo o prevenendo il reclutamento di proteine che regolano la trascrizione. $H2A$ e $H2B$ presentano anche code C-terminali che regolano la trascrizione.

Varianti istoniche

Le varianti istoniche sono altre proteine con stabilità diverse, domini specialisti che cambiano la funzione del cromosoma, sequenze diverse alle terminazioni. Hanno amminoacidi diversi che possono essere diversamente modificati. Le varianti istoniche sono depositate da complessi di rimodellamento della cromatina dipendenti da ATP e possono essere dipendenti o indipendenti dalla replicazione.

Interazioni con il DNA

Gli istoni interagiscono con il DNA attraverso interazioni elettrostatiche tra i gruppi fosfato del legame fosfodiester e gli amminoacidi basici negli istoni e attraverso legami a idrogeno tra l'atomo di ossigeno nei gruppi fosfato e gli atomi di idrogeno nei gruppi ammino degli istoni. Modifiche chimiche delle basi o code istoniche attraverso enzimi modificano le cariche locali e le interazioni.

1.3.2 Livelli di compattazione

L'impacchettamento cromatinico ha diversi livelli di compattamento.

Primo livello

Il primo livello di compattazione è la fibra di 10nm, che nasce dall'associazione con il DNA dei nucleosomi con apparenza di perline su un filo.

Secondo livello

La fibra è ulteriormente compattata da una quinta proteina istonica *H1* in una fibra di 30nm nel secondo livello di compattamento, un ordinamento regolare che avvicina i nucleosomi. *H1* si lega al DNA linker tra due nucleosomi successivi diminuendo la lunghezza di 7 volte. Anche le code istoniche sono coinvolte nella formazione di questo secondo livello.

Terzo livello

Il terzo livello di compattamento, con diametro di 300nm si forma grazie a domini di loop radiali e al legame con la matrice nucleare nelle cellule in interfase. La matrice nucleare è composta dalla lamina nucleare composta da fibre proteiche della matrice interna e da proteine che legano ad essa i cromosomi. Le proteine attaccano la base di un loop di DNA alla fibra proteica grazie a sequenze specifiche *MAR* (matrix-attachment region) e *SAR* (scaffold attachment region). Si nota come ogni cromosoma occupa nel nucleo un territorio determinato

Quarto livello

I loop radiali diventano altamente compattati e rimangono ancorati alla matrice nucleare. Mentre la cellula entra in profase la membrana nucleare si dissolve e non si trova più una matrice nucleare: la compattazione aumenta drammaticamente nel quarto livello di compattazione o condensazione. Alla fine della profase i cromosomi sono interamente eterocromatici con un diametro di 700nm. Pertanto i cromosomi in metafase subiscono poca trascrizione e unicamente nel centromero. In questo momento i cromosomi hanno accesso al fuso mitotico.

Quinto livello

Il quinto livello di compattamento avviene con la formazione dei cromosomi visibili e grazie alla condensina. La condensina è una proteina che si sposta nel nucleo durante l'inizio della fase *M*, si lega ai cromosomi e compatta i loop radiali riducendo il loro diametro. Un'altra proteina coinvolta è la coesina caricata durante la fase *S* per tenere uniti i cromatidi fratelli.

1.4 Modifiche covalenti degli istoni

1.4.1 Epigenetica

Si intende per epigenetica l'ereditarietà di fenotipi non causati da cambi nella sequenza del DNA. È un fenomeno principalmente eucariote ed è causata da cambi strutturali nella composizione dei nucleosomi (varianti istoniche), modifiche chimiche della coda o nucleo istonico che altera lo stato di compattazione della cromatina e l'attività del nucleo del nucleosoma, metilazione del DNA alla citosina e dal legame di DNA o RNA con RNA non codificanti. Queste opzioni alterano l'espressione genetica. Cambi epigenetici sono trasferiti da cellula madre e figlia durante la replicazione del DNA e un numero di sindromi e cancro sono dovuti alla mal-regolazione di attività epigenetiche. Nei batteri la trascrizione dipende principalmente dall'RNA polimerasi e la sua regolazione allo stadio di iniziazione. Metilazione di adenosina e citosina intervengono nell'espressione genica, nella replicazione e riparazione del DNA e come difesa contro attacchi virali. Le modifiche chimiche più comuni sono alle code istoniche ma anche gli amminoacidi del nucleo globulare degli istoni possono essere modificati. Le modifiche sono principalmente acetilazione, metilazione, fosforilazione, ubiquitinazione e sumoilazione: "PUMAS". Tali modifiche vanno a colpire la struttura cromatinica e il reclutamento di proteine specifiche su di essa. Le modifiche epigenetiche sono molto veloci e reversibili attraverso enzimi e sono alla base di una veloce e precisa regolazione dell'attività genica.

1.4.2 Acetilazione

La maggior parte della cromatina possiede istoni acetilati, specialmente nelle code *H3* e *H4*. È associata con una trascrizione attiva: l'eucromatina è più acetilata. L'acetilazione di code e nuclei ha effetto sulla struttura cromatina:

- Direttamente: neutralizza le cariche positive sulla lisina sulla coda istonica riducendo le interazioni tra le code e il DNA rendendo la cromatina più accessibile da proteine leganti il DNA.
- Indirettamente: la lisina acetilata agisce come un sito di riconoscimento e legame per proteine contenenti bromodomini o lettori che possono reclutare altre proteine, componenti di grandi complessi che regolano la trascrizione come *HAT*, complessi di rimodellamento della cromatina e fattori di trascrizione che agiscono come *HAT*.

L'enzima responsabile per l'aggiunta di un gruppo acetile (mono-acetilazione) al gruppo ammino + NH_3 della lisina è l'istone acetiltrasferasi *HAT*, mentre l'istone deacetilasi *HDAC* lo rimuove. L'acetilazione della lisina pertanto neutralizza direttamente la carica positiva di essa riducendo l'attrazione tra DNA PO_4^- e lisina NH_3^+ . Inoltre diventa un sito di legame per proteine con bromodominio e rimodellatrici della cromatina aprendola e attivando la trascrizione. La deacetilasi agisce come repressione della trascrizione. La (de)acetilazione in regioni promotrici ha un ruolo nell'iniziazione della trascrizione. Altre acetilazioni sono presenti lungo sequenze codificanti, con

ruolo nell'allungamento della trascrizione. Se ne trovano ancora in enhancers o in varianti istoniche che presentano trascrizione attiva. Le proteine contenenti un bromodominio possono legarsi a una o più lisine acetilate attraverso il dominio e contengono altri domini come un dominio *PHD* che si lega a lisine metilate.

1.4.3 Metilazione

La metilazione sugli istoni avviene grazie a un istone metiltrasferasi *HMT* che può aggiungere 1, 2 o 3 gruppi metile sul gruppo amminico della lisina *K* o 1 o 2 gruppi metile sul gruppo amminico dell'arginina *A*. La metilazione della coda e del core istonico ha due effetti sulla struttura cromatinica:

- Diretto: mantiene la carica locale della lisina positiva compattando il legame tra istoni e DNA.
- Indiretto: proteine contenenti cromodomini (*HP1*, *Polycomb*) riconoscono e legano a specifiche lisine metilate e reclutano proteine che causano il silenziamento trascrizionale (togliendo spazio al legame con fattori di trascrizione) o la sua attivazione.

La metilazione è associata sia con attivazione che con repressione della trascrizione in base al residuo che è metilato:

- Mono-metilazione di *K9* nella coda *H3* causa una cromatina attiva trascrizionalmente.
- Mono- o tri-metilazione di *K4* nella coda *H3* causa una cromatina attiva trascrizionalmente.
- Di- o tri-metilazione di *K9* nella coda *H3* causa una cromatina silente trascrizionalmente.

1.4.4 Fosforilazione

I fosfati sono aggiunti da chinasi e rimossi da fosfatasi. La fosforilazione aggiunge una carica negativa alla coda istonica. Fosforilazione di *S10* nella coda *H3* permette la crescita cellulare e trascrizione promuovendo l'acetilazione di *K14* sulla coda *H3*. La fosforilazione di *S10* e *S27* nella coda *H3* è correlata con la condensazione dei cromosomi durante la mitose. È importante per la replicazione e riparazione del DNA e per l'apoptosi.

1.4.5 Ubiquitinazione e sumoilazione

L'ubiquitinazione delle lisine consiste nell'aggiunta di una proteina di 76 amminoacidi catalizzata dall'ubiquitina ligasi e rimossa dalla de-ubiquitinasi. Il suo ruolo non è compreso a fondo e avviene specialmente nelle code C-terminali di *H2A* e *H2B*. Regola la trascrizione reclutando rimodellatori e risposte al danno del DNA. Una mono-ubiquitinazione di *H2A* causa repressione trascrizionale mentre se avviene a *H2B* causa un'attivazione indiretta in quanto richiesta per la mono-metilazione di *H3K4* e di *H3K79*. La sumoilazione della lisina è una modifica simile all'ubiquitinazione e gioca un ruolo nella regolazione di trascrizione e riparazione di DNA.

1.4.6 Codice istonico

Le grandi possibili modifiche in aggiunta con le loro interazioni porta alla definizione di un codice istonico in cui modifiche uniche definiscono certi stati di cromatina e di espressione genica.

1.5 Complessi rimodellatori dei nucleosomi

La cromatina compattata rappresenta una barriera per le proteine che devono accedere al DNA e pertanto inibisce processi come trascrizione. La composizione del nucleosoma, la compattezza del suo legame con il DNA e la sua locazione possono essere fisicamente cambiati da complessi di rimodellamento dei nucleosomi dipendenti da ATP. Questi complessi possono introdurre loop nel DNA avvolto intorno a un nucleo istonico, far scivolare il DNA lungo l'ottamero istonico o rimuovere l'intero ottamero o 1-2 proteine istoniche e trasferirle da qualche altra parte. Possono attivare o reprimere la trascrizione ma non sono usati per la replicazione.

1.5.1 Ruoli dei rimodellatori della cromatina

I complessi di rimodellamento della cromatina hanno diversi ruoli nello stato cromatinico. Possono intervenire dopo la deposizione degli istoni durante la maturazione dei nucleosomi portando a una loro spaziatura regolare. Possono inoltre alterare lo stato cromatinico riposizionando i nucleosomi, espellendoli completamente o solo alcune loro subunità. Possono inoltre compiere installazioni o rimozioni di varianti istoniche.

1.5.2 Sottofamiglie

Esistono diverse classi di rimodellatori dei nucleosomi, ma tutte contengono dei domini chiave:

- Dominio motore ATPasi come *Dexx* e *HELICc*.
- Bromodominio o cromodominio.
- Dominio legante actina *HSA*.
- Dominio per il legame alla coda istonica *SANT* e *SLIDE*.

Switch/sucrose non-fermentable

Il complesso *SWI/SNF* facilita l'accesso alla cromatina: fa scivolare ed espelle i nucleosomi per l'attivazione o repressione genica.

Imitation switch

Il complesso *ISWI* assembla e spazia i nucleosomi principalmente per la repressione della trascrizione.

Cromodominio elicasi legante il DNA

Il complesso *CDH* è usato per l'assemblaggio dei nucleosomi e la loro spaziatura, per l'accesso ai geni esponendo i promotori e l'editing attraverso l'incorporazione di *H3.3*. Aiuta i repressori a legarsi alla cromatina e reprimere i geni attraverso *HDAC* associate.

Richiedenti inositolo

Il complesso *INO80* interagisce con *HAT* per attivare la trascrizione. Interviene anche nell'assemblaggio e spaziatura dei nucleosomi oltre a sostituire *H2A* con *H2A.Z* per la riparazione del DNA.

1.6 Variazione nella struttura cromatinica

I cromosomi subiscono varie fasi di compattazione diversa durante il ciclo cellulare. Durante l'interfase, quando i cromosomi sono relativamente poco condensati, i geni sono trascritti e il genoma è replicato si trova un gran numero di compattazione lungo il cromosoma. Della trascrizione può avvenire nelle regioni eterocromatiche, ma la traslocazione di un gene da una regione eucromatica a una eterocromatica può prevenire attivamente la sua trascrizione. Il livello di compattamento della cromatina non è uniforme e l'epigenetica rappresenta il suo ultimo livello di regolazione.

1.6.1 Eucromatina

Si dicono eucromatiniche le regioni dove le fibre di 30nm formano domini radical loop formando cromatina a 300nm. Questa zona è trascrizionalmente attiva.

1.6.2 Eterocromatina

Nell'eterocromatina i domini radical loop sono ulteriormente compattati attraverso metilazione della coda istonica a formare una cromatina a 700nm. L'eterocromatina si divide in costitutiva, o regioni sempre eterocromatiche permanentemente disattivate rispetto alla trascrizione o silenti e facoltativa, o regioni di cromatina che cambiano stato tra eucromatina ed eterocromatina. Alcune zone dei cromosomi sono altamente eterocromatiche:

- Telomeri: regioni di DNA alla terminazione dei cromosomi.
- Peri-centromeri.
- Regioni con sequenze di DNA altamente ripetute come l'rDNA nei nucleoli.

1.6.3 Effetti della cromatina

La cromatina ha effetto su trascrizione, replicazione, ricombinazione e trasmissione dei cromosomi. Riarrangiamenti che spostano un'origine di replicazione nell'eterocromatina causano una replicazione tardiva, arrivando fino a ritardare la divisione cellulare. La ricombinazione coinvolge rotture e riunioni di DNA di diverse molecole. Le regioni eterocromatiche ne subiscono di meno, proteggendo la regione contro tale modifica, cosa che avviene come nei geni di ripetizione di DNA ribosomiale. I cromosomi devono essere completamente compattati affinché avvenga la trasmissione e segregazione dei cromosomi.

1.6.4 Nucleolo

Il nucleolo è la parte del nucleo che contiene i geni di rDNA. Gli esseri umani possiedono cinque cluster di rDNA vicino la fine di cinque cromosomi. Si dice regione organizzatrice dei nucleoli i trascritti di rRNA prodotti dalle ripetizioni dall'rDNA. rDNA codifica per l'rRNA ribosomiale e molte cellule possiedono migliaia di ripetizioni di rDNA per riuscire a soddisfare la richiesta di rRNA e produzione di ribosomi. Il nucleolo non è separato da una membrana: sono le proteine e le RNA ad esso specifiche che gli conferiscono diversi pattern di colorazione. Un sottoinsieme di ripetizioni di rDNA sono silenti trascrizionalmente ed eterocromatiche in modo da aumentare la stabilità delle regioni ripetute.

1.7 Metilazione del DNA

Il DNA può essere modificato chimicamente attraverso la metilazione, che avviene in batteri ed eucarioti. I gruppi metile possono essere aggiunti a residui di citosina per creare la *5-metil citosina* attraverso DNA metiltransferasi o DNA metilasi. La modifica è reversibile grazie alla DNA demetilasi. La metilazione è rischiosa in quanto può alterare il DNA permanentemente. Le citosine metilate infatti possono subire una spontanea deamminazione idrolitica che cambia la citosina in timina con cambio mutagenico.

1.7.1 DNA metilasi

Le DNA metilasi utilizzano un base flipping per accedere alla citosina: una citosina è fatta uscire dalla doppia elica: un amminoacido dell'enzima è inserito temporaneamente al suo posto. La citosina viene poi metilata e reinserita nel DNA.

1.7.2 Effetti della metilazione

Nei procarioti

Nei procarioti la metilazione del DNA distingue il DNA appena sintetizzato nel processo di riparazione: appena dopo la replicazione solo il filamento genitore è metilato: questa regione si dice emi-metilata. Quando gli enzimi di riparazione del mismatch ne trovano uno leggono lo stato metilato per identificare correttamente il filamento parentale e riparare quello appena sintetizzato. La metilazione permette anche ai batteri di distinguere il DNA genomico da quello virale invasivo: enzimi di restrizione tagliano il DNA del fago a siti di riconoscimento specifici e durante il taglio il batterio protegge il proprio DNA metilando i siti di restrizione.

Negli eucarioti

La metilazione del DNA negli eucarioti silenzia la trascrizione. È pertanto un'altra forma di silenziamento epigenetico. Non cambia la carica della base e l'effetto repressivo è indiretto in quanto comporta il reclutamento di proteine lettrici che riconoscono e legano la base metilata. La metilazione avviene tipicamente a siti *CpG* o *CpXpG*, dove *p* è il legame fosfodiesterico e *X* una base qualsiasi. Circa il 60% delle *CpG* umane sono metilate. La metilazione può anche essere ereditata. Alcuni complessi si legano specificatamente a DNA metilato come enzimi di modifica istonica e complessi di rimodellamento della cromatina. Alcune proteine leganti istoni possono reclutare DNA metiltransferasi.

Isole *CpG* Le sequenze *CpG* non sono distribuite uniformemente nel genoma ma si trovano in lunghezze di 1-2kb dove il 60% del contenuto di DNA forma queste isole *CpG*. Sono studiate principalmente per la disattivazione del cromosoma *X*, si trovano in tutti i geni housekeeping, principalmente nella zona 5' nel promotore. Sono principalmente hypo-metilate, protette dalla metilazione e si correla con un'alta attività di trascrizione. *CpG* sono riconosciute da proteine *MBD* (metil-CpG-binding domain) con un dominio di legame di DNA e di un dominio repressore della trascrizione che possono reclutare complessi di rimodellazione della cromatina che disattivano la trascrizione. La metilazione può anche proibire il legame con fattori di trascrizione alle proprie sequenze di riconoscimento del DNA in un processo di mascheramento di *C*. La demetilazione avviene quando un gene deve essere trascritto. La metilazione di *CpG* è ereditata grazie all'enzima DNA metiltransferasi *DNMT1* che riconosce il sito emimetilato e lo rende completamente metilato.

1.7.3 La disattivazione del cromosoma X è un esempio di silenziamento epigenetico della metilazione di cromatina nei mammiferi

Un cromosoma X in ogni cellula è disattivato nelle femmine in modo che abbiano la stessa quantità di prodotto di gene X come nei maschi che ne possiedono uno solo. Il DNA è altamente metilato, $H2A$ è sostituito con *MacroH2A-Z*, gli istoni sono modificati come in eterocromatina ed avviene una regolazione basata su long non-coding RNA. La disattivazione del cromosoma X è casuale ed avviene alla gastrulazione nell'embrione, ognuna delle cellule possono scegliere individualmente quale dei due cromosomi X disattivare e la scelta viene ereditata. Uno dei due cromosomi si presenterà pertanto più denso, compatto e su un lato del nucleo. Circa il 15% dei geni legati a X non vengono disattivati completamente e la loro attività genica varia tra i cromosomi disattivati. La maggior parte di questi si trova nelle regioni pseudoautosomiali *PAR*, dove X e Y si accoppiano durante la meiosi.

Non corretta disattivazione di X durante la gastrulazione

Gatti calico La colorazione rossa del pelo dei gatti è dovuta a un gene nel cromosoma X , in cui l'allele rosso sintetizza un enzima che crea il pigmento arancio, mentre un altro non lo esprime e causa una colorazione nera. Nel caso in cui un gene X non sia disattivato e i maschi presentano XXY presentano una colorazione arancio e nera, oltre ad essere sterili.

1.7.4 Imprinting genetico

Una parte dell'attività genetica è controllata dall'imprinting genetico, che regola l'espressione di geni materni e paterni nell'embrione, casualmente in alcune cellule è silenziata la copia materna, in altre quella paterna. Se una delle copie di un gene è silenziata e l'altra è stata deleta non si trova espressione genica.

Disordini fisici e neurologici dovuti alla misregolazione dei geni soggetti a imprinting attraverso metilazione di citosina

Sindrome di Rett Questa sindrome è dovuta a una mutazione disattivante in un allele del *MECP2*. Avviene quando *MECP2* in un allele non è espresso a causa di metilazione. Uno di questi geni deve essere sempre espresso per la vitalità.

Sindrome di Prader-Willy e di Angelman In queste due sindromi sono colpiti gli stessi alleli del cromosoma 15. *PWS* avviene quando una regione paterna di 7 geni è eliminata. *AS* avviene quando è eliminata la regione materna. Le sindromi si manifestano quando l'altro allele parentale è espresso sub-ottimamente a causa dell'imprinting. Un insieme allelico parentale deve essere intatto per la sopravvivenza dell'embrione.

1.8 La separazione dei domini cromatinici da elementi barriera

1.8.1 Variegazione da effetto di posizione

Un effetto epigenetico è la variegazione da effetto di posizione. Un suo esempio è il colore dell'occhio di *Drosophila* in cui si presentano rossi grazie all'espressione del gene *white*⁺. In alcuni casi gli occhi

possono presentare sfaccettature bianche se il gene viene convertito in una regione eterocromatica in qualche cellula in cui risulta silenziato.

1.8.2 Elementi di barriera

Le cellule possiedono elementi di barriera che separano eu ed eterocromatina. Questi elementi possono prevenire la diffusione dell'eterocromatina. In *S. pombe* due elementi di barriera affiancano una regione di eterocromatina silente intorno al centromero. Gli *H3* negli elementi di barriera sono altamente metilati a *K9* silenziando la regione, mentre quelli fuori la barriera sono altamente metilati a *K4* attivando la regione. La rimozione di questi elementi permette la diffusione di metilazione *K9* e della zona silenziata. L'eterocromatina può infatti diffondersi attraverso modifiche di istoni successive come deacetilazione di *H3* la sua metilazione a *K9* e il legame della proteina di silenziamento *Swi6*. Gli elementi di barriera agiscono come barriere fisiche e possono essere sequenze specifiche a cui si legano proteine regolatrici delle modifiche istoniche o grandi loop di cromatina. Elementi di sequenze di barriera possono ancorare gli anelli nella lamina nucleare: l'eterocromatina si trova nelle regioni periferiche del nucleo in quanto *SAR/MAR* affiancano spesso elementi di sequenza di barriera.

1.9 Elementi richiesti per la funzione dei cromosomi

1.9.1 Origine di replicazione

Le origini di replicazione sono regioni del DNA con sequenze specifiche richieste per la replicazione in batteri ed eucarioti. Le *Ori* sono dove il dsDNA è svolto e separato per prepararsi all'attacco delle proteine di replicazione. La replicazione è bidirezionale:

- Nei batteri si trova un *Ori* per cromosoma e si indica con *ter* gli elementi di terminazione della replicazione.
- Negli eucarioti si trovano diverse *Ori* lungo il cromosoma in quanto si ha più cromosoma da replicare

1.9.2 Centromeri

I centromeri si trovano in tutti i cromosomi eucarioti. Sono sequenze che si trovano tipicamente al centro del cromosoma e sono necessarie per la segregazione dei cromosomi durante la divisione cellulare. La maggior parte delle specie ne possiedono 1 per cromosoma. Si trova in una regione eterocromatica e dopo la replicazione del DNA 2 cromatidi fratelli si formano e sono uniti dai complessi di anelli di coesina. Il centromero appare come una costrizione dovuta all'arricchimento locale di coesina che si trova in una quantità minore lungo l'intero paio di cromatidi. Il centromero ha dimensioni variabili a seconda della specie e si distingue in:

- Centromero puntiforme con sequenze definite di poche centinaia di basi.
- Centromero regionale con centinaia di kilobasi.

Alcuni organismi possiedono molti centromeri lungo il cromosoma e sono detti olocentrici. I microtubuli si attaccano su tutta la lunghezza del cromosoma. Durante la mitosi si possono segregare frammenti di cromosomi.

Il cinetocore

Il centromero recluta più di 100 proteine che formano il cinetocore che attacca i cromatidi fratelli ai microtubuli che si estendono da poli opposti del fuso, permettendo ad esso di separare i cromatidi attraverso la depolimerizzazione dei microtubuli. Questo meccanismo di segregazione è altamente conservato.

Esempi di centromeri

S. cerevisiae Il centromero è lungo 125bp e possiede tre regioni *CDEI*, *CDEII* e *CDEIII*, la prima e la terza possiedono sequenze altamente conservate e singole mutazioni possono rompere la funzione del centromero. *CDEII* invece è una regione ricca di *AT* e la sequenza esatta non è fondamentale.

S. pombe Ogni cromosoma di *S. pombe* presenta un cromosoma con una sequenza di centromero leggermente diversa con un nucleo unico di 5-6kb con lunghe sequenze di ripetizioni inverse che lo affiancano.

Esseri umani I centromeri sono lunghi 1Mb e sono fatte di sequenze ripetute dette ripetizioni α -satellite, lunghe 171bp ordinate in ripetizioni di ordine più alto. I nucleosomi centromerici possiedono varianti istoniche di *H3* *CENP-A* particolarmente nelle regioni ricche di *AT* che potrebbe riconoscere gli *i-motivi* e diadi. Il centromero marcato da *CENP-A* è dove il cinetocore si assembla. Una sovraespressione di *CENP-A* causa un legame del cinetocore con tutto il cromosoma e una sua rottura durante la segregazione.

1.9.3 Telomeri

I telomeri sono regioni alle terminazioni dei cromosomi lineari e funzionano come cappucci protettivi. Negli esseri umani sono formati da sequenze ripetute centinaia di migliaia di volte di *TTAGGGG*, marcano la terminazione del cromosoma definendolo e impedendo la fusione di cromosomi alle loro terminazioni. Infatti più un cromosoma è lungo più è propenso a subire rotture. Il DNA dei telomeri consiste di un filamento ricco di *G* e uno ricco di *C*. La lunghezza totale delle ripetizioni varia tra i 50 000 e i 30 000bp in base alla specie. La sequenza ricca di *G* si estende 5'-3' verso la terminazione del cromosoma dove termina in una regione corta a filamento singolo. Negli organismi con telomeri lunghi questa regione può essere processata in un rolled back T-loop formato dall'invasione e accoppiamento di basi del filamento singolo con la sequenza a doppio filamento a monte. Le ripetizioni sono un sito di legame per proteine che le marcano come terminazioni naturali distinguendoli dalle rotture del DNA. La DNA polimerasi non può copiare la terminazione di una molecola di DNA e pertanto interviene la telomerasi per mantenere le terminazioni dei cromosomi.

Telomerasi

Le proteine *TRF1* e *TRF2* (*TTAGGGG* repeat binding factor), *TIN2* e *RAP1* si legano ai telomeri e proteggono le loro terminazioni. La terminazione di ogni telomero forma un T-loop composto da una ripetizione *TTAGGGG* 3' a filamento singolo che lega una sequenza complementare in una sequenza a monte denaturata detta D-loop (displacement loop) e viene stabilizzata da copie multiple di *POT1* che vi si lega. La telomerasi è una speciale DNA polimerasi che possiede una proteina e una componente a RNA: forma un *RNP*. L'RNA della telomerasi fornisce un corto stampo che specifica la sequenza della ripetizione telomerica che deve essere aggiunta. La telomerasi pertanto sintetizza il DNA telomerico usando l'RNA come stampo. La lunghezza dei telomeri è mantenuta

1.9. ELEMENTI RICHIESTI PER LA FUNZIONE DEI CROMOSOMI

nelle cellule staminali e germinali, mentre nei tessuti maturi si trova una telomerasi insufficiente e avviene un accorciamento dei telomeri. Che limita il numero di divisioni cellulari che la cellula può avere. Una sovraattivazione della telomerasi è implicata in molti cancro e permette alle cellule di continuare a crescere e a dividersi.

Capitolo 2

Replicazione

2.1 Replicazione del DNA semi-conservativa

Durante la divisione cellulare l'informazione genetica deve essere copiata e distribuita equamente tra le cellule figlie. Dopo la scoperta della struttura a doppia elica del DNA si ragionò come i due filamenti complementari sono copiati e replicati.

2.1.1 Modelli di replicazione

- Replicazione conservativa: il DNA rimane intatto come un doppio filamento e agisce come stampo.
- Replicazione semi-conservativa: un filamento agisce come stampo per sintetizzare un nuovo filamento complementare.
- Replicazione dispersiva: il doppio filamento si rompe nella sua lunghezza e frammenti sovrapposti servono come stampi per la sintesi.

Determinazione del modello semi-conservativo

Per determinare il modello di replicazione Meselson e Stahl idearono un esperimento che utilizzava un gradiente di densità e ultra-centrifugazione.

Tecnica del gradiente di densità Gli scienziati presero un tubo di plastica in cui era presente un gradiente del sale cloruro di cesio. In questo modo aggiungendo componenti di diversa densità in cima al gradiente e centrifugandoli ad alta velocità la forza gravitazionale trasporta le componenti nel gradiente in modo che si fermino quando la loro densità è uguale alla densità locale della soluzione di *CsCl*. Le componenti a bassa densità sono posizionate più in alto nel gradiente, mentre quelle a densità più alta in basso. Quando le componenti si sono mosse attraverso il gradiente durante la centrifugazione si prendono campioni dal basso all'alto attraverso frazionamento.

Rendere il DNA più pesante Per rendere il DNA di nuova sintesi più pesante rispetto a quello originale viene utilizzato un isotopo dell'azoto ^{15}N in quanto è la massa dell'elemento più frequente e può essere sintetizzato in forma radioattiva.

L'esperimento Gli scienziati fecero crescere una coltura di *E. coli* per 4 divisioni cellulari in un medio minimale contenente glucosio e con $^{15}\text{NH}_4\text{Cl}$ come l'unica fonte di azoto. Il DNA alla fine pertanto conterrà ^{15}N nelle basi nucleotidiche. Prendendo un campione della coltura della quarta divisione cellulare e isolandolo. Successivamente si isola il resto dei batteri attraverso centrifugazione, li si lava e risospende nel medio minimale con $^{14}\text{NH}_4\text{Cl}$ come unica fonte di azoto. Si lascia crescere e dividere la coltura così ottenuta prendendo campioni ogni divisione. Si isola il DNA dai vari campioni e lo si carica su un tubo a gradiente di CsCl separato. Successivamente si ultracentrifugano tutti i tubi in parallelo, si fraziona i campioni e si fa correre il DNA su un gel di agarosio e li si trasferisce su membrana di nitrocellulosa. Infine si espone la membrana a un foto film.

Conclusioni Si nota come in base al modello si osserverebbero comportamenti diversi:

- Modello conservativo: il numero di batteri con ^{15}N rimarrebbe costante e aumenterebbe quella con ^{14}N , presentando pertanto due bande, una per l'isotopo e una per ^{14}N .
- Modello dispersivo: i batteri presenterebbero tutti del DNA ibrido contenente sia ^{15}N che ^{14}N , presentando pertanto una banda unica all'ibrido.
- Modello semi-conservativo: si troverebbe nella popolazione un numero di molecole contenenti uno strand con ^{15}N e l'altro ^{14}N , mentre il resto tutto a ^{14}N , pertanto si noterebbero due bande, una per ^{14}N e una per l'ibrido.

Si osserva che avviene il terzo caso, determinando che la replicazione è semi-conservativa.

2.2 Il modello dei repliconi

Il modello dei repliconi è stato proposto nel 1963. Si indica con replicone la parte del DNA che sta venendo replicata. La replicazione inizia a una particolare sequenza di origine o replicatore. Una proteina iniziatrice si lega al replicatore per iniziare il processo di replicazione.

2.2.1 Scoperta del modello

La scoperta del modello dei repliconi avviene grazie a Cairns nel 1963 attraverso un'analisi autoradiografica del genoma in replicazione di *E. coli*. Si cresce la cellula in un medium contenente glucosio e azoto. Si aggiunge ad essa $[^3\text{H}]\text{-timidina}$ e si fanno avvenire due replicazioni del DNA in modo che questa si incorpori due volte nel filamento di nuova sintesi. Si lisa la cellula e la si espone a un foto film per due mesi.

Osservazioni

Si nota come dopo una replicazione un filamento non è radioattivo mentre l'altro lo è. All'inizio della seconda replicazione si forma una sezione con entrambi i filamenti radioattivi, permettendo di visualizzare come il DNA si replica in maniera semi-conservativa nella cellula. Nel replicone o *Ori* il DNA si apre formando bolle tra due forcelle di replicazione. La bolla si estende in maniera bidirezionale. Lo si nota osservando la radioattività ai due estremi della bolla di replicazione.

2.2.2 Origini di replicazione

Origine di replicazione singola

In caso di una singola *Ori* in DNA circolare questo si comincia a svolgere in tale sequenza producendo una bolla di replicazione con una forcella ad ogni terminazione. Le forcelle procedono lungo il cerchio producendo il modello Θ . Successivamente interviene la topoisomerasi II *girasi* che separa le due molecole di nuova formazione.

Origini di replicazioni multiple

In caso di multiple *Ori* in DNA lineare si formano varie bolle di replicazione con forcelle ad ogni estremità. Le bolle mano a mano che ne incontrano altre si fondono tra di loro.

2.3 Identificazione delle origini di replicazione

Nei genomi di batteri, batteriofagi, virus e plasmidi si trova un *Ori* per molecola di DNA, nei primi in quanto il loro DNA si replica in maniera indipendente da quella dell'ospite. L'origine di replicazione in *E. coli* o *OriC* è stata trovata attraverso un esperimento.

2.3.1 Esperimento

Si prende una coltura di *E. coli* e la si trasforma con un plasmide contenente del DNA di *E. coli* ottenuto attraverso enzimi di restrizione e un gene che codifica la resistenza all'ampicillina. In questo modo ogni colonia che cresce in presenza dell'antibiotico contiene un *Ori* nel plasmide. Si continua a ridurre la lunghezza del frammento fino a che non si verifica più la resistenza. In questo modo si riesce a determinare la sequenza minima e specifica dell'*Ori*. In *E. coli* è lunga 245bp e la parte che si apre formata da $3 \times 13bp$ è ricca in *A* e *T* in quanto le basi formando solo due legami a idrogeno sono più facili da aprire. La sequenza contiene inoltre $5 \times 9bp$ siti di legame per 5 proteine iniziatrici *DnaA*.

2.3.2 Origini di replicazione negli eucarioti

Si nota come i lunghi cromosomi lineari degli eucarioti possiedono multiple origine di replicazione in modo da replicare il DNA in un tempo ragionevole. Attraverso autoradiografia si identificano diverse origine di replicazione attraverso le bolle con diversa dimensione in base al tempo di formazione: se precoce o tardiva. Le forcelle di replicazione si muovono comunque in maniera bidirezionale e si uniscono tra di loro quando si incontrano. La lunghezza di repliconi individuali è di 100bp in lievito e mosche e tra i 75 000 e 175 000 in cellule animali e umane. Il tasso di replicazione negli eucarioti è di $2000 \frac{bp}{min}$, molto più lento rispetto ai batteri. Si nota come dalla velocità di replicazione il genoma di un mammifero potrebbe essere replicato in un'ora. Nonostante questo la fase *S* dura più di 6 ore in una cellula somatica. Questo avviene in quanto non più del 15% dei repliconi sono attivi in un dato momento. Ci sono eccezioni come le divisioni degli embrioni di *Drosophila*, con fase *S* molto più breve.

Identificazione degli *ARS* nel lievito *S. cerevisiae*

Si intende con *ARS* la sequenza replicante autonomamente o *Ori*. L'identificazione avviene in maniera simile a quella dell'*Ori* di *E. coli*: frammenti di DNA ottenuti attraverso enzimi di restrizione

2.4. PANORAMICA DELLA REPLICAZIONE DEL DNA

vengono introdotti un un plasmide e in cellule del lievito incapaci di crescere in coltura priva di istidina. Le colture in grado di crescervi contenevano un'origine di replicazione. Si nota come si trova un *ARS* ogni 135 000bs.

Ruolo della struttura cromosomica nella replicazione

Negli eucarioti non tutte le *Ori* sono utilizzate durante la replicazione: la loro attivazione è regolata nella fase *S* da proteine che regolano il ciclo cellulare, l'ambiente locale di cromatina (effetto di posizionamento). *ARS1* si trova vicino al centromero e *ARS501* vicino a un telomero. Il cambio di posizione cambia il momento di firing dell'*ARS*. Nel lievito le *ARS* vicine al centromero si attivano precocemente. Fattori di replicazione, modificatori della cromatina e rimozioni degli istoni leggono il codice istonico aprendo e chiudendo la cromatina determinando domini di replicazione precoce e altri di replicazione tardiva.

Mappatura fisica di *Ori* attraverso elettroforesi su gel d'agarosio

Questa tecnica, detta anche ibridizzazione del Southern blot traccia le aperture e i movimenti di un *Ori* in un pezzo di DNA durante la replicazione. Per farlo si isola il DNA genomico da una cultura asincrona e lo si taglia con un enzima di restrizione specifico. Si traccia il DNA con una sonda con un Southern blot su gel 2D. Ogni cellula rappresenta uno stato intermedio di attività di replicazione locale. Sul medium a 2D con agarosio con presente *EtBr* più denso in modo che la forma del frammento influisca sulla sua posizione. Il DNA osservato può assumere diverse forme in base allo stato replicativo:

- Y semplice con un grafico ad arco Y che rappresenta un replicatore passivo.
- Y doppia con un grafico ad arco a doppia Y che rappresenta una terminazione.
- Bolla simmetrica con un arco a bolla che rappresenta un'iniziazione.
- bolla asimmetrica con un arco a bolla e transizione ad arco ad Y che rappresenta un'iniziazione con un'origine non centrata.

2.4 Panoramica della replicazione del DNA

Affinchè la cellula si divida deve avvenire la replicazione del DNA. Si intende per replicazione la completa e fedele copia del DNA nei cromosomi della cellula. La replicazione è semi-conservativa: ogni filamento della doppia elica parentale agisce come stampo per la sintesi di un nuovo filamento per la cellula figlia. In modo da copiare lo stampo la base dello stampo deve essere identificata e deve essere aggiunta la base complementare. Questo garantisce a meno di mutazione che ogni doppia elica figlia sia identica a quella parentale e che ogni cellula figlia riceva molecole di DNA identiche.

2.4.1 Le fasi della replicazione del DNA

Iniziazione

Durante l'iniziazione viene riconosciuta l'origine di replicazione da una proteina iniziatrice che apre la doppia elica localmente e recluta elicasi. Le DNA elicasi continuano a svolgere l'elica per esporre DNA a singolo filamento che è circondato da proteine leganti *ssDNA*. L'iniziazione è controllata in modo che avvenga una sola volta per ogni ciclo cellulare. La sintesi del DNA necessita di un primer

in quanto può aggiungere nucleotidi a una terminazione $3' - OH$ esistente. Il primer è un piccolo filamento di RNA sintetizzato da una DNA primasi.

Allungamento

Dopo la sintesi del DNA primer la pinza scorrevole simile ad un anello è reclutata all'ibrido a doppio filamento ssDNA RNA primer. La DNA polimerasi si lega al DNA attraverso la pinza e il macchinario di replicazione o replisoma si muove lungo il DNA copiando i filamenti. Ogni base nel DNA parentale è letta dalla DNA polimerasi che aggiunge basi complementari al filamento in crescita in una direzione $5' - 3'$.

Terminazione

La terminazione avviene quando due forcelle diverse si incontrano o quando questa raggiunge la terminazione del cromosoma lineare. Il complesso di replicazione è disassemblato, i primer a RNA sono rimossi e sostituiti con DNA e la DNA ligasi connette le sequenze di DNA di nuova sintesi.

2.5 Iniziazione

Le origini di replicazione sono i siti in cui il DNA è inizialmente svolto. Le proteine iniziatrici si legano alle origini a siti di legame degli iniziatori permettendo il legame con l'elicasi e continuando a svolgere il DNA. Alcuni organismi hanno specifiche sequenze come origine ma è l'abilità di legare la proteina iniziatrice che definisce un'origine. Le proteine iniziatrici sono proteine leganti $ATP AAA^+$. In *E. coli* si chiama *DnaA*. Negli eucarioti l'iniziatore è il complesso di riconoscimento dell'origine *ORC*, in *S. cerevisiae* ha 6 subunità e si chiama *Orc1-6*. L'*ATP* regola il legame dell'iniziatore. Il legame di *ATP* con *Orc1* è richiesto per il legame di *ORC* con il DNA. Spesso le origini di replicazione possiedono una sequenza di DNA definita con DNA unwinding element, regioni ricche di *AT* che facilitano lo svolgimento in quanto possiedono solo 2 legami a idrogeno. In queste regioni l'iniziatore separa i due filamenti quando si lega al DNA e recluta altre proteine.

2.5.1 Svolgimento dell'*Ori* nei procarioti - *E. coli*

In *E. coli* l'origine di replicazione *OriC* è una sequenza di 245bp con 5 *DnaA* box di 9bp, 3 con alta affinità e 2 con bassa che legano 15 molecole di *DnaA*. Nelle 3 ad alta affinità il *DnaA* è sempre legato ad essi, mentre in quelle a bassa affinità si lega *DnaA-ATP* solo quando la replicazione deve iniziare. Tutte le proteine *DnaA* legano *ATP* e si multimerizzano in un filamento a spirale. Il filamento distorce il DNA producendo un superavvolgimento positivo locale svolgendo a valle la regione ricca di *AT* di $3 \times 13bp$ che subisce invece un superavvolgimento negativo. Successivamente *DnaA-ATP* e 6 molecole di *DnaC-ATP* caricano l'anello omoesamerico formato dall'elicasi *DnaB* sui singoli filamenti dell'origine. *DnaC* successivamente lascia l'*OriC* a seguito dell'idrolisi dell'*ATP*. *DnaB* recluta la DNA primasi *DnaG* che sintetizzerà il RNA primer. La pinza scorrevole si lega alla sequenza ibrida ssDNA-RNA primer.

Regolazione dello svolgimento di *OriC*

Inattivazione regolata di *DnaA* (*RIDA*) La replicazione del DNA è un punto di non ritorno: si devono prevenire rireplicazioni alla stessa origine e deve essere integrata con le altre attività di divisione cellulare. In *E. coli* il discriminante è la presenza di *DnaA-ATP* contro *DnaA-ADP*: solo

il primo può multimerizzarsi e svolgere il DNA. Dopo l'iniziazione l'*ATPasi* AAA^+ *HDA* lega la pinza e stimola l'idrolisi in *DnaA-ADP* che si dissocia dall'*OriC* e non può riattivare la replicazione. *RIDA* è il principale meccanismo regolatorio che previene la ri-replicazione nei procarioti.

Metilazione di *OriC* L'iniziazione può essere prevenuta attraverso metilazione del DNA: la DNA adenine metilasi *DAM metilasi* metila i residui *A* nella sequenza *GATC* lungo il genoma di *E. coli*. 11 siti *GATC* in *OriC* sovrappongono i siti di legame di *DnaA*. Dopo la replicazione solo un filamento di DNA è metilato (emimetilazione). La proteina *SeqA* si lega ai siti *GATC* emimetilati e blocca il legame della *Dam metilasi*. In questo modo previene la metilazione di entrambi i filamenti e il legame di *DnaA* con le sue box. Il blocco è temporaneo: l'origine è completamente metilata sul nuovo filamento dopo 10 minuti causando la dissociazione di *SeqA* dalla doppio filamento. I cromosomi completamente metilati si segregano nelle cellule figlie e sono in grado di legare *DnaA*.

Sequestro di *DnaA* a *datA* Il *DnaA* può legarsi alla sequenza di DNA *datA* che si trova vicino all'*OriC*. La regione può legare 370 molecole di *DnaA*. Dopo la replicazione *DnaA* viene sequestrato alla regione *datA* e non è disponibile per il legame in *OriC*. All'inizio di un nuovo ciclo di replicazione cambi locali nella sequenza *datA* causano una dissociazione di *DnaA* che può legare *OriC*. Il *DnaA* è attivato in *DnaA-ATP*.

2.5.2 Svolgimento dell'*Ori* negli eucarioti

Le origini di *S. cerevisiae* sono simili a quelle di *E. coli*. Sono lunghe tra 100 e i 200bp, siti *A* e *B1* a cui si lega *ORC*. I siti *B2* e *B3* sono adiacenti ad esse e ricchi di *AT*. *ORC* si lega intorno ai siti *A* e *B1*. La struttura cromatinica è importante per le origini negli organismi multicellulari: la replicazione inizia in regioni cromosomiche specifiche, ma le sequenze di DNA non sono conservate. In *Drosophila* *ORC* si lega a sequenze *Ori* come le code istoniche sono iperacetilate, pertanto *HAT* potrebbero essere coinvolti nella regolazione delle origini di replicazione.

Regolazione dell'attivazione di *Ori* negli eucarioti

Gli eucarioti possiedono multiple origini di replicazione. Ognuna di esse deve attivarsi una volta per ciclo cellulare. Sono selezionate in G_1 e attivate in *S*. Le origini non possono essere riutilizzate fino alla rifezione nella fase G_1 successiva.

Selezione Durante la fase G_1 *ORC* si lega a un'origine. *ORC* è costituito da 6 subunità: *Orc1-6*. 5 di esse sono AAA^+ *ATPasi*. *ORC* sottostà alla formazione del complesso *pre-RC*, formato da *ORC*, *Cdt1*, *Cdc6* e l'elicasi *MCM2-7*.

Attivazione Durante la fase *esse* nel *pre-RC* le elicasi *Mcm2-7* vengono fosforilate dalla chinasi *Dbf4 DDK*. Sono reclutate *Sld2* e *Sld3* per formare il complesso *SDS* al *pre-RC* e vengono fosforilate dalla chinasi dipendente dalla fase *S* o *S-Cdk*. Entrambe le fosforilazioni risultano in una completa attivazione del *pre-RC*. *Cdt1* e *Cdc6* lasciano il complesso. Avviene il reclutamento dei fattori di iniziazione *Cdc45* e dei complessi *GIN5* e *SDS* che permette l'apertura e svolgimento dell'origine. La replicazione deve essere completa prima che avvenga la segregazione. Se avviene uno stallo delle forcelle viene attivata la risposta al danno del DNA e l'entrata in mitosi è bloccata fino alla correzione degli errori.

Rif1 (Rap1 interacting factor 1) Questa proteina regola positivamente l'attivamento delle origini precoci e negativamente quello delle tardive. Previene il reclutamento del fattore di iniziazione *Cdc45* al *pre-RC*. È conservata nelle cellule umane ed è il regolatore chiave del programma di replicazione del DNA, ovvero dell'ordine temporale di attivazione delle *Ori*.

Regolazione dell'iniziazione della replicazione negli eucarioti multicellulari

Essendoci molte origini lungo un cromosoma lineare l'ordine di attivazione dipende da:

- *Rif1*.
- Lo stato di acetilazione della cromatina: *HAT* sono richieste per il reclutamento delle proteine del *pre-RC*.
- Lo stato di trascrizione: le *Ori* si trovano spesso vicino a siti di inizio della trascrizione e la forcella di trascrizione produce uno stato di superavvolgimento positivo a monte e negativo a valle, semplice da aprire per la replicazione.

In G_1 le origini selezionate che verranno attivate nella fase *S* sono determinate dal punto di decisione delle origini *ODP*. Specifici domini di replicazione sono regioni di DNA cromosomiale che contengono origini attivate nello stesso momento. Le proteine richieste per la replicazione sono concentrate in strutture nucleari dette replication factories, dove i domini di replicazione co-localizzano con 14 forcelle di replicazione per factory. La pinza scorrevole resa fosforescente *PCNA-GFP* rende visibile i foci di replicazione al microscopio. Le origini dormienti presentano un *pre-RC* assemblato ma non sono attive, ma possono essere attivate velocemente quando una forcella di replicazione vicina entra in stallo. Nei mammiferi l'eucromatina nell'interno del nucleo è replicata precocemente, mentre l'eterocromatina nella periferia è replicata tardivamente.

2.5.3 DNA elicasi

Dopo l'apertura e attivazione all'*Ori* il dsDNA deve essere ancora svolto. Questa operazione viene catalizzata dalla DNA elicasi, un esamero che si lega a un DNA a singolo filamento aperto alla sequenza *Ori* che si muove in basso verso il doppio filamento per svolgerlo alla forcella di replicazione in svolgimento. L'energia necessaria è fornita dall'*ATP*. L'elicasi è inoltre responsabile del reclutamento di altre proteine richieste per la replicazione: il complesso replisomico. Il filamento rimanente viene legato da una proteina. In *E. coli* l'elicasi è la *DnaB* e possiede 6 subunità identiche (omoesamero), negli eucarioti ed archea l'elicasi è complesso *MCM2-7* e comprende 6 subunità diverse (eteroesamero). Ognuna delle 6 subunità lega *ATP* a coppie causando un cambio di conformazione, mentre l'idrolisi e rilascio di *ADP* causa un ritorno alla conformazione iniziale. L'elicasi assume un comportamento pulsante: svolge il DNA e si spinge in avanti. L'elicasi batterica si muove lungo il filamento principale, mentre quella eucariotica in quello lagging.

Risoluzione delle strutture secondarie

Essendo che ssDNA può formare strutture secondarie che rende la sua copia difficoltosa e più sensibile a danno proteico *SSB* nei batteri e *RPA* (replication protein A) negli eucarioti si devono legarsi come omotetrameri al ssDNA in modo da tenerlo aperto. Sono fisicamente rimosse durante la polimerizzazione del DNA.

Risoluzione del superavvolgimento

Mentre il DNA viene svolto viene introdotto uno stress torsionale in quanto la separazione dei filamenti risulta in un superavvolgimento a valle di essa. Questo rende più difficile per l'elicasi proseguire nella separazione. Per questo devono intervenire topoisomerasi che risolvono il problema rompendo transientemente il DNA e permettendo il rilassamento del superavvolgimento. La ligasi successivamente chiude il DNA. Nei procarioti sono presenti solo le topoisomerasi II, negli eucarioti anche le I.

2.5.4 Sintesi di primer a RNA o RNA-DNA dalla DNA primasi

Dopo l'apertura dell'*Ori* la DNA primasi *DnaG* in *E. coli* è reclutata dalla DNA elicasi. Agisce esclusivamente alla forcella di replicazione producendo una corta sequenza RNA o RNA-DNA. Non richiede un esistente 3' – OH per la sintesi a differenza della polimerasi. La primasi batterica possiede due subunità e crea un primer di 10-30 basi, mentre quella eucariotica ne possiede 3 e crea un primer misto DNA-RNA. Due subunità funzionano come primasi mentre una come DNA polimerasi α aggiungendo un primer di DNA all'RNA. L'attività della primasi è unita a quella dell'elicasi e insieme formano il complesso primosoma. Dopo che la subunità DNA polimerasi α della primasi eucariotica crea una corta lunghezza di DNA la DNA polimerasi replicativa III per i batteri e δ o ϵ negli eucarioti la sostituisce e sintetizza il resto del DNA nel processo di polimerase switching. Questo avviene ogni volta che un frammento di Okazaki viene creato tra i primer. La polimerasi replicativa viene reclutata dalla pinza scorrevole e determina l'inizio dell'allungamento.

2.6 Allungamento

2.6.1 Pinza scorrevole

L'allungamento inizia con il reclutamento della pinza scorrevole che permette l'alta processività della DNA polimerasi mantenendola stabilmente legata al DNA. A una sequenza di ssDNA stampo viene caricata la pinza da una proteina di caricamento. Sia la pinza che il suo caricatore sono conservati in batteri, archea ed eucarioti:

- Pinza scorrevole: pinza β nei procarioti e *PCNA* (proliferating cell nuclear antigen) negli eucarioti.
- Proteina caricatrice della pinza: fattore di replicazione *C*: *RFC*.

La pinza scorrevole è un anello con un buco da 35Å che racchiude il primer ssDNA. È molto stabile e rimane associata con il DNA una volta caricata. Affinchè possa associarsi al DNA il caricatore della pinza, una struttura ad anello a 5 subunità deve aprirla. Specifiche subunità del caricatore sono AAA^+ *ATPasi* e quando lega *ATP* causano cambi conformazionali che guidano il legame con la pinza, la sua apertura e il reclutamento al DNA. La pinza scorrevole, una volta legata al primer recluta la DNA polimerasi attraverso un motivo a 8 amminoacidi. Il caricatore della pinza ha bassa affinità per la pinza scorrevole fino a che non è legato all'*ATP*. Quando lo lega il caricatore lega la pinza. Il complesso ha un alta affinità per il primer a ssDNA.

Legame con il primer

Il legame con il primer stimola l'attività *ATPasica* del caricatore che chiude la pinza e si rilascia. L'idrolisi dell'*ATP* riduce l'affinità di *RFC* per il DNA. La pinza rimane associata con il DNA e

recluta l'oloenzima DNA polimerasi permettendo l'inizio dell'allungamento. Il caricatore della pinza rilasciato può essere ricaricato con *ATP* per ripetere il processo a un altro primer. Si dice oloenzima un complesso multiproteico in cui un enzima centrale è associato con componenti addizionali che ne aumentano la funzione.

2.6.2 Sintesi del DNA

Materiali richiesti

La sintesi del DNA richiede il complesso stampo a ssDNA e primer. In vitro il primer è a DNA, mentre in vivo è a RNA o RNA-DNA. Lo stampo è il filamento di ssDNA che è letto dalla DNA polimerasi. Oltre al complesso sono richiesti deossiribonucleotidi: i monomeri *dNTP* come *dATP*, *dCTP*, *dGTP* e *dTTP*. La sintesi avviene sempre nella direzione 5'-3', mentre la lettura nella direzione inversa.

Polimerizzazione

La polimerizzazione del DNA consiste nella formazione dei legami fosfodiesterici. αP in *dNTP* si lega alla terminazione 3'*OH* del primer e viene rilasciato pirofosfato $\gamma P-\beta P$. L'energia che spinge la reazione in avanti deriva dall'idrolisi del pirofosfato da parte della pirofosfatasi che causa l'irreversibilità della reazione.

2.6.3 DNA polimerasi

Le principali DNA polimerasi processive coinvolte nella replicazione del DNA sono la DNA polimerasi III nei batteri e le DNA polimerasi γ e ϵ negli eucarioti. Le DNA polimerasi rimangono attaccate al DNA grazie alla pinza scorrevole per lunghe sequenze prima di dissociarsi rendendo la polimerasi processiva. Le polimerasi processive sono altamente conservate e contengono multipli domini e regioni con diverse funzioni tra cui la ricerca di errori. I tre domini della DNA polimerasi sono detti pollice, dita e palmo che insieme assomigliano a una mano destra. Il DNA in crescita a doppio filamento "il braccio" si trova nel palmo mentre il ssDNA passa attraverso le dita. I domini delle dita aiutano a posizionare il nucleotide in arrivo, il pollice mantiene il dsDNA allungato ma non contribuisce alla reazione di polimerizzazione.

Funzione

L'aggiunta corretta di un nucleotide al filamento in sintesi attiva la polimerasi attraverso un cambio conformazionale: la mano rilascia il DNA dopo aver aggiunto il nucleotide e si muove di una base per leggere lo stampo da 3' a 5'. Il sito attivo della DNA polimerasi possiede gruppi carbossilati di due residui di aspartato con due ioni $2Mg^{2+}$. I siti attivi catalizzano un trasferimento di fosforile unendo il 5'*P* del nucleotide in arrivo al 3'*OH* del DNA in crescita per formare un legame fosfodiesterico. La reazione consiste nell'attacco nucleofilo dal 3'*OH* al α -fosfato del *dNTP* in arrivo, rilasciando i fosfati β e γ come pirofosfato. Gli ioni magnesio sono critici: uno attiva il priming di 3'*OH* abbassando il suo *pKa* e favorendo l'attacco nucleofilo, l'altro interagisce con l'ossigeno negativo dei gruppi fosfato $\beta\gamma$ e posiziona α -fosfato vicino al priming 3'*OH*.

2.6.4 Fedeltà della polimerizzazione del DNA

La DNA polimerasi processiva fa un errore ogni 100 000 nucleotidi. L'identità dei monomeri è controllata durante e dopo l'aggiunta di un *dNTP* al filamento in crescita. La fedeltà è mantenuta

da “proofreading”. Durante l’addizione la polimerasi riconosce il nucleotide corretto grazie alla forma precisa nel palmo quando accoppiato con lo stampo. Nucleotidi scorretti hanno forme diverse e non entrano nel sito attivo con la stessa precisione. Non è richiesta alcuna energia per questo processo.

Riparazione dell’errore

Dopo l’aggiunta del nucleotide la DNA polimerasi si muove al nucleotide successivo nel filamento di stampo ma rallenta quando viene aggiunto un nucleotide scorretto. L’operazione di proofreading serve a riconoscere questi errori. Successivamente la terminazione scorretta è rotta nei legami a idrogeno e viene “flipped out” nel sito dell’esonucleasi che rimuove le basi mal-accoppiate attraverso attività di delezione 3’-5’ di esonucleasi: le funzioni di polimerasi ed esonucleasi sono spazialmente separate. La terminazione 3’OH nel nuovo filamento è successivamente riimmessa nel sito attivo e reinizia la sintesi. Per questo passaggio è richiesta energia.

Cause di errori

Tautomeria dei nucleotide L’inserzione di un nucleotide errato può essere dovuta a un riposizionamento transiente dei doppi legami delle quattro basi all’azoto che cambia la posizione dell’idrogeno legato al gruppo amminico o forme tautomeriche. Una volta accoppiate le basi tautomeriche possono riconvertirsi in posizione normale ma la coppia rimane mal-accoppiata.

- La timina si lega con la forma enolica della guanina.
- L’adenina si lega con la forma imminica della citosina.
- La guanina si lega con la forma enolica della timina.
- La citosina si lega con la forma imminica dell’adenina.

Un altro caso è la depurinazione delle purine, ovvero rimozione o perdita del nucleotide che può accadere a bassi pH e può causare transizioni e trasversioni.

- Transizioni: una purina è sostituita da un’altra purina.
- Transversioni: una purina è sostituita da una pirimidina.

I mal-accoppiamenti tautomerici sono riconosciuti dalla DNA polimerasi e corretti dall’attività dell’esonucleasi o da sistemi di riparazione post-replicazionali. Quando non corretti causano delle mutazioni.

- Transizioni (di entrambi i tipi) una purina è sostituita da un’altra purina ($A \leftrightarrow G$).
- Transversioni (del secondo tipo) una purina è sostituita da una pirimidina: ($A \leftrightarrow C \vee T$, $G \leftrightarrow C \vee T$).

Inclusione di un *NTP* Le concentrazioni intracellulari di *NTP* sono molto più alte rispetto a quelle di *dNTP*. Questi vengono discriminati in base alla posizione 2’OH. La DNA polimerasi contiene anche una tirosina nel dominio catalitico che si scontra con il 2’OH del ribosio prevenendo il loro posizionamento nel sito catalitico e la formazione del legame fosfodiesterico. La capacità di riconoscimento di questi errori è diversa per ogni DNA polimerasi. Viene comunque inserita la base corretta per l’accoppiamento e viene corretta attraverso un processo di riparazione post-replicatorio attraverso *RNasi H1* e *RNasi H2* che rimuovono il ribonucleotide dal filamento e una DNA polimerasi non processiva e una DNA ligasi chiudono il buco.

2.6.5 Sintesi del DNA discontinua

La DNA polimerasi è in grado di sintetizzare in direzione 5'-3' ma entrambe le forcelle si muovono bidirezionalmente. Pertanto per forcella un filamento può essere sintetizzato 5'-3' usando un primer a RNA nel mezzo della bolla e la polimerasi segue l'elicasi. Questo viene detto filamento continuo o guida. Il secondo filamento deve essere sintetizzato in maniera discontinua da 5' a 3' e viene detto filamento ritardato. Nessun primer a monte è disponibile sul DNA svolto mentre l'elicasi si muove. Viene detto "lagging" in quanto la sintesi è più lenta rispetto all'altro filamento in quanto dipende dalla produzione di molti primer. Corti primer a RNA o RNA-DNA sono posti alla forcella di replicazione da una DNA primasi legata alla DNA elicasi nel complesso primosoma. Nei procarioti i primer a RNA sono posti ogni 1000-2000 basi con superavvolgimenti a monte. Negli eucarioti invece gli intervalli sono di 150-200 basi in quanto lo svolgimento è più difficile della cromatina con i nucleosomi. La terminazione 3'OH del primer è il sito di inizio per la DNA polimerasi e la sintesi di un pezzo di DNA da un primer al successivo viene detto frammento di Okazaki.

Maturazione dei frammenti di Okazaki in *E. coli*

Il DNA viene sintetizzato come frammenti corti e discontinui e i frammenti sono uniti quando i primer a RNA sono rimossi. In *E. coli* la DNA polimerasi III sintetizza il DNA che si estende dal primer e si dissocia dal DNA quando raggiunge il primer successivo, mentre la pinza scorrevole rimane attaccata. La DNA polimerasi I rimuove il primer attraverso "nick translation" e riempie il buco lasciato dal primer con DNA, mentre il nick viene unito da una DNA ligasi.

DNA polimerasi I La DNA polimerasi I di *E. coli* è una polimerasi ad alta fedeltà e non processiva che viene usata per rimuovere i primer a RNA e per riparare i danni del DNA. Ha una struttura diversa rispetto a quella processiva. È una proteina con capacità di polimerasi 5'-3' ed esonucleasi in entrambe le direzioni. L'enzima può essere rotto con la proteasi subtilisina, e l'esonucleasi genera il frammento di Klenow con attività esonucleasica che viene usata per rimuovere overhang in 3'OH o riempirlo in 5' per rendere blunt le terminazioni di DNA trattato con enzimi di restrizione.

DNA ligasi La DNA ligasi è un enzima che catalizza la formazione del legame fosfodiesterico tra due molecole di DNA. Lo fa rimuovendo γ -fosfato in 5' e favorendo l'attacco nucleofilo grazie allo ione magnesio.

Maturazione dei frammenti di Okazaki negli eucarioti

La rimozione del primer e il riempimento dello spazio nel nuovo DNA è diversa negli eucarioti. La DNA polimerasi δ sintetizza il DNA partendo da un primer fino a che raggiunge il successivo. Successivamente sposta il primer a valle insieme al frammento di Okazaki creando un "flap". La Flap endonucleasi *Fen1* lo taglia (con omologo presente in archea). La DNA polimerasi δ successivamente si stacca dal DNA e la DNA ligasi unisce il nick. La DNA polimerasi δ viene reclutata da una pinza scorrevole alla terminazione 3'OH al primer successivo.

2.6.6 Attività del replisoma alla forcella di replicazione

La DNA polimerasi replicativa è parte di un grande complesso: il replisoma. Questo è composto da 3 DNA polimerasi, 3 pinze scorrevoli, 3 proteine Tau, un caricatore della pinza, una DNA elicasi, una DNA primasi, *SSB/RFC*, e solo nei procarioti da una DNA polimerasi I. La proteina Tau τ

lega la DNA polimerasi al caricatore della pinza. Si trova un replisoma per forcella di replicazione e sintetizza contemporaneamente sia il filamento principale che quello ritardato.

- Negli eucarioti la DNA polimerasi ϵ sintetizza il filamento guida, mentre le DNA polimerasi α e δ sintetizzano quello ritardato.
- Nei procarioti una DNA polimerasi III sintetizza il filamento guida, mentre due DNA polimerasi III sintetizzano quello ritardato.

Il modello a trombone

Le DNA polimerasi si muovono in direzioni diverse sui due filamenti, ma si muovono seguendo la forcella di replicazione e la stessa elicasi. Questo viene permesso dalla creazione di un loop da parte del filamento ritardato. Il loop deve essere rilasciato dalla polimerasi ogni paio di secondi. Si nota come le polimerasi sul filamento guida e ritardato si trovano insieme spazialmente e sono regolate insieme.

2.7 Terminazione

Quando la replicazione inizia il processo non si ferma fino a quando il replicone è replicato. Nei batteri le forcelle a destra e a sinistra si fondono su un plasmide o sul genoma batterico circolare, mentre negli eucarioti una bolla si fonde con un'altra in arrivo da destra e un'altra in arrivo da sinistra sul cromosoma lineare.

2.7.1 Terminazione nei batteri

Nei batteri una sequenza di terminazione *Ter* determina dove entrambe le forcelle si fondono. È composta da 10 ripetizioni di 23bp ed è localizzata diametricamente opposta a *OriC*. Le ripetizioni di *Ter* sono 5 in senso orario e 5 antiorario. La proteina terminatrice *Tus* si lega a ognuna delle 10 sequenze *Ter* fermando la forcella di replicazione, causando il disassemblaggio del complesso di replicazione e il corto pezzo di DNA non replicato è riempito dalla DNA polimerasi I e chiuso dalla DNA ligasi. *Tus* possiede due facce: una non permissiva che blocca la replicazione e una permissiva che permette la sua continuazione. Dopo la terminazione e ligazione i genomi sono incastrati e vengono liberati da topoisomerasi durante la decatenazione.

2.7.2 Terminazione negli eucarioti

In alcune regioni dei cromosomi eucariotici la replicazione è bloccata in una direzione per impedire scontri con il macchinario di trascrizione che arriva da un altro lato. Nel lievito tale sito è detto barriera della forcella di replicazione e si trova nel rDNA. La proteina *FOB1* si lega al sito e previene la progressione della forcella di replicazione.

Replicazione alla fine di un cromosoma lineare

Il meccanismo per la sintesi del filamento ritardato non può replicare la terminazione di un cromosoma lineare in quanto la rimozione dell'ultimo primer da parte di *RNAasi H* lascia un vuoto che non può essere riempito, pertanto una porzione del cromosoma (telomero) potrebbe essere accorciata durante diversi cicli di replicazione.

2.8 Replicazione dei telomeri

Il problema della replicazione delle terminazioni dei cromosomi è risolto dai virus *T4* unendo diverse copie del cromosoma lineare che poi si dissociano. Gli eucarioti invece possiedono una telomerasi per prevenire il problema della replicazione delle terminazioni. I telomeri variano tra i 100bp e i 20 000bp in base alla specie. Si accorciano ad ogni ciclo di replicazione e possono essere allungati attraverso la telomerasi che aggiunge nuove sequenze telomeriche alla terminazione. Nella *Drosophila* i *non-LTR* retrotrasposoni *HeT-A* e *TART* si traspongono ripetutamente nelle terminazioni cromosomiche per produrre una regione simile ai telomeri. Il mantenimento della lunghezza del telomero attraverso eventi di trasposizione addizionali.

2.8.1 Telomerasi

La telomerasi è una speciale DNA polimerasi, una ribonucleoproteina *RNP* formata da un'enzima e una molecola di RNA. L'enzima o *TERT*, telomerasi trascrittasi inversa è conservata negli eucarioti. L'RNA fornisce lo stampo per la sintesi delle ripetizioni della telomerasi. La telomerasi si lega al overhang di DNA nel telomero a singolo filamento. L'overhang alla terminazione 3' si lega con l'RNA che viene usato come stampo per sintetizzare DNA. La sintesi viene ripetuta più volte.

2.8.2 Mantenimento della lunghezza

Il mantenimento della lunghezza dei telomeri viene raggiunto in due passi:

1. Allungamento dell'overhang 3': la telomerasi si lega all'overhang 3', sintetizza del DNA, lo allunga, si trasloca e riallunga il filamento.
2. Replicazione del filamento complementare: viene sintetizzato un primario da una primasi, il gap viene riempito da una polimerasi, il primer viene rimosso e una ligasi lega i filamenti.

Si noti come rimane comunque un overhang che è coinvolto nella formazione del D-loop.

2.8.3 Problemi nel mantenimento della lunghezza del telomero

Nelle cellule somatiche adulte il gene che codifica *hTERT* non è ben espresso a causa di repressione epigenetica, ma altamente espresso in cellule fetali e staminali. I telomeri si accorciano nelle cellule causando la risposta del danno al DNA e un arresto *p53* dipendente e morte cellulare. Inoltre la repressione epigenetica può impedire il legame della telomerasi con la cromatina del telomero (approfondire).

2.8.4 Il limite di Hayflick

La lunghezza dei telomeri alla nascita è di 15 000bp e per divisione cellulare il telomero si accorcia di 300-1000 coppie di basi. Ad un certo punto tra le 50 e le 60 divisioni si raggiunge il limite di Hayflick in cui una cellula smette di dividersi per evitare ulteriore erosione del cromosoma e raggiunge la senescenza. Molte cellule cancerogene possiedono una telomerasi sovraespressa e continuano a dividersi e proliferare o uno stato epigenetico alterato ai telomeri che porta alle stesse conseguenze.

2.9 Correzione degli errori post-replicativa

Il DNA mismatch repair *MMR* non è parte dell'attività della DNA polimerasi e riduce il tasso di errore di 100 volte. Un nucleotide mismatch causa una distorsione locale nella doppia elica riconosciuta dal dimero *MutS* che si lega ad essa e recluta il dimero *MutL* che stabilizza il legame con il DNA distorto. La riparazione avviene sul nucleotide di nuova sintesi, riconosciuto grazie alla metilazione sullo stampo di *GATC*. Il complesso *MutS-MutL* recluta l'endonucleasi *MutH* che si lega alla sequenza metilata sul filamento stampo più vicino al mismatch, taglia il filamento di nuova sintesi vicino alla sequenza, aiutata occasionalmente dall'elicasi *UvD*. L'esonucleasi 3'-5' rimuove il DNA fino al sito di errore e la DNA polimerasi III lo risintetizza. Quando si incontra il DNA dopo il taglio i filamenti sono riattaccati da una ligasi. L'efficienza di riparazione si abbassa allontanandosi da *GATC*.

2.10 Mantenimento delle modifiche istoniche

Si nota come i nucleosomi sono in parte rimossi dal replisoma in movimento. Le proteine istoniche devono pertanto essere reclutate ai posti corretti nel DNA appena reclutato e mantenere le modifiche epigenetiche (eredità epigenetica). I nucleosomi sono distrutti dalla forcella di replicazione e devono rilegarsi dopo che questa li ha superata e le modifiche epigenetiche sono aggiunte su essi. Questo avviene in quanto il nucleosoma parentale non è rimosso: 50% sono distribuiti equamente tra i filamenti singoli: avviene una segregazione degli istoni parentali. Di questi solo quelli marcati epigeneticamente 2(*H3-H4*) sono mantenuti, mentre gli altri sono rimossi anche se rimangono vicini. Quelli non marcati sono inclusi nei nuovi nucleosomi attraverso il chaperone *CAF1* insieme ai 2(*H2A-H2B*) parentali o nuovi attraverso il chaperone *NAP1*. Pertanto si trovano 4 tipi di nucleosomi nei filamenti sintetizzati. Infine il nucleosoma parentale originale serve come stampo per produrre lo stato epigenetico su tutti i nucleosomi che sono rimasti vicini.

2.11 DNA polimerasi specializzate

Tutte le DNA polimerasi catalizzano la stessa sintesi del DNA ma differiscono nella velocità e tasso di errore.

- DNA polimerasi replicative: sono processive e catalizzano la replicazione dei genomi durante la fase *S*, hanno un'alta fedeltà di polimerizzazione e un'attività esonucleasica 3'-5' altamente efficiente.
- DNA polimerasi non-processive o distributive: rappresentano la maggior parte di DNA polimerasi, polimerizzano corte lunghezze per mantenerlo in uno stato sano (riparazione), hanno bassa fedeltà, sintetizzano il DNA sul filamento danneggiato o sintesi di translesione *TLS*.

Certe DNA polimerasi hanno altre attività e tutte sono reclutate dalla pinza scorrevole sul DNA che permette loro di stare attaccate al DNA, altrimenti dopo ogni addizione di nucleotide potrebbero separarsi.

2.11.1 DNA polimerasi batteriche

Altamente conservate in tutti i procarioti *E. coli* possiede 5 DNA polimerasi:

- DNA polimerasi I: attiva nella riparazione del DNA ed estende e matura i frammenti di Okazaki.
- DNA polimerasi II: attiva nella riparazione del DNA.
- DNA polimerasi III: polimerasi replicativa che catalizza la polimerizzazione del DNA a grande lunghezza.
- DNA polimerasi IV e V: riparano danni al DNA che bloccano il replisoma.

2.11.2 DNA polimerasi specializzate

DNA polimerasi *TLS*

Esistono in tutti gli organismi e sono le DNA polimerasi per la sintesi di traslesione. Il DNA in una cellula subisce danno costante e deve essere riparato, in particolare il dimero di timidina nello stampo del filamento singolo causa un arresto della forcella, rottura ed il doppio filamento e morte cellulare. Le DNA polimerasi *TLS* promuovono la replicazione, ma hanno alto tasso di errore a causa dell'assenza dell'attività nucleasica. Quando il dimero di timidina causa uno stacco della DNA polimerasi III ad esso si lega la DNA polimerasi *TLS* che avendo un sito di legame più aperto riesce a superare il blocco e continuare per un breve tratto la sintesi con minore fedeltà prima di essere di nuovo sostituita dalla DNA polimerasi III.

Trascrittasi inversa

La trascrittasi inversa è una DNA polimerasi dipendente da RNA: copia l'RNA in ssDNA o cDNA. Usa un singolo filamento di RNA come stampo e necessita di un primer. Sono molto simili alle DNA polimerasi e sono codificate da virus e retrotrasposoni, elementi di DNA mobile negli eucarioti. La telomerasi è una trascrittasi inversa specializzata. I virus a RNA o retrovirus come *HIV-1* e influenza A hanno un genoma a RNA che codifica per una trascrittasi inversa nella cellula ospite. Il gema è trascritto in dsDNA che si integra nel genoma ospite.

Capitolo 3

Trascrizione

3.1 Panoramica della trascrizione

L'informazione conservata nel DNA è utilizzata per creare proteine o molecole di RNA funzionale. Si intende per trascrizione il processo di copia di un filamento di DNA in una molecola di RNA detta trascritto. Il processo di trascrizione viene svolto da una RNA polimerasi. Un cofattore esterno RNA polimerasi elicasi separa i filamenti di DNA e permette i ribonucleosidi trifosfato di accoppiarsi con il filamento stampo. Per produrre una proteina da una molecola di DNA la sequenza di RNA è letta dal ribosoma nella traduzione. L'RNA in questo processo viene detto RNA messaggero *mRNA*. Si nota come la RNA polimerasi non possieda attività esonucleasica e pertanto non possa correggere errori di mal-accoppiamento. Il tasso di errore è di 10^{-4} e il tasso di da 40 a 80 nucleotidi al secondo. Si nota come rispetto alla DNA polimerasi è più lenta, inefficiente e meno accurata.

3.1.1 Il processo di trascrizione

La trascrizione può essere divisa in iniziazione, allungamento e terminazione. Inizia quando la RNA polimerasi si lega a una sequenza di DNA che precede il gene: il promotore. Il sito di inizio di trascrizione *TSS* è la prima base ad essere trascritta ed è notata con +1. L'RNA viene trascritto nella direzione 5'-3' con il filamento letto in direzione opposta come nella sintesi del DNA. Pertanto si indica come basi a monte quelle 5' e a valle quelle 3'.

Iniziazione

Durante l'iniziazione la RNA polimerasi separa i filamenti di DNA per creare una bolla di trascrizione tra le 12 e le 14bp e inserisce i primi ribonucleoside trifosfati *NTP* mentre si trova al promotore. Quando il RNA è di lunghezza sufficiente la RNA polimerasi lascia il promotore "promoter clearance" e cambia conformazione per essere più stabilmente associata con il DNA permettendo l'allungamento del RNA.

Allungamento

L'allungamento inizia dopo la clearance del promotore e la RNA polimerasi si muove lungo il DNA aggiungendo ribonucleotidi e allungando il trascritto di RNA. La RNA polimerasi svolge il DNA a valle e chiude quello a monte mantenendo la dimensione della bolla di trascrizione di dimensione

costante per impedire la formazione di R-loop. Nella bolla una regione del trascritto di 8-10bp è accoppiata con il DNA mentre il resto è estruso dalla polimerasi.

Terminazione

L'allungamento continua fino a che la polimerasi incontra una sequenza di DNA detta terminatore che segnala la fine della sintesi di RNA. Il RNA è rilasciato e la RNA polimerasi si dissocia dal DNA.

3.1.2 Nomenclatura dei geni

In un gene viene indicato con +1 il sito di inizio della trascrizione *TSS*, con numeri negativi la zona del promotore e con *TTS* il sito di terminazione della trascrizione. Si dice con prossimale la zona più vicina al *TSS*, distale quella che si trova allontanandosi verso il *TTS*. Si intende per sequenza codificante *CDS* la sequenza del gene senza introni, mentre con open reading frame *ORF* la sequenza con gli introni.

3.1.3 Regolazione della trascrizione

La trascrizione è regolata per produrre il RNA richiesto al tempo corretto. La cromatina negli eucarioti presenta una sfida per la trascrizione in quanto i nucleosomi prevengono il legame e il movimento del macchinario di trascrizione attraverso la cromatina. Sono pertanto richiesti:

- Rimodellamento dei nucleosomi: per riposizionare gli istoni lontano dal DNA che deve essere trascritto.
- Chaperone degli istoni per riassemblare e disassemblare i dimeri nucleosoma-istone.
- Enzimi che modificano le proteine istoniche epigeneticamente per permettere o prevenire il legame di proteine che regolano la trascrizione.

3.2 L'enzima centrale della RNA polimerasi

Se i procarioti possiedono 1 RNA polimerasi gli eucarioti ne possiedono 3 principali:

- RNA polimerasi I *Pol I*: trascrive i geni di grandi RNA ribosomali *rDNA*. Agisce nei nucleoli 5-6 negli umani, 1 nel lievito.
- RNA polimerasi II *Pol II*: trascrive i geni di RNA messaggero *mRNA*, RNA non codificanti corti e lunghi, *miRNA* con ruolo nella regolazione dell'espressione genica attraverso interferenza e *snRNA* RNA piccoli nucleari con un ruolo nel processamento di *mRNA*. Agisce nel nucleoplasma.
- RNA polimerasi III *Pol III*: trascrive una varietà di RNA come tutti gli RNA transfer *tRNA*, il piccolo 5S RNA ribosomale, *snRNA U6* componente dello spliceosoma, *7SL RNA* o long non-coding RNA parte della particella di riconoscimento del segnale che regola la traduzione. Agisce sia nei nucleoli che nel citoplasma.

Le piante possiedono una quarta RNA polimerasi con ruolo di trascrizione degli RNA non codificanti con ruolo nell'espressione genica. Gli eucarioti possiedono anche RNA polimerasi mitocondriali. Si intende per nucleoli raggruppamenti non circondati da membrana di 6 ripetizioni di rDNA trascritti da *Pol I* e *Pol III*.

3.2.1 Struttura

Le RNA polimerasi sono formate da diverse subunità che formano il nucleo polimerasico.

RNA polimerasi batteriche

Le RNA polimerasi batteriche sono le più piccole con 5 subunità che si assemblano in un complesso nucleare con lobi simili a mascelle che formano una pinza. Le mascelle sono formate dalle subunità β e β' con le due α e ω alla base. La base della fessura è il sito attivo dell'enzima. Ogni subunità α possiede un dominio N terminale αNTD e un dominio C terminale αCTD unite da un collegatore flessibile.

Subunità e loro ruolo

- α (2): responsabile per l'assemblaggio del complesso.
- Domini N terminali: interagiscono con le subunità β e β' .
- Domini C terminali: si legano al DNA promotore: elemento a monte.
- β (1): contengono il sito catalitico, formano i legami fosfodiesteri legando Mg_2^+ e svolgono la correzione degli errori.
- β' (1): mantiene l'enzima legato al filamento stampo, svolge e riavvolge il dsDNA.
- ω (1): è un chaperon: promuove la stabilità strutturale della RNA polimerasi.

RNA polimerasi di eucarioti ed archea

Tutte le RNA polimerasi possiedono le 5 subunità centrali viste prima altamente conservate, specialmente al sito attivo. Si trovano ulteriori subunità in RNA polimerasi di archea ed eucarioti ordinate intorno alle 5 centrali. L'attività catalitica pertanto rimane conservata in tutti e tre gli alberi della vita.

Ruoli della RNA polimerasi II Oltre a trascrivere il DNA la RNA polimerasi II accoppia la trascrizione con il processamento del trascritto di RNA: il dominio C terminale CTD di *Pol II* è cruciale per questa funzione: è la terminazione della subunità *Rpb1* e esiste come ripetizioni di una sequenza di 7 amminoacidi: *Tyr-Ser-Pro-Thr-Ser-Pro-Ser* con 26 ripetizioni nel lievito e 52 negli umani.

3.3 Riconoscimento dei promotori

Il nucleo della RNA polimerasi può sintetizzare il RNA ma non può riconoscere e legarsi alla sequenza promotrice di un gene. Si rendono pertanto necessarie ulteriori subunità che si legano direttamente al promotore. Si forma pertanto un oloenzima dalla loro unione con l'enzima nucleare.

3.3.1 Batteri

Nei batteri le subunità che si legano al promotore sono dette fattori σ . Ne esistono diverse che riconoscono promotori specifici per promuovere la trascrizione di geni specifici in base a particolari condizioni di crescita. I fattori σ si legano a sequenze che definiscono i promotori batterici composti tipicamente da due elementi: uno a -35 e uno a -10 detto box di Pribnow. Ogni fattore sigma ha una sequenza di legame specifico e una particolare spaziatura tra i due elementi. Questo in modo da regolare la trascrizione: più vicine le sequenze e la spaziatura a quella particolare del fattore σ più forte il legame e più alti tassi di trascrizione.

Fattori sigma

Nell'oloenzima della RNA polimerasi il fattore σ^{70} è costitutivo e 3 dei suoi 4 domini riconoscono specifici elementi dei promotori:

- Il dominio 2 si lega alla regione -10 e aiuta a separare il dsDNA “promoter melting”.
- Il dominio 3 riconosce le due basi della regione estesa -10 .
- Il dominio 4 riconosce l'elemento -35 , è attaccato a una parte flessibile del nucleo dell'enzima che permette diversa spaziatura tra -35 e -10 .

Alcuni fattori sigma sono regolati in risposta a condizioni ambientali o di sviluppo. Possono essere regolati sia a livello trascrizionale o traduzionale o alterando la stabilità del proprio mRNA. Sono inoltre regolati da:

- Fattori pro-sigma: proteine con domini inibitori che devono essere rotte prima che il fattore σ possa associarsi con l'enzima RNA polimerasi.
- Fattori anti-sigma: proteine che legano i fattori σ inibendo la loro funzione.

Fattori anti-sigma e regolazione dell'assemblaggio del flagello in *Salmonella typhimurium* Mentre le proteine che formano la base del flagello sono sintetizzate il fattore anti-sigma *FlgM* si lega a σ^F impedendo il suo legame con la RNA polimerasi- σ^{70} . σ^F promuove la trascrizione di geni necessari per il completamento dell'assemblaggio del flagello. Negli ultimi passi della sintesi delle proteine del flagello *FlgM* è esportata dalla cellula in modo che σ^F possa legarsi alla RNA polimerasi sostituendo σ^{70} e promuovere la trascrizione dei geni dell'assemblaggio permettendo la creazione del flagello.

3.3.2 Eucarioti

Anche i promotori eucariotici e di archea necessitano di ulteriori proteine per direzionare la RNA polimerasi ai promotori. Questi sono detti fattori di trascrizione generali *TF* seguito dal numero della *Pol*. Negli eucarioti *TFII* si assemblano al promotore e con l'enzima del nucleo formano il complesso di pre-iniziazione. *Pol I* e *Pol III* richiedono diverse proteine per formare il complesso di pre-iniziazione. I promotori per la RNA *Pol II* possiedono spesso una TATA box con sequenza di consenso *TATAA* tra le 25 e le 30bp a monte dell'inizio della trascrizione. Tutte le polimerasi eucariotiche necessitano della TATA binding protein *TBP* per iniziare la trascrizione che viene riconosciuta da *TBP* associated factors *TAF* formando il complesso *TFIID*. Alcuni altri elementi includono l'elemento di riconoscimento *TFIIB BRE*, l'elemento iniziatore *INR* e l'elemento promotore a valle *DPE*. Molti promotori non hanno questi elementi e la loro variabilità rende difficile la

3.4. INIZIAZIONE DELLA TRASCRIZIONE E TRANSIZIONE A UN COMPLESSO DI ALLUNGAMENTO

loro predizione. Il complesso di pre-iniziazione contiene 32 proteine fattori di trascrizione generali e la RNA *Pol II* con 12 subunità. I promotori per *Pol II* possono dividersi in:

- Prossimali: distanti meno di 200bp da *TSS* nei lieviti unicellulari.
- Distali: distanti fino a 10kb dal *RSS* come negli eucarioti multicellulari.

3.3.3 Formazione del complesso di pre-iniziazione

Il primo passo nell'assemblaggio del complesso di pre-iniziazione è il legame di *TFIID* alla *TATA* box attraverso *TBP* che si lega alla fessura minore del DNA causando una piegatura locale. Altre componenti di *TFIID* i *TBP-associated factors TAF* mediano il riconoscimento di altri elementi del promotore con *INR* e *DPE* per *TFIID*. Dopo l'associazione di *TFIID* con il DNA viene reclutata *TFIIB* che riconosce l'elemento *BRE* e ci si lega asimmetricamente aiutando a determinare la direzione di trascrizione. *TFIIB* è simile al fattore σ batterico. Dopo il legame di *TFIID* e *TFIIB* viene reclutata *TFIIA* che lega e stabilizza l'interazione *TBP-DNA*. Successivamente si legano *TFIIF*, *TFeIIE* e *TFIIH*. L'ultima di queste catalizza attivamente lo svolgimento del DNA. In vivo questo complesso necessita di un complesso *Mediatore* per attivare molti geni trascritti da *Pol II* (non agisce con le altre RNA polimerasi). Altre proteine necessarie contengono enzimi modificatori degli istoni e rimodellatori dei nucleosomi.

3.3.4 Differenze tra le RNA polimerasi eucariotiche

Il nucleo delle RNA polimerasi è simile in *Pol I*, *Pol II* e *Pol III* e tutte usano *TBP* per iniziare la trascrizione. *TBP* è un monomero con due domini simmetrici attraverso i quali lega il DNA e recluta diversi complessi iniziatori della trascrizione per ogni polimerasi. Le tre RNA polimerasi infatti usano proteine diverse per iniziare la trascrizione che riconoscono promotori diversi e on diversi fattori di trascrizione generali.

RNA polimerasi I

La RNA polimerasi I trascrive il rRNA e si lega a un promotore con un elemento di nucleo riconosciuto da *TP* e da un elemento di controllo a monte *UCE*.

RNA polimerasi III

La RNA polimerasi III trascrive 5S RNA e tutti i tRNA e riconosce elementi promotori chiave: box *A* e *C/B* nella sequenza codificante insieme ad altri elementi a monte.

3.4 Iniziazione della trascrizione e transizione a un complesso di allungamento

Una volta che la RNA polimerasi è in posizione l'olocomplesso di RNA polimerasi è il promotore a doppio filamento chiuso vengono detti il complesso chiuso. Le RNA polimerasi batteriche aprono da 12 a 14bp di dsDNA formando la bolla di trascrizione. Invece in *Pol II* la bolla di trascrizione è aperta dalle subunità elicasiche di *RFIIH* e richiede *ATP*. A questo punto la RNA polimerasi unita a un promotore aperto viene detta complesso aperto. La RNA polimerasi non crea un RNA di lunghezza completa al primo tentativo: inizialmente produce un RNA di 9-12 nucleotidi detto trascritto abortivo di iniziazione.

3.4.1 Modello di inizio abortivo

- Inizio a passaggio transiente: la RNA polimerasi produce un RNA abortivo, si ferma e torna indietro.
- Inizio a bruco: la RNA polimerasi produce il RNA abortivo allungandosi flessibilmente prima di tornare indietro.
- Inizio ad accartoccamento: la RNA polimerasi accartoccia il DNA al suo interno per produrre il RNA abortivo.

Sia il fattore σ batterico che l'eucariote *TFIIB* sono coinvolti nell'iniziazione abortiva: possiedono infatti un loop che si estende nel sito attivo delle RNA polimerasi in modo che blocchi il trascritto in allungamento dopo 9-12 nucleotidi. Lo spostamento "chiusura" del loop aiuta la polimerasi a separarsi da promotore nella promoter clearance e andare nella modalità di allungamento.

3.4.2 Sintesi del RNA

Una volta che si forma la bolla di trascrizione un timone nella RNA polimerasi mantiene il filamento di stampo e l'altro separati. Il filamento di stampo si lega alla RNA polimerasi nel sito attivo e il RNA formato esce attraverso un canale di uscita tra il muro e il coperchio. I ribonucleotidi *NTP* entrano nel sito attivo attraverso un poro e si accoppiano con le basi del filamento stampo. I nucleotidi successivi sono attaccati attraverso attacco nucleofilo alla terminazione 3' della molecola di RNA in crescita, formando un legame fosfodiesterico e rilasciando pirofosfato. Due ioni di magnesio sono presenti nel sito attivo e catalizzano l'addizione di nucleotidi attivando il 3'OH dell'ultimo nucleotide nel filamento di RNA e stabilizzando una carica negativa sull'ossigeno che sta lasciando sul γ -pirofosfato.

3.4.3 Promoter clearance

Durante la promoter clearance la RNA polimerasi subisce un cambio conformazionale nel coperchio e nel timone nella subunità β' che associa l'enzima molto stabilmente con il DNA e riduce il legame con i fattori di iniziazione dissociandosi da essi. Le polimerasi eucariotiche sono anche fosforilate mentre si convertono al complesso di allungamento. L'enzima del nucleo della RNA polimerasi svolge la sintesi del RNA attraverso il sito attivo nella subunità β .

3.5 Allungamento della trascrizione

La trascrizione è processiva una volta che la fase di allungamento inizia. La RNA polimerasi rimane associata con il DNA e aggiunge centinaia o migliaia di basi sul RNA crescente con una velocità tra i 20 e i 50 nucleotidi al secondo. La bolla di trascrizione si muove a 12-14 basi costanti. La RNA polimerasi infatti separa il filamento di DNA a valle senza elicasi ulteriori e le coppie di basi si riformano quando questa si muove nella base successiva. I ribonucleotidi di addizione più recente si trovano nella bolla di trascrizione.

3.5.1 Pause e arresti nella trascrizione

Le RNA polimerasi possono fermarsi a causa di ostruzioni fisiche nel "transcriptional pausing" che può avvenire quando il filamento stampo forma una forcina a causa di corte sequenze complementari

3.5. ALLUNGAMENTO DELLA TRASCRIZIONE

o attraverso la formazione di deboli ibridi DNA-RNA nella bolla. Il “transcriptional arrest” avviene quando l’enzima non può riprendere la sintesi. Il transcriptional pausing può essere aumentato o diminuito da fattori di allungamento.

3.5.2 Processamento del mRNA

La RNA polimerasi II crea mRNA ma non si trova nella forma attiva *pre-mRNA*. Questo deve essere processato in modo da diventare attivo. Negli eucarioti l’allungamento trascrizionale è associato con il processamento del mRNA. La regione *CTD* fosforilata della subunità *Rpb1* viene coinvolta: la quinta serina nella ripetizione di 7 amminoacidi è fosforilata *S5-P* mentre la RNA polimerasi II esce dalla regione del promotore e inizia il processo di allungamento. Questa recluta enzimi di processamento del RNA come quelli che aggiungono un cap di guanosina metilata alla terminazione 5’ del mRNA. L’allungamento è brevemente messo in pausa per permettere questo evento. Successivamente il capping porta alla fosforilazione della seconda serina sulla *CTD S2 – P* che permette la ripresa dell’allungamento da parte della polimerasi.

3.5.3 Backtrack

Quando c’è una pausa nella sintesi del RNA il complesso di allungamento può fare del backtrack. Questo processo può avere un ruolo nella correzione di errori: se sono incorporate basi mal-accoppiate il doppio DNA-RNA può essere distorto e causare pausa fisica della RNA polimerasi. La regione mal-accoppiata può essere poi rotta permettendo alla RNA polimerasi di avere un’altra occasione per essere incorporata nel nucleotide corretto. La RNA polimerasi inverte la sua direzione e il RNA appena sintetizzato protrude dal complesso. La pausa della trascrizione e fattori di rottura *TFIIS* per gli eucarioti e *GreB* in *E. coli* rompono il RNA protrudente 3’ aumentando l’attività endonucleasica della RNA polimerasi e poi la trascrizione può riprendere. Questi fattori si legano nella regione a imbuto della RNA polimerasi attraverso cui protrude il RNA in caso di backtrack. I fattori posizionano uno ione magnesio al sito attivo che attiva una molecola di acqua per l’idrolisi del legame fosfodiesterico. Il taglio è fatto 3 nucleotidi prima della fine.

3.5.4 Problemi dell’allungamento

I nucleosomi impediscono il movimento della RNA polimerasi

Gli eucarioti usano chaperones degli istoni per disassemblare i nucleosomi delicatamente a valle della RNA polimerasi e li riassemblano dietro essa. Questi chaperones sono detti *FACT* (facilitano la trascrizione della cromatina) e un esempio è *Spt6*. I *FACT* disassemblano parzialmente il nucleosoma rimuovendo un singolo dimer *H2A-H2B* che allenta sufficientemente il DNA intorno al nucleosoma in modo che la RNA polimerasi II possa trascriverlo. *FACT* assiste inoltre con il riposizionamento del dimer associato mentre *Spt6* aiuta a garantire l’assemblaggio corretto dietro la RNA polimerasi II. La ristorazione dell’organizzazione cromatinica corretta potrebbe prevenire l’occupazione di siti TATA critici nelle regioni codificanti da *TBP* e il reclutamento di RNA polimerasi II.

Superavvolgimento a valle del DNA trascritto

La bolla di trascrizione si muove lungo il DNA svolgendo la doppia elica continuamente. Questo porta a cambi nel superavvolgimento: un aumento di positivo a valle e di negativo a monte. Tali cambi potrebbero causare pause della RNA polimerasi e la tensione deve essere sollevata da topoisomerasi I o II.

3.6 Terminazione della trascrizione

La RNA polimerasi deve fermarsi a trascrivere il gene nel posto corretto, rilasciare il trascritto e dissociarsi dal DNA. Sequenze di terminazioni nel DNA segnalano la RNA polimerasi di fermare la trascrizione.

3.6.1 Batteri

Le classi di sequenze terminatrici batteriche sono intrinseche Rho-indipendenti e Rho-dipendenti.

Intrinseche o semplici

I terminatori intrinseci terminano la trascrizione senza altri fattori. Quelli batterici presentano una sequenza di ripetizioni invertite che quando trascritte formano uno stem-loop nel RNA. Sono spesso ricche di *G – C* per legami forti. Presentano inoltre stringhe di 8-10 residui che si accoppia con il trascritto *poli-U* nella bolla di trascrizione. La formazione di forcine di RNA forzata tira fuori il RNA dal sito attivo grazie all'accoppiamento con il DNA debole nella bolla di trascrizione finale.

Rho-dipendenti

Dove la sequenza terminatrice non è sufficiente intervengono fattori proteici: in *E. coli* questi geni necessitano dell'elicasi Rho per terminare la trascrizione: *terminatori ρ-dipendenti*. Questi geni contengono ripetizioni invertite.

Rho Questo è un anello formato da *ATPasi* esameriche che legano a un'area terminale del RNA ricca in *C*: la sequenza *RUT* (rho utilization site) lunga fino a 40 nucleotidi. L'anello ha una struttura aperta che si lega al RNA. Una volta legato l'anello si chiude e l'indrolisi del *ATP* spinge il RNA attraverso l'anello. La formazione di forcine risulta in una pausa della RNA polimerasi. Rho successivamente raggiunge la RNA polimerasi e la fa separare dal DNA. Rho si svolge e rilascia il RNA dallo stampo ssDNA a cui è legato.

3.6.2 Eucarioti

Terminazione della RNA polimerasi I

Sono presenti sequenze *Poli-A* ma il grande gene rRNA contiene una sequenza ripetuta riconosciuta da una proteina terminatrice *TTF-1* (transcription termination factor for RNA polymerase I). *TTF-1* legato blocca fisicamente la trascrizione causando la disassociazione di *RNA polimerasi I* e il rilascio del RNA di nuova sintesi.

Terminazione della RNA polimerasi II

La terminazione dei geni trascritti dalla RNA polimerasi II è accoppiata con un processamento della terminazione 3' del mRNA. La RNA polimerasi II continua a trascrivere il DNA dopo il segnale di rottura e poliadenilazione e il mRNA è rotto e viene aggiunta una coda *poli-A* alla terminazione *C*. Questa coda non è codificata dal DNA.

3.7. PRINCIPI DELLA REGOLAZIONE DELLA TRASCRIZIONE

Modello allosterico La RNA polimerasi II trascrive attraverso il segnale di rottura e poliadenilazione, proteine di procesamiento del RNA si associano con i segnali di procesamiento e il *CTD*. La rottura o riconoscimento delle proteine di procesamiento causa cambi conformazionali che portano alla dissociazione di *Pol II* dal DNA.

Modello torpedo Dopo la rottura come nel modello allosterico il RNA a valle è digerito dalla endonucleasi *Rat1*, il torpedo che continua a degradare il RNA fino a che incontra la RNA polimerasi II che causa la sua dissociazione dal DNA.

Conclusioni È possibile che una combinazione dei due modelli sia responsabile per la terminazione. Inoltre fosforilazione di *CTD* a *S2* e *S7* potrebbe ridurre addizionalmente l'affinità della RNA polimerasi II per il DNA dopo la rottura e polieadenilazione.

Terminazione della RNA polimerasi III

La RNA polimerasi III riconosce siti di terminazione intrinseci ricchi di *A* che destabilizzano l'ibrido DNA-RNA come quelli batterici.

3.7 Principi della regolazione della trascrizione

La regolazione della trascrizione è la chiave per fornire a una cellula la corretta quantità di prodotto genico al momento corretto. I regolatori di trascrizione come i fattori di trascrizione possono avere effetti drammatici sull'espressione genica. Inoltre tale regolazione sottostà alla differenziazione cellulare e allo sviluppo in tutti gli eucarioti grazie all'utilizzo di una grande varietà di meccanismi.

3.7.1 Meccanismi di regolazione

L'oloenzima RNA polimerasi può trascrivere qualunque gene con un promotore funzionale, pertanto un livello di regolazione è la forza del promotore (per esempio la fedeltà delle sequenze -35 e -10). La maggior parte della regolazione avviene da regolazione su un gene obiettivo durante l'iniziazione (più efficiente e utilizzata), all'allungamento o terminazione o regolando il RNA trascritto. Le proteine che aumentano la trascrizione sono dette attivatori e stimolano il legame della RNA polimerasi al promotore attraverso contatto fisico. Le proteine che diminuiscono la trascrizione sono dette repressori e impediscono il legame della RNA polimerasi con il DNA.

Sequenze regolatorie

Molti geni sono regolati da proteine che si legano vicino al gene, principalmente a monte del promotore. Le sequenze regolatrici sono regioni specifiche del DNA a cui la proteina regolatrice si lega. Si possono trovare vicino al promotore o a molte kilobasi di distanza.

Batteri Nei procarioti le sequenze riconosciute dai regolatori sono dette siti operatori e sono vicine al promotore o si sovrappongono ad esso. Se gli operatori sono distali al gene il DNA deve formare un loop in modo che la proteina regolatrice possa interagire con la polimerasi aiutata da proteine piegatrici del DNA.

Eucarioti Negli eucarioti i geni sono controllati da sequenze regolatorie distanti fino a migliaia di basi. Sono classificate come enhancers (amplificatori), silencers (repressori) o insulators. Le sequenze regolatorie legano diverse proteine aumentando la capacità e l'intensità della regolazione. Le sequenze possono trovarsi a monte, valle o nel gene. Come nei batteri sequenze distanti regolano il gene attraverso loop.

3.7.2 Enhancer

Si dice enhancer un effettore positivo a lunga distanza che potrebbe trovarsi fino a 10 000bp dal TSS. È una sequenza di DNA promotrice agente in *cis*. I regolatori che si legano a un enhancer si dicono agenti in *trans*. Questi comprendono fattori di trascrizione base e attivatori trascrizionali. È lungo 500bp e contiene tra i 10 e i 12 siti di legami per fattori di trascrizione. Il legame dei fattori è cooperativo o esclusivo ma tutti hanno un effetto additivo sulla trascrizione. Le sequenze possono agire in entrambi gli orientamenti rispetto alla direzione della trascrizione. Il complesso formato da sequenza enhancer e dagli attivatori legati ad esso viene detto *enhanceosome*. Questo complesso forma un anello e connette con il PIC attraverso il complesso mediatore. Il livello di trascrizione è direttamente proporzionale con il numero di attivatori legati.

3.7.3 Silencer

Il silencer è assolutamente analogo al enhancer: lega repressori trascrizionali formando il *silenceosome* e compete con il enhanceosome per il legame con il complesso mediatore che attiva la trascrizione.

3.7.4 Insulator

Lo insulator è una sequenza di DNA a cui proteine repressori si legano che hanno effetto negativamente tra il enhancer e il promotore, di fatto isolandolo. Si può trovare a monte, a valle o in una sequenza codificante di un gene. Spesso separano zone eterocromatiche da quelle eucromatiche, sono pertanto gli elementi di barriera.

3.7.5 Struttura delle proteine regolatrici

Le proteine regolatrici devono essere in grado di riconoscere specificatamente la corretta sequenza regolatoria: ogni regolatore ha un dominio legante il DNA che riconosce una sequenza specifica. Sono spesso modulari, con domini addizionali che aiutano l'oligomerizzazione, coattivano o coreprimono la trascrizione o interagiscono con altri regolatori e RNA polimerasi. I regolatori eucariotici controllano la trascrizione reclutando coattivatori o corepressori che non sono in grado di legare direttamente il DNA.

3.7.6 Eventi con effetto sulla trascrizione

Programmi di sviluppo o cambiamenti nelle condizioni di crescita possono influenzare la trascrizione di molti geni. Inoltre piccole molecole come effettori allosterici (estrogeno) possono legarsi direttamente a proteine regolatrici cambiando la loro conformazione. La fosforilazione o altre modifiche covalenti possono modificare come proteine regolatorie interagiscono con il DNA e altre proteine. Queste modifiche con altri fattori come abbondanza e localizzazione del regolatore servono per aggiustare i livelli di trascrizione.

La struttura cromatinica

La struttura cromatinica ha un ruolo cruciale nella trascrizione eucariote: cromatina iper-acetilata tende ad essere attivamente trascritta, mentre i livelli di trascrizione sono più bassi in cromatina ipo-acetilata. L'attivazione dei geni è accoppiata con il reclutamento di acetil-trasferasi istoniche, mentre la deacetilasi istonica può reprimerla. Altre modifiche istoniche come metilazione, fosforilazione, ubiquitinazione e SUMOilazione possono avere un effetto sulla trascrizione. I codici istonici sono proposti in certe combinazioni di modifiche istoniche che portano a eventi di trascrizione.

3.7.7 Confronto con replicazione

Si nota come la trascrizione è più lenta della replicazione in quanto:

- La DNA polimerasi processiva non cade dal ssDNA.
- La RNA polimerasi si ferma quando si formano deboli ibridi RNA-DNA, forcine e altre strutture che possono portare allo stop e alla rimozione della RNA polimerasi.
- Il replisoma che incontra una RNA polimerasi sul DNA causa una sua rimozione.
- R-loop possono formarsi nella bolla di trascrizione e bloccare la sua progressione fino alla risoluzione da parte di RNAasi.
- ssDNA può formare strutture secondarie.
- Ci sono difficoltà nel riavvolgimento del dsDNA dietro la RNA polimerasi.
- La correzione degli errori della RNA polimerasi è debole e richiede altre proteine, rallentando il processo.
- L'inizio della trascrizione è più complicato rispetto alla replicazione.
- Negli eucarioti sono importanti modifiche epigenetiche a valle che richiedono enzimi modificatori degli istoni attaccati alla RNA polimerasi.
- Il processamento del RNA avviene in contemporanea con la sua trascrizione.

3.8 Domini leganti il DNA in proteine che regolano la trascrizione

3.8.1 Helix-turn-helix

Un dominio comune delle proteine regolatrici è il helix-turn-helix. La prima α -elica e la α -elica di riconoscimento R -elica che entrano nella fessura principale. Molti agiscono come dimeri con le eliche di riconoscimento spaziate di $3.4nm$ (un giro di elica) per potersi trovare nella fessura principale vicina. La struttura a dimero permette di aumentare la selettività e l'affinità con la sequenza di $3-7bp$ riconosciuta. La α -elica di riconoscimento legge la sequenza attraverso interazioni con le coppie di basi. Vicino ai siti di legame e della loro connessione si trova un segnale di localizzazione nucleare *NLS* attraverso cui viene localizzata nel nucleo.

Homeodomain

Il homeodomain è un comune dominio eucariote legante il DNA. È monomerico. La elica 3 si trova nella fessura principale e catene laterali interagiscono con le basi del DNA. Il braccio N terminale fa contatto con la fessura minore aumentando la specificità e stabilità del legame.

3.8.2 Zinc finger

Gli zinc finger *ZnF* sono domini che comprendono 30 amminoacidi con 1 α -elica e due β -filamenti intorno a uno ione di zinco centrale che interagisce con due cisteine e due istidine. 12 amminoacidi separano *Cys*₂*His*₂, 3 amminoacidi separano 2*Cys* da 2*His* affiancati da 4 amminoacidi. Le proteine spesso hanno diversi zinc finger, ognuno dei quali inserisce la sua α -elica nella fessura maggiore. Sono il dominio legante il DNA più comune nel genoma umano. I motivi *ZnF* sono spesso presenti in ripetizioni nei fattori di trascrizione. Ognuno dei quali aiuta affinità e specificità del legame. *TFII2* contiene 9 ripetizioni.

3.8.3 Leucine zipper

I domini leucine zipper *bZIP* comprende due lunghe α -eliche di 60 amminoacidi. Queste sono formate come coiled-coils, ovvero si arrotolano tra di loro. Le due eliche possiedono leucine idrofobiche sulla superficie interna che si uniscono insieme. Alla terminazione N le eliche si separano e si trovano nella fessura maggiore. Si legano a sequenze di DNA di 4bp invertite separate da 1 nucleotide. I monomeri dei leucine zipper possono combinarsi in diverse combinazioni di eterodimeri per attivare la trascrizione. Il legame può avvenire al di fuori del DNA o sul DNA quando avviene la dimerizzazione tra monomeri legati al DNA. La proteina *C/EBP AP-1* è un eterodimero formato dal proto-oncogene *c-Jun* e *c-Fos*. Regola l'espressione genica in risposta allo stress o infezioni virali o batteriche.

3.8.4 Helix-loop-helix

Le proteine helix-loop-helix *bHLH* hanno coiled-coils e un'elica con un residuo basico che si trova nella fessura principale. È tipicamente dimerico con ogni monomero contenente 2 α -eliche unite da un loop. L'elica N-terminale che lega il DNA è più grande rispetto a quella C-terminale. Potrebbe legarsi come monomero o formare eterodimeri con domini *HLH* o di altre proteine per funzione specifiche. La flessibilità del loop permette di impacchettamento con altre eliche e dimerizzazione attraverso interazioni idrofobiche. Si lega a sequenze di DNA palindromiche dette *E-box*. Attiva l'espressione genica *Myo-D* per lo sviluppo muscolare e *c-Myc* per il ciclo cellulare, la biogenesi dei ribosomi o il metabolismo.

3.8.5 Ribbon-helix-helix

Il dominio Ribbon-helix-helix *RHH* contiene due β -foglietti sul ribbon, ognuno che arriva da un monomero proteico coinvolto nella formazione del dimero e nelle interazioni specifiche con le basi del DNA nella fessura principale.

3.8.6 Interazioni con gli enhancer

Gli elementi di sequenze enhancer e le loro sequenze nella regione promotrice si legano ad attivatori per stimolare l'espressione genica. La loro varietà riflette la presenza di diverse proteine attivatrice. La loro combinazione può avere un forte effetto sulla trascrizione genica. Un elemento enhancer

3.9. MECCANISMI PER REGOLARE L'INIZIAZIONE DELLA TRASCRIZIONE NEI BATTERI

può contenere fino a 12 siti di legame per diversi fattori di trascrizione permettendo un controllo combinatorio della trascrizione genica. Specifici enhanceosomi si legano a *TAF*, al mediatore e alla RNA polimerasi II per regolare la trascrizione. Mutazioni in uno di questi attivatori o nella sequenza sottostante possono ridurre o aumentare la trascrizione.

Il sistema del lievito a due ibridi basato sui regolatori trascrizionali

Il fattore di trascrizione attivatore contiene un dominio legante il DNA e un sito attivante della RNA polimerasi. Un esempio sono i *Gal4* dei zinc finger. Attiva l'espressione di geni per la digestione del galattosio in *S. cerevisiae*. Il repressore *LexA* lega il DNA nella sequenza clonata di fronte a *lacZ*. Reprime l'espressione di *lacZ* quando il dominio legante il DNA di *LexA DBD* lega. L'attivazione di *lacZ* avviene quando il dominio di attivazione *Gal4* è legato a *LexA's DBD*, l'ibrido *Gal4-LexA*. Si nota come i domini di *Gal4* possono essere separati porta alla tecnica di screen del lievito a due-ibrido.

3.9 Meccanismi per regolare l'iniziazione della trascrizione nei batteri

I geni nei procarioti sono organizzati in operoni, gruppi di geni (fino a 12) le cui proteine codificate agiscono in un pathway. I geni sono organizzati in maniera back-to-back e trascritti da un promotore. Viene prodotto un lungo trascritto di mRNA detto mRNA poli-cistronico che non viene spliced: i ribosomi lo traducono in proteine separate. Anche i virus lo utilizzano. Negli eucarioti invece i geni sono singole unità trascrizionali tranne che nei nematodi. In *C. elegans* i trascritti poli-cistronici vengono spliced.

3.9.1 Operone

Nell'operone a monte dei geni si trovano un promotore e un regolatore. L'operatore è una sequenza regolatrice che agisce in cis a cui si possono legare proteine codificate da altri geni regolatori o sequenze regolatrici agenti in trans. Un modo per regolare la trascrizione è di prevenire il legame della RNA polimerasi al promotore o inibire il repressore che compie tale operazione.

3.9.2 Regolazione della trascrizione

Regolazione positiva

Si dice regolazione positiva quando un attivatore determina il destino dell'operone. Si dice trascrizione inducibile positiva quando un attivatore legato all'operone deve essere attivato per permettere la trascrizione. Si dice trascrizione reprimibile positiva quando tale attivatore viene disattivato e si separa dall'operone,

Regolazione negativa

Si dice regolazione negativa quando un repressore determina il destino dell'operone. Si dice trascrizione inducibile negativa quando il repressore attivo viene spento e si separa dal DNA e trascrizione negativa reprimibile quando il repressore inattivo viene attivato.

3.10 L'operone *lac* in E. coli

L'operone *lac* di E. coli è un esempio di un operone negativo inducibile. Pertanto sull'operone agisce un repressore che deve essere inibito per permettere l'inizio della trascrizione. In assenza di lattosio il gene *lacI* viene trascritto creando il repressore che si lega all'operatore *O*, pertanto i geni seguenti *lacZYA* sono molto poco espressi. In presenza di lattosio invece il repressore lega un induttore *allolattosio* che impedisce il legame con il sito *O*. La RNA polimerasi è pertanto libera di legarsi all'operone permettendo la trascrizione dei geni *lacZ*, *lacY* e *lacA*. Questi poi produrranno le proteine β -galattosidasi, permeasi e transacetilasi. La fine della trascrizione è ρ -indipendente. Il promotore appena a monte del sito *O* o sito *P* viene bloccato fisicamente da *LacI* che si lega alla sequenza palindromica dell'operatore *O*.

3.10.1 Genetica e analisi funzionale

Isolando mutanti di E. coli si separano quelli che producono β -galattosidasi in maniera costitutiva e si nota come ne esistono di due gruppi:

- Mutazioni localizzate a un gene lontano da *lacZ*: *lacI^c*.
- Mutazioni localizzate a una sequenza di DNA non codificante a monte di *lacZ*: operatore *O^c*.

I mutanti sono divisi in mutazioni che agiscono in cis e in trans. Quelle in cis avvengono in una sequenza non codificante sull'operone e non possono essere complementate esprimendo la sequenza wild type da un plasmide. Le altre in trans sono mutazioni sulla sequenza codificante una proteina e possono essere complementate esprimendo il gene wild type da un plasmide. Inoltre possono essere recessive o dominanti. Successivamente avviene l'analisi delle complementazioni delle mutazioni introducendo un operone *lac* wild-type sul plasmide *F'* detto operone esogeno. Il plasmide viene costruito in modo che faccia complementazione 1 : 1: contiene sia il gene *lacI* che l'intero operone.

Analisi dei mutanti complementari: crescita senza lattosio

Mutazioni a *lacI* Il mutante costitutivo del repressore che agisce in trans causa una produzione costitutiva del mRNA poli-cistronico da P. Con il plasmide la produzione viene repressa in quanto *F'* è un operone 1 : 1 complementare. La trascrizione endogena pertanto si blocca a causa di *LacI*.

Mutazioni al sito *O* Il mutante costitutivo del sito *O* che agisce in cis causa anch'esso una produzione costitutiva del mRNA poli-cistronico da P. L'aggiunta del plasmide non è in grado comunque di disattivare la sua produzione in quanto la proteina *LacI* non è in grado di attaccarsi al sito *O^c*. Si dice pertanto che questa mutazione è dominante.

3.10.2 Il repressore *LacI*

Il repressore *LacI* contiene un dominio helix-turn-helix per legarsi al DNA con un'elica di riconoscimento. Forma un tetramero quando si lega al sito *O* e contiene inoltre dei domini per il legame con l'allolattosio. L'inibizione allosterica di *LacI* causata dall'allolattosio è dovuta a cambi conformazionali che riducono l'affinità per il sito *O*.

Mutazioni di *LacI*

Si riconoscono 3 categorie di mutanti di *LacI*.

3.11. L'OPERONE TRIPTOFANO TRP IN E. COLI

lacI^d Questo mutante può di e tetramerizzare ma non può legarsi all'operatore. Con complementazione di *F'* *lacZYA* si nota come non c'è quasi possibilità di avere un di/tetramero completamente funzionale.

lacI⁻ Questo mutante non può di-tetramerizzare e non può legarsi all'operatore. Con complementazione non può legare a *O* e reprimere la trascrizione, mentre quello espresso dal plasmide può.

lacI^s Questo mutante non può legare l'allolattosio ma si lega fermamente all'operatore. Pertanto la complementazione non può competere per il legame e i geni saranno costitutivamente disattivati.

3.10.3 L'operone *lac* e induzione positiva

L'operone *lac* è anche un esempio di un operone inducibile positivamente: un attivatore agisce e deve essere attivato affinché avvenga la trascrizione. L'operone subisce regolazione positiva solo in presenza di lattosio e assenza di glucosio.

CAP e *cAMP*

Si nota come *E. coli* preferisca il glucosio come fonte di carbonio ed energia, pertanto in presenza di glucosio la trascrizione di *lacZYA* rimane repressa da *LacI* senza che il glucosio lo inibisca allostericamente. In assenza di glucosio *E. coli* cerca un altro carboidrato e se il lattosio è presente induce la trascrizione di *lacZYA*. Si nota come nonostante l'induzione dall'allolattosio l'intensità di trascrizione è bassa, pertanto *E. coli* deve aumentarla per consumare il lattosio efficientemente e velocemente in modo da sopravvivere. Viene pertanto aumentata l'affinità per la RNA polimerasi al promotore attraverso un attivatore trascrizionale *CAP* (catabolic gene activating protein) la cui attività è indotta da *cAMP*. Infatti quando i livelli di glucosio sono bassi e il lattosio è presente *ATP* viene convertito in *cAMP*. A questo punto si forma il complesso *CAP+cAMP* che si lega a monte del promotore nel proprio operatore. Questo aumenta l'affinità con RNA polimerasi legandosi al dominio *C* terminale *CTD* della subunità α dell'enzima.

3.11 L'operone triptofano *trp* in *E. coli*

L'operone *trp* è un operone negativo reprimibile: un repressore agisce e deve essere attivato per reprimere la trascrizione. L'operone *trpE-A* codifica l'enzima che converte l'acido corismico in triptofano. In *E. coli* la regolazione trascrizionale della sintesi del triptofano avviene a due fasi: all'iniziazione della trascrizione (regolazione reprimibile negativa) quando si hanno alti livelli di triptofano intracellulari. Nella seconda fase alla terminazione della trascrizione e della traduzione attraverso attenuazione: quando si trovano bassi o medi (la trascrizione ha iniziato ma deve essere fermata) livelli intracellulari di triptofano. Entrambi i meccanismi sono utilizzati per gli altri amminoacidi.

3.11.1 Regolazione all'iniziazione della trascrizione

In questa fase si trovano livelli di triptofano medio alti, pertanto non si rende necessaria sintesi attiva di triptofano: quello intracellulare induce l'attività del repressore inattivo *TrpR* per bloccare la propria sintesi. Funge pertanto da co-repressore. Un modo per regolare la trascrizione negativamente è impedire che la RNA polimerasi possa accedere al promotore come *TrpR*, una proteina helix-turn-helix. Il legame del repressore al DNA dipende dal livello di triptofano nella cellula. Quando i livelli

3.12. REGOLAZIONE DELLA TRASCRIZIONE DA PARTE DI RIBOSWITCHES DEL TRASCRITTO

sono bassi l'operone *trpE-A* viene trascritto per sintetizzarlo, mentre quando sono alti il legame di *TrpR* all'operatore blocca il legame della RNA polimerasi. *TrpR* può legarsi al DNA quando è legata a triptofano (indicatori che i livelli sono alti).

3.11.2 Regolazione della trascrizione attraverso attenuazione

Questa regolazione avviene quando i livelli di triptofano sono bassi. L'attenuazione è il controllo della trascrizione quando è già cominciata. Quando i livelli sono bassi non avviene la repressione da parte di *TrpR*, pertanto la trascrizione inizia. Quando il trascritto è lungo 161nt la trascrizione può continuare (livelli bassi di triptofano) o fermarsi (livelli medi). L'attenuazione controlla la sintesi dei geni batterici necessari per la sintesi di triptofano e degli altri amminoacidi sfruttando il piegamento di una sequenza di RNA leader in strutture secondarie come forcine.

Attenuazione mediata da *trpL*

Il mRNA *trp* possiede una sequenza leader *TrpL* vicino all'inizio con un terminatore intrinseco *poli-U*. Questa sequenza codifica parzialmente per un peptide di 14 amminoacidi che include due codoni per il triptofano. La sequenza leader contiene 4 blocchi di sequenza che possono formare accoppiamenti alternativi: 1 : 2 un debole stem-loop per il ribosoma, 3 : 4 un terminatore ρ -indipendente molto forte per la RNA-polimerasi, 2 : 3 un debole stem-loop per il ribosoma.

Livelli bassi di triptofano nella cellula In presenza di questi livelli bassi il ribosoma si blocca quando traduce il peptide leader: la regione 1 è occupata dal ribosoma, e la 2 e la 3 formano uno stem-loop impedendo il legame di 3-4 e la formazione del terminatore permettendo la continuazione della trascrizione, pertanto i geni nel mRNA *trpE-A* verranno trascritti e singolarmente tradotti.

Medi livelli di triptofano nella cellula In presenza di livelli medi il ribosoma procede occupando 1 e 2 permettendo la formazione del terminatore 3 : 4: la RNA polimerasi lascia il RNA leader e non trascrive oltre l'operone.

3.12 Regolazione della trascrizione da parte di riboswitches del trascritto

I riboswitches sono porzioni di un trascritto che legano piccole molecole che controllano la struttura secondaria del RNA regolando la trascrizione. Questi hanno due regioni:

- L'aptamero che lega un metabolita.
- La piattaforma di espressione che controlla la trascrizione.

Si nota come i riboswitches sono in grado di controllare sia la trascrizione che la traduzione.

3.12.1 Riboswitch di adenina di *B. subtilis*

Il riboswitch di *B. subtilis* regola la sintesi e il trasporto di adenina: l'espressione genica dipende se si forma un terminatore o un anti-terminatore.

Bassi livelli di adenina

Con bassi livelli di adenina le regioni 2 e 3 del RNA formano deboli loop e la trascrizione procede.

Alti livelli di adenina

Con alti livelli di adenina le regioni 3 e 4 formano un forte terminatore ρ -indipendente. Il RNA è rimosso dalla RNA polimerasi che si separa dal filamento di DNA stampo.

3.13 Regolazione dell'espressione genica del batteriofago λ in *E. coli*

La maggior parte dei fagi come *T1*, *T4*, *T7*, *SPO1* sono litici o virulenti: iniettano il loro DNA nel batterio senza che interagisca con il genoma batterico. Il DNA è trascritto dalla RNA polimerasi batterica. Gli mRNA poli-cistronici del fago sono tradotti da ribosomi batterici in proteine fagiche. Il DNA utilizza enzimi batterici che permettono l'assemblaggio della progenie fagica. Il rilascio dei fagi avviene dopo la rottura e lisi della cellula ospite. Questo ciclo è l'unico path di vita e riproduzione dei fagi litici. Il fago λ invece è più versatile: non uccide necessariamente la cellula ospite e può utilizzare due cammini di riproduzione: il cammino litico o virulento o quello lisogenico o temperato.

3.13.1 Il path lisogenico

Nel path lisogenico il DNA del fago è iniettato nel batterio e geni precoci sono trascritti e tradotti. Viene prodotta una proteina λ repressore e *cI* interrompe la trascrizione di tutti i propri geni in modo da non produrre progenie. Il DNA del fago si integra nel genoma dell'ospite replicando insieme ad esso. Il batterio con il DNA fagico integrato è detto lisogeno e il DNA fagico detto profago. I profaghi possono avere varie conseguenze sull'ospite: possono causare distruzione dei geni batterici, riarrangiamenti cromosomici attraverso ricombinazione omologa, causare l'espressione di nuove funzioni come tossine, proteine effettrici o regolatrici, possono indurre o silenziare geni batterici, alterazione delle capacità di creare biofilm o autoimmunità contro infezioni da fagi imparentati. In un qualunque momento il ciclo lisogenico può riconvertirsi in quello litico, specialmente in condizioni di stress per il batterio.

3.13.2 I due cicli vitali del fago λ

Il repressore λ mantiene il virus in uno stato dormiente e lisogenico, una condizione stabile che può esistere per un numero di generazioni. Fattori di stress come mutageni, chimici o fisici possono rompere lo stato lisogenico ed escindere il genoma λ e attivare il ciclo litico che in 20 minuti causa il rilascio della nuova progenie fagica.

3.13.3 Il genoma del fago λ e sue interazioni

Il dDNA del fago λ è lungo 48 502bp e codifica per 50 geni. Si presenta lineare nella testa di λ ma circolarizza dopo l'iniezione chiudendosi attraverso estremità di 12 basi complementari e coesive dette *overhangs*. Una regione di controllo o immunità codifica per i geni che determinano lo stato litico o lisogenico e comprende il repressore λ *cl*. Si trovano operoni precoci destri o sinistri che codificano

3.13. REGOLAZIONE DELL'ESPRESSIONE GENICA DEL BATTERIOFAGO λ IN *E. COLI*

le proteine per la ricombinazione del DNA di λ , la sua escissione e replicazione. Si trovano poi le proteine richieste per la lisi batterica.

Il genoma di λ circolarizza in *E. coli*

Dopo l'iniezione il genoma circolarizza e segue il ciclo litico. Il genoma si deve circolarizzare per integrarsi nel genoma ospite e anche dopo l'escissione da esso. La circolarizzazione lo protegge dalla degradazione da parte del batterio attraverso esonucleasi, permette l'integrazione attraverso ricombinazione omologa del genoma λ in quanto la sequenza *attP* si trova nel mezzo del genoma. La circolarizzazione permette anche la trascrizione dei geni terminali richiesti per il ciclo litico e la replicazione del genoma λ basata sul rolling-circle.

Ricombinazione omologa sito-specifica

I siti della ricombinazione omologa sito-specifica sono corti tra i 20 e i 200nt ma altamente specifici. Comprendono due motivi: una ripetizione invertita parziale simmetrica a cui si lega la ricombinasi e una sequenza di crossover centrale dove avviene la ricombinazione. Entrambi i siti sono tipicamente identici, con alcune eccezioni: nel caso di λ i siti sono *attP POP'* e su *E. coli* *attB BOB'*. La ricombinasi agisce come un dimero e taglia un ssDNA, scambia le due eliche coinvolte e lega i filamenti. Nel caso di λ ed *E. coli* i siti *attP* e *attB* vengono coinvolti e dopo la ricombinazione si formano due nuovi siti *attL* e *attR* che affiancano il genoma λ integrato a sinistra e destra. In questo caso la ricombinasi è la λ integrasi e utilizza proteine accessorie:

- *IHF*: integration host factor, utilizzata per l'integrazione e l'escissione in quanto piega il DNA per unire insieme i siti di legame.
- *FIS*: factor for inversion stimulation, utilizzata per l'escissione.
- *XIS*: escissionasi., utilizzata per l'escissione.

Per rompere i filamenti rimanenti e completare il processo interviene una speciale topoisomerasi di tipo I. Si nota inoltre come dopo la ricombinazione avviene una permutazione dell'ordine dei geni di λ : nel profago si trova dopo il sito *attL* i geni a destra di *POP'*, seguiti dai geni alla sua sinistra e poi *attR*.

3.13.4 Il ciclo litico del fago λ

Dopo l'iniezione e la circolarizzazione del DNA di λ il ciclo litico inizia automaticamente, ma può essere fermato e convertito nel ciclo lisogenico. La regolazione di questo ciclo è fatta attraverso anti-terminazione. I geni λ richiesti sono espressi in tre passi temporizzati attraverso la trascrizione e si dividono in tre gruppi disposti sequenzialmente sul DNA e la loro espressione temporale è controllata regolando la trascrizione attraverso anti-terminazione a livello del trascritto o del promotore.

Controllo temporale dell'espressione dei geni λ

Passo 1 La RNA polimerasi di *E. coli* reclutata da σ^{70} trascrive 2 geni precoci immediati: il gene *N*, trascritto dal promotore sinistrorso P_L sul filamento superiore e il gene *cro*, trascritto dal promotore destro P_R sul filamento inferiore. Quando la RNA polimerasi raggiunge la fine di *N* e *cro* incontra i loro terminatori t_L e t_R ρ -indipendenti e si ferma a fronte dei geni precoci ritardati. I trascritti sono poi tradotti dai ribosomi di *E. coli* creando le proteine *N* e *Cro* cruciali per la continuazione del ciclo litico:

3.13. REGOLAZIONE DELL'ESPRESSIONE GENICA DEL BATTERIOFAGO λ IN *E. COLI*

- *Cro*: è un repressore che blocca la trascrizione del gene repressore λ *cI* che deve essere fatto per iniziare e mantenere il ciclo lisogenico.
- *N*: una proteina anti-terminatrice che aiuta la RNA polimerasi a ignorare i terminatori dei geni *cro*, *N* e *P*: la sua attività porta alla trascrizione da P_L e P_R dei geni precoci ritardati: i trascritti *N* e *cro* si estendono e diventano poli-cistronici.

Quando avviene la trascrizione mediata da *N* inizia la fase precoce ritardata. Permette inoltre la trascrizione attraverso t_1 per i geni *xis* e *int* dopo l'espressione da P_1 .

Passo 2 La RNA polimerasi trascrive i geni precoci ritardati come due RNA poli-cistronici da P_L e P_R includendo i costroni dei geni *O* e *P* che codificano proteine necessarie per iniziare la replicazione del DNA del fago λ . La proteina *O* è un analogo della proteina iniziatrice *DnaA* e forma un complesso all'*Ori* che si trova nella sequenza codificante di *O*. La proteina *P* recluta *DnaB* all'origine per iniziare la replicazione in un meccanismo simile a quello che avviene a *oriC*. La replicazione precoce segue il modello Θ e quella tardiva il meccanismo del rolling-circle. Viene trascritto il gene *cII* che può iniziare il passaggio al ciclo lisogenico se necessario. Inoltre la proteina *Q* è un anti-terminatore che aiuta la RNA polimerasi a ignorare il terminatore t_R che permette l'espressione dei geni tardivi da P_R .

Passo 3 I geni tardivi vengono trascritti in direzione destra dal promotore tardivo P_R che si trova a valle di *Q*. La trascrizione da P_R si ferma dopo 194bp a causa del terminatore ρ -indipendente t_R a meno che l'anti-terminatore *Q* intervenga. I geni tardivi codificano enzimi litici della parete cellulare *S* e *R* che causano la lisi di *E. coli*. L'enzima di restrizione *A* per tagliare il DNA λ circolare ai siti *cos* e le proteine *W* e *J* che formano la testa, il corpo e le gambe del fago.

Regolazione della trascrizione di *N* e *Q*

N e *Q* regolano la trascrizione attraverso due meccanismi diversi.

Anti-terminazione mediata da *N* L'anti-terminazione mediata da *N* avviene a livello della RNA polimerasi e del trascritto. Inizialmente nel passo precoce non è ancora stata sintetizzata la proteina *N* la RNA polimerasi si lega al sito $O_L P_L$ e trascrive il mRNA per la proteina *N*. Nel passo mediano in presenza della proteina *N* vengono reclutati su di essa *NusA*, *NusB*, *NusG* e *S10* (subunità della 30S piccola subunità ribosomiale). Si trova all'inizio dell'mRNA che sta per essere trascritto una sequenza *nutL* (*N* utilization site) a cui si lega *N*. *NusA* e *NusG* mediano l'interazione con la RNA polimerasi di *N*. *nutL* è formata da una box *A* e una *B*. La box *B* forma una ripetizione invertita che forma uno stem-loop, mentre la box *A* è il sito di legame per *NusB*. Si nota come il processo di anti-terminazione è processivo: tutti i fattori di anti-terminazione *NusA*, *G*, *B* e *S10* rimangono associati con la RNA polimerasi mentre si muove e trascrive il DNA permettendole di passare attraverso t_L e t_R . La stessa attività di anti-terminazione avviene ai terminatori di *cro* e *P* dopo la trascrizione di P_R e di *int* dopo P_1 . Siti *nutL* sono presenti nelle sequenze di *cro*, *P* e *int*.

Anti-terminazione mediata da *Q* L'anti-terminazione mediata da *Q* avviene a livello del promotore P_R . *Q* si lega alla sequenza *qut/QBE* in P_R e non a una sequenza interna al mRNA.

Assenza di Q In assenza di Q la RNA polimerasi si lega a P_R e si ferma per diversi minuti al sito di pausa, uno pseudo-sito a -10 dopo aver trascritto il segnale di pausa uno pseudo-sito a -35 . Dopo aver lasciato il sito di pausa la RNA polimerasi trascrive fino al terminatore ρ -dipendente t_R dove si ferma.

Presenza di Q In presenza di Q la RNA polimerasi si lega a P_R e si ferma per diversi minuti al sito di pausa. Q si lega alla sequenza *qut/QBE* vicino alla RNA polimerasi in pausa. *qut* legato a Q sposta σ^{70} della RNA polimerasi e lo forza a legarsi alle box pseudo -35 e -10 . Pertanto Q si lega alla RNA polimerasi e altera la sua struttura per permetterle di procedere con la trascrizione e di ignorare il terminatore t_R producendo così il mRNA poli-cistronico dei geni tardivi.

3.13.5 Passaggio al ciclo lisogenico

Affinchè il ciclo litico si trasformi in lisogenico per il fago λ :

- Il repressore λ deve essere espresso. Questa proteina può agire come attivatore o repressore: reprime la trascrizione di N e cro da P_L e P_R , inoltre auto-attiva la trascrizione del proprio gene cI da P_{RM} .
- Il repressore λ deve fermare l'espressione genica di cro : la trascrizione di cI da P_{RM} a O_{R3} è bloccata da Cro .

Ciclo lisogeno

Il ciclo lisogeno è uno stato che si deve stabilire e mantenere ed entrambi i passaggi richiedono la proteina λ repressore. cI è trascritto da 2 promotori: P_{RE} e P_{RM} e agiscono uno dopo l'altro:

- P_{RE} promoter for repressor establishment: si trova a destra di P_R sul filamento superiore tra i geni cro e cII . Direzione la trascrizione sinistrorsa attraverso cro e cI : vengono create le prime piccole quantità del repressore λ .
- P_{RM} promoter for repressor maintenance: si trova a sinistra di P_R sul filamento superiore e la trascrizione da P_{RM} avviene dopo lo stabilimento del repressore e durante il ciclo lisogeno per garantire una continua fonte di repressore λ per mantenere il ciclo lisogeno.

Passo 1: stabilimento della lisogenia Dopo l'attività di anti-terminatore di N la trascrizione da P_R produce un mRNA *cro-cII-O-P-Q* nella metà del ciclo litico. P_{RE} produce un mRNA che è un trascritto di senso per cI ma antisenso per il RNA che si sovrappone con *cro-cII-O-P-Q*: il singolo filamento cI di questo mRNA è tradotto e forma il repressore λ . Il cistrone cro non può essere tradotto in quanto il mRNA da P_{RE} localmente lega il mRNA *cro-cII-O-P-Q*. La parte cII rimane a filamento singolo ed è tradotta nella proteina CII che stimola la trascrizione dal promotore sinistrorso P_{AQ} in Q e produce un RNA anti-senso per Q che impedisce la traduzione per Q bloccando il ciclo litico e favorendo quello lisogeno. Inoltre stimola il legame della RNA polimerasi a P_{RE} aumentando la quantità del repressore λ che reprime P_L e P_R fermando il ciclo litico.

Attivatore della trascrizione cII cII agisce come un omotetramero e si lega al DNA attraverso un dominio helix-turn-helix. P_{RE} ha box a -35 e -10 con debole somiglianza a entrambe le sequenze di consenso riconosciute da σ^{70} . Non può essere trascritto dalla RNA polimerasi da sola ma cII deve legarsi all'olocomplesso della RNA polimerasi. Ora riesce a legarsi a P_{RE} e trascrive cI .

3.13. REGOLAZIONE DELL'ESPRESSIONE GENICA DEL BATTERIOFAGO λ IN *E. COLI*

cII stimola il legame della RNA polimerasi anche con P_I , il promotore per i geni *xis* e *int*, proteine richieste per la lisogenia e l'integrazione ed escissione del DNA λ nel genoma di *E. coli*.

Passo 2: mantenimento della lisogenia Il mantenimento della lisogenia avviene grazie all'autoregolazione della trascrizione di *cI* da P_{RM} da parte del repressore λ . Una volta che una piccola quantità di repressore λ è sintetizzata da P_{RE} questo si lega come omodimero all'operatore O_R e O_L che fiancheggiano P_{RM} da ambo le parti. Il repressore forma un tetramero, la sua forma attiva. Il legame del repressore λ a O_R e O_L ha due conseguenze: il repressore λ stimola la propria trascrizione legandosi a bassi livelli alla RNA polimerasi a O_R2 attivandolo a P_{RM} , inoltre reprime la trascrizione da P_L e P_R fermando il ciclo litico. Fermare la produzione di *Cro* da P_R permette il legame della RNA polimerasi con O_R3 a P_{RM} .

O_R e O_L Si nota come O_R e O_L sono formati da 3 sezioni che possono legare il repressore λ . O_R controlla la trascrizione sinistrorsa di *cI* da P_{RM} e quella destra di *cro* da P_R . Ha diverse affinità ordinate con 1, 2 e 3. Avviene un legame cooperativo di λ con O_R1 e O_R2 ma non con O_R3 .

Struttura e legame al DNA della proteina repressore λ *cI* *cI* è un selfish regulator: auto-attiva la propria espressione reprimendo tutti gli altri geni P_L e P_R sul genoma di λ . Possiede un dominio C-terminale con i domini per dimerizzazione e tetramerizzazione. Questo è legato da un linker flessibile a un dominio N-terminale che contiene una regione regolatoria che si lega al dominio σ^4 di σ^{70} e una regione che si lega alla sequenza operatrice come un omodimero con motivo helix-turn-helix. La tetramerizzazione permette il legame al DNA cooperativo.

Legame del repressore λ con O_R1-3 Il repressore λ si lega a O_R1-3 con diverse affinità. Quando si lega a O_R1 e O_R2 non avviene trascrizione da P_R e non viene trascritto il RNA *cro* e *cII*, con il secondo richiesto per la trascrizione da P_{RE} . La scomparsa di *cII* non è un problema in quanto la lisogenia è stata già stabilita e una piccola quantità di repressori λ è sufficiente per mantenerla. Questa viene fornita fino a che O_R3 non è occupato da alti livelli di repressore λ permettendo la trascrizione di *cI* da P_{RM} e il repressore λ legato a O_R1 e O_R2 blocca la trascrizione di *cro* impedendo il legame della RNA polimerasi con P_R in quanto *Cro* reprime P_{RM} . Si nota come se il repressore λ aumentasse tutta la trascrizione di *cI* verrebbe bloccata che diminuirebbe i livelli di repressore λ che si dissocerebbe da O_R3 permettendo il riinizio della trascrizione di *cI*, un sistema di autoregolazione che impedisce che i livelli di repressore λ diventino troppo alti.

Coinvolgimento di O_L La formazione di loop tra O_R e O_L da parte dell'ottamero del repressore λ causa la repressione di P_L e P_R . Con alti livelli si lega anche a O_R3 e O_L3 causando la repressione di P_L , P_R e P_{RM} .

3.13.6 Determinare il destino dell'infezione

Si nota come anche nell'ambiente locale di una placca delle cellule infettate soffrono il ciclo litico mentre altre lisogenizzano, crescono e dividono. Si nota inoltre come se un fago λ infetta una cellula lisogenizzata la proteina repressore λ è già presente e legherà al DNA circolarizzato del fago reprimendo P_L e P_R : un lisogeno è pertanto immune a super-infezioni e lisi da un fago λ in entrata con la stessa regione di controllo/immunità del profago. Si nota come gli *E. coli* in ogni cellula sono identici geneticamente, così come i fagi λ utilizzati per l'infezione, pertanto la scelta del ciclo non è genetica ma risulta dall'equilibrio tra i livelli intracellulari del repressore λ e di *Cro*: se il primo è più presente avviene il cammino lisogeno, altrimenti quello litico.

Cross-regolazione dell'espressione tra il repressore λ (*cI*) e *Cro* (*cro*)

Se il gene *cI* produce abbastanza repressore λ questo legherà a O_R e O_L di P_R e P_L impedendo la trascrizione dei geni precoci di *N* e *Cro*. L'assenza del loro prodotto impedisce l'espressione degli altri geni, non vengono fatti altri fagi e non avviene la lisi. Se il gene *cro* produce abbastanza proteine *Cro* si legherà a O_{R3} e reprime P_{RM} e la trascrizione di *cI* e non avverrà lisogenia. L'abilità di *Cro* di bloccare la trascrizione di *cI* sta nella sua affinità con O_R e O_L : si lega infatti ad entrambe e l'affinità per le parti è di ordine inverso rispetto al repressore λ . *Cro* si lega come un omodimero attraverso i domini helix-turn-helix alle regioni O_R e la sua incapacità di tetramerizzare impedisce legami cooperativi.

λ e *Cro* regolano la regione di controllo in maniera mutualmente esclusiva Si nota come *Cro* agisce sempre come repressore e lega O_{R3} fermando la trascrizione di *cI* da P_{RM} . Quando i livelli di *Cro* aumentano copre sia O_R che O_L impedendo la trascrizione di tutti i geni da P_R e P_L come *cII* e *cIII* senza i quali P_{RE} non può funzionare. Non viene pertanto prodotto il repressore λ e inizia il ciclo litico. Il fatto che *Cro* disattivi la trascrizione precoce è richiesto per la crescita litica in quanto la continua produzione di proteine precoci ritardate nell'infezione tardiva abortisce il ciclo litico.

La concentrazione della proteina *cII* determina se vince il repressore λ o *Cro* Nel passaggio intermedio la proteina *N* anti-terminatrice causa la formazione di un lungo trascritto da P_R che contiene sia *cro* che *cII*. Il fattore fondamentale è la concentrazione della proteina *cII* nella cellula: se è alto il rapporto di concentrazione tra *cII* e *Cro* la lisogenia segue immediatamente il passaggio intermedio litico. Infatti si nota come *cII* attiva P_{RE} producendo RNA *cI* senso stabilendo il programma lisogeno e RNA *cro* anti senso impedendo la traduzione del RNA *cro* nel mRNA poli-cistronico.

La proteina *cIII* determina la concentrazione di *cII* nella cellula Il prodotto del gene intermedio *cIII* si ottiene dopo l'anti-terminazione del passo precoce della proteina *N* a t_L . *cIII* lega *cII* formando un complesso per proteggere il primo dalla distruzione da parte di processi intracellulari. Alti livelli di proteasi *FtsH* infatti possono distruggere *cII* forzando il ciclo litico. Alti livelli della proteasi si trovano in buone condizioni di crescita, pertanto un medium ricco causa il cammino litico, mentre quello povero quello lisogeno. Questo avviene in quanto il pathway litico richiede considerevole energia e risorse dalla cellula per replicare il DNA λ , mentre quello lisogeno richiede unicamente la sintesi del repressore λ .

3.13.7 Attivazione del profago e ciclo litico nei batteri lisogeni

Elementi chimici mutageni e *UV* causano l'escissione del profago dal genoma del batterio lisogeno. Questo viene causato dall'azione del sistema di risposta del danno al DNA che coinvolge più di 40 proteine come *RecA* la cui espressione è sotto il controllo negativo del repressore *LexA*. *RecA* è una DNA ricombinasi, ma danni ambientali la attivano come co-proteasi. Questa si lega e attiva l'attività di autorottura del repressore λ che si taglia a metà ed è rilasciato dalle sequenze operatrici. Inizia la trascrizione da P_R e P_L . Uno dei primi geni trascritti è *cro* il cui prodotto impedisce trascrizione ulteriore di *cI*. Viene rotta la lisogenia e inizia il ciclo litico.

Controllo dell'integrazione e dell'escissione del DNA λ

Si nota come per l'integrazione all'inizio del ciclo lisogeno la proteina *cII* induce la trascrizione di *int* ma non di *xis* causando l'integrazione insieme a *int* e *IHF*. Affinchè avvenga l'escissione si deve produrre *int*, *xis* e *IHF*. *xis* viene controllata da una sequenza *sib* che durante il ciclo litico forma una forcina che viene distrutta dalla *RNAasi III*. In caso di irradiazione da *UV* il repressore λ si autodistrugge dopo il danno al DNA e in quanto il genoma è integrato *sib* si trova lontana dai geni *int-xis* e diventa parte del trascritto di RNA prodotto da P_L .

3.14 Regolazione della trascrizione da sistemi di trasduzione del segnale a due componenti

La risposta a segnali esterni coinvolge spesso la fosforilazione di una proteina recettore che porta a una cascata di segnalazione come il pathway di trasduzione del segnale a due componenti. I sistemi a due componenti possiedono un recettore chinasi sensoriale e un partner di risposta regolatore di trascrizione. Una istidina nel dominio della chinasi si auto-fosforila utilizzando *ATP* sulla recezione del segnale. Il gruppo fosforile è poi passato a un residuo di acido aspartico in un regolatore di risposta che viene attivato e procede ad attivare o reprimere il proprio gene obiettivo con cui forma un regulon. Si nota come questo processo porta a un adattamento della cellula allo stimolo.

3.15 Regolazione dell'iniziazione della trascrizione ed allungamento negli eucarioti

3.15.1 Regolazione dell'iniziazione della trascrizione

Gli eucarioti tipicamente regolano l'iniziazione attraverso proteine leganti il DNA che reclutano co-attivatori o co-repressori. Un esempio è *ELK1* che recluta il complesso mediatore necessario per la trascrizione della RNA polimerasi II. Un mitogeno o fattore di crescita è un segnale extracellulare (proteina o ormone) che promuove la divisione cellulare e l'entrata in mitosi temporizzata. In assenza di un mitogeno *ELK1* si lega al fattore di risposta serum *SRF* e non attiva la trascrizione. Il legame del mitogeno sulla superficie attiva chinasi che fosforilano *ELK1* che a sua volta recluta il complesso mediatore promuovendo la trascrizione del gene.

Metabolismo del galattosio

Il metabolismo del galattosio nel lievito è regolato in maniera più complessa a livello di iniziazione della trascrizione: il fattore di trascrizione *Gal4* regola l'espressione dei geni per il metabolismo dello zucchero: attiva la trascrizione legandosi alla sequenza *UAS*. La sua attività è regolata dal co-repressore *Gal80* e dal co-attivatore *Gal3*, entrambi i quali rispondono al galattosio nella cellula.

Assenza di galattosio In assenza di galattosio *Gal80* si lega a *Gal4* prevenendo la trascrizione degli enzimi metabolizzatori del galattosio.

Presenza di galattosio In presenza di galattosio *Gal3* si lega a *Gal80* per impedire che questo si leghi a *Gal4* che recluterà il complesso *SAGA* acetilatore degli istoni e il mediatore attivando la trascrizione.

3.16. IL RUOLO DELLE CASCATE DI SEGNALAZIONE NELLA REGOLAZIONE DELLA TRASCRIZIONE

Risposta ai livelli di azoto e carbonio

Nel lievito *Ume6* risponde alla disponibilità di azoto e carbonio. Può attivare o reprimere la trascrizione genica.

In presenza di azoto e carbonio Quando *N* e *C* sono presenti non si rende necessario sintetizzare, pertanto *Ume6* si lega al DNA e recluta co-repressori:

- *Sin3* recluta il complesso *Rpd3*
- *Rpd3* è una deacetilasi istonica.
- *Isw2* è un enzima rimodellatore del nucleosoma che aiuta a stabilire il pattern cromatinico alterato.

In assenza di azoto e carbonio Quando *N* e *C* non sono presenti *Ume6* viene fosforilata dalla chinasi *Rim15*, successivamente *Sin3* e *Rpd3* si dissociano da *Ume6-P* e il co-attivatore *Ime1* è reclutato e recluta a sua volta il complesso di acetil-trasferasi istonico.

3.15.2 Regolazione dell'allungamento della trascrizione

Si consideri come esempio l'allungamento della trascrizione di *hsp70* in *Drosophila*: la proteina *Hsp70* è coinvolta nella protezione della cellula durante l'esposizione a temperature elevate.

In assenza di heat shock A temperature normali 25°C il fattore *GAGA* si lega a monte di *hsp70* e recluta il rimodellatore del nucleosoma *NURF* che mantiene il promotore libero da nucleosomi e permette il legame della RNA polimerasi. La trascrizione inizia ma si blocca immediatamente in quanto la RNA polimerasi non è abbastanza fosforilata "proximal pausing". Nonostante questo la RNA polimerasi è presente e pronta ad allungare il trascritto *hsp70* in risposta a un possibile heat shock.

In presenza di un heat shock Quando la temperatura si alza 37°C il fattore heat-shock *Hsf* si attiva: si trimerizza e lega all'elemento heat shock *HSE* e successivamente interagisce con il mediatore e recluta una chinasi che fosforila il *CDT*, permettendo alla RNA polimerasi di riprendere la trascrizione di *hsp70*.

3.16 Il ruolo delle cascate di segnalazione nella regolazione della trascrizione

I pattern di espressione genica cambiano costantemente in risposta alle necessità di sviluppo e indizi ambientali. Le cascate di segnali sono il risultato di cambiamenti di condizioni, con cambi trascrizionali come una delle conseguenze finali. Le proteine recettori nucleari rispondono specificatamente a effettori come gli ormoni. Queste possiedono un dominio che lega tali ligandi e uno che lega il DNA. Il legame del ligando induce un cambio conformazionale nel recettore, permettendogli di reclutare diversi co-repressori o co-attivatori. Possono essere coinvolte reti più complesse: per esempio un recettore nucleare fuori dal nucleo potrebbe entrarvi solo quando legato a un ligando.

3.16.1 Esempio di risposta immunitaria

Un esempio di cascata a passaggi multipli è il fattore di trascrizione *NF- κ B*, un eterodimero di *p50* e *p65* importante nella risposta immunitaria dei mammiferi.

Cellule non infette Nelle cellule non infette *NF- κ B* è mantenuto inattivo nel citoplasma dalla chinasi *I- κ B* che si lega al segnale di localizzazione nucleare che imporrebbe il trasporto di *NF- κ B* nel nucleo.

Cellule infette Dopo l'infezione della cellula questa causa una cascata che attiva attraverso fosforilazioni attraverso la chinasi *IKK* la chinasi *I- κ B*. Questa una volta fosforilata viene ubiquitinata dalla *E3* ubiquitina ligasi ed è spedita per la degradazione dal proteasoma. La distruzione di *I- κ B* espone il segnale di localizzazione nucleare di *NF- κ B* che può spostarsi nel nucleo e attivare la trascrizione.

3.17 Silenziamento genico attraverso imprinting genomico

Il silenziamento trascrizionale attraverso de-acetilazione o metilazione istonica o metilazione del DNA avviene in *IGF2* e *H19*. Il primo produce un fattore di crescita necessario per l'embrione in sviluppo e il secondo crea un RNA non codificante che potrebbe svolgere un ruolo fondamentale nel cancro, entrambi si trovano sul cromosoma 11. Si nota come *IGF2* è espresso solo sul cromosoma paterno e *H19* solo su quello materno. Questo è dovuto alla metilazione del DNA alla regione di controllo isolatoria *ICR*. La proteina *CTCF* (*CCCTC*-binding factor) può legarsi unicamente al *ICR* solo quando questo non è metilato come nel cromosoma materno. Pertanto:

- Nel cromosoma materno quando *CTCF* è legata al *ICR* le proteine legate al enhancer possono stimolare la trascrizione a *H19* ma sono bloccate dallo stimolare l'espressione di *IGF2*.
- Nel cromosoma paterno *CTCF* non può legarsi pertanto *IGF2* non è bloccata e può essere trascritta. *H19* è metilato e inibisce la sua trascrizione.

La trascrizione è mediata da loop del DNA e determinata da attivatori legati al enhancer che permettono o bloccano l'accesso. Un fallimento nell'imprinting corretto può portare alla sindrome di Beckwith-Wiedemann che causa bambini più grandi del normale e più suscettibili al cancro.

3.17.1 DNA metilato

A differenza di *CTCF* il DNA metilato può reclutare specifiche proteine: i cromodomini leganti i metile si legano specificatamente alle metil-citosine. Un esempio è *MeCP2* che lega il DNA metilato e recluta il repressore trascrizionale *Brm* (Brahm) e *Sin3A*, un co-repressore trascrizionale che è parte di un complesso di deacetilasi istonico. *meCP2* reprime un numero di geni umani e mutazioni in questa proteina causano la sindrome di Rett, come se dovesse essere eliminato dal cromosoma *X*.

3.17.2 Silenziamento trascrizionale ai mating loci fissione del lievito

Il silenziamento trascrizionale ai mating loci della fissione del lievito è mediata da interferenza a RNA che induce uno stato eterocromatico: piccole regioni del DNA peri-centromerico sono trascritte in entrambe le direzioni e possono accoppiarsi formando molecole di RNA a doppio filamento che sono

3.17. SILENZIAMENTO GENICO ATTRAVERSO IMPRINTING GENOMICO

rotte in frammenti di 21bp che sono incorporati nel complesso proteico *RITS*, o RNA-induced initiation of transcriptional silencing. Il complesso *RITS* recluta la deacetilasi istonica *Clr3* e la metilasi istonica *Clr4* e si lega alle code istoniche locali. Questi enzimi creano eterocromatina in *S. pombe* attraverso ipo-acetilazione degli istoni metilazione della lisina 9 di *H3*. La coda *H3* metilata recluta *Swi6* che porta a più metilazione e deacetilazione, mantenendo e diffondendo l'eterocromatina.

Capitolo 4

Processamento dell'RNA

4.1 Panoramica del processamento del RNA

Gli RNA prodotti sono spesso non funzionali. Questi RNA precursori *pre-RNA* devono essere modificati per diventare funzionali. Questo avviene durante il processamento del RNA o maturazione. Il processamento avviene nel nucleo per impedire che i pre-RNA siano tradotti nel citoplasma. Dopo il processamento sono trasportati nel citoplasma per traduzione seguente da parte dei ribosomi. Il processamento avviene per tre ragioni:

- Regolazione dell'attività genica.
- Diversità: molti RNA diversi possono essere prodotti da un gene attraverso splicing alternativo rimuovendo diverse combinazioni di introni.
- Controllo della qualità: mRNA difettivi sono individuati e degradati.

4.1.1 Modifiche al RNA

Le modifiche al RNA coinvolgono un grande numero di complessi molecolari e molti di essi contengono sia proteine che RNA e sono ribonucleoproteine *RNP*. Il RNA negli *RNP* può avere ruolo strutturale ma anche attività catalitiche come i ribozimi e i ribosomi. Alcuni *RNP* contengono guide a RNA che si accoppiano con le basi dei pre-RNA e li guidano alla sequenza corretta per il processamento del pre-RNA obiettivo.

4.2 Processamento di rRNA e di tRNA

4.2.1 Procarioti

rRNA

Il rRNA nei procarioti è prodotto come lunghi pre-rRNA 30S. Questi sono rotti in un numero di rRNA da endonucleasi. Gli rRNA rotti sono successivamente raffinati da esonucleasi alle estremità per produrre gli rRNA finali. Questi non sono tradotti ma diventano il backbone strutturale delle subunità grande e piccola dei ribosomi.

4.2. PROCESSAMENTO DI RRNA E DI TRNA

E. coli In *E. coli* il pre-rRNA 30S forma degli stem loop in corrispondenza del rRNA 16S, del 23S e una struttura a forcina con due stem-loop in corrispondenza del 5S. Le RNAasi *RNAasi III*, *RNAasi M16* e da *RNAasi M23* rilasciano il 16S e il 23S, mentre la *RNAasi E* rilascia il 5S. La sequenza che viene processata contiene anche dei tRNA interni che vengono elaborati diversamente e specificatamente.

Ribonucleasi Le ribonucleasi rompono o raffinano gli RNA in pezzi più piccoli.

Esonucleasi Le esonucleasi rimuovono nucleotidi dalle terminazioni di un trascritto, non sono specifiche alla sequenza e la maggior parte agiscono in direzione 3'-5'. La maggior parte sono processive. La *PNPasi* e l'esosoma sono esonucleasi 3'-5' di *E. coli*. *Xrn1* e *Exol*, sempre di *E. coli* sono esonucleasi 5'-3' e la seconda è processiva.

Endonucleasi Le endonucleasi rompono il RNA nel filamento. Alcune rompono dsRNA come *RNAasi III*, mentre altre ssRNA come *RNAasi P* o *tRNAasi Z*. Possono rompere a 3' o a 5'. La *RNAasi P* possiede componenti a RNA e proteine, in quella batterica solo la parte a RNA può rompere il RNA, mentre la parte proteica aumenta l'attività e l'intervallo di substrati. In quella eucariotica, di archea e mitocondriale invece il componente a RNA da solo non può tagliare il RNA ma è essenziale per la funzione.

tRNA

I tRNA hanno una struttura variabile composta da un sito accettore *ACC*, un braccio *TΨC* con pseudo-iridina, un braccio variabile, un braccio dell'*AC* e un braccio *D* contenente di-idro-uridina. Si nota la presenza di molte basi modificate. In *E. coli* vengono prodotti come un pre-tRNA policistronico 30S che viene processato. A 5' interviene il ribozima *RNAasi P* che taglia il pre-tRNA, mentre una endonucleasi 3' lo separa e esonucleasi di tRNA invece affinano la terminazione 3' fino alla sequenza di stop prima di *CCA*. Il caricamento degli amminoacidi avviene sul *CCA-3'OH* grazie a un aminoacil-tRNA sintetasi che idrolizza *ATP* per creare un legame estere tra la terminazione e l'amminoacido.

Degenerazione del codice genetico Si nota come in *E. coli* per 20 amminoacidi si trovano 64 codoni e 43 tRNA. Pertanto esistono più codoni per 1 amminoacido che può usare multipli tRNA per la sua inclusione in una catena peptidica in base al suo codone.

La degenerazione diminuisce gli effetti deleteri delle mutazioni Una mutazione può rimanere silente o non causare uno stop nella sintesi della proteina.

Gerarchia dei codoni Si trova una gerarchia di importanza tra i diversi codoni per la codifica di uno stesso amminoacido: certi codoni sono usati in proteine a bassa priorità che non sono sintetizzate in caso l'amminoacido sia poco presente. Altri sono utilizzati in proteine ad alta priorità fatte senza tener conto della disponibilità di amminoacido. Pertanto diverse sequenze di codoni non hanno la stessa importanza creando il codon bias. In questo modo la cellula conosce quali proteine devono essere fatte e quali possono essere ignorate quando la loro disponibilità è bassa. Si nota come l'utilizzo dei codoni influenza il tasso di traduzione del RNA e la produzione di proteine: più sono frequentemente usati più veloce è la traduzione.

4.2.2 Eucarioti

rRNA

Negli umani nel nucleolo si trova il rDNA array formato da un non-transcribed spacer *NTS* e il gene per il rRNA. La RNA polimerasi I sintetizza il 47S pre-rRNA che contiene tre rRNA con a 5' un external transcribed spacer *ETS*, seguito dal rRNA 18S, un internal transcribed spacer *ITS1*, poi un rRNA 5.8S, un *ITS2* e il 28S. Codificare gli rRNA diversi in un solo precursore assicura che le quantità di ogni RNA siano bilanciate. Il pre-rRNA viene elaborato: prima viene eliminato il *ETS*, poi rotto il *ITS1* e infine viene affinato il 18S e vengono separati il 5.8S e il 28S. Il 5.8S andrà poi a legarsi al 28S. Inoltre piccoli RNA nucleolari *snoRNA* con proteine associate formano i *snoRNP*, processosomi delle piccole e grandi subunità ribosomali che catalizzano il processo di maturazione del rRNA.

tRNA

Gli eucarioti producono trascritti di tRNA mono-cistronici sintetizzati dalla RNA polimerasi III. Un gene per un tRNA contiene una metà 5' separata dalla metà 3' da un introne. Dopo la trascrizione il macchinario di splicing dei tRNA rimuove l'introne, la *RNAasi P* rimuove la terminazione 5' mentre la *tRNAasi Z* la 3'. Successivamente dopo questo processamento viene aggiunto il *CAA* dalla *tRNA nucleotidil trasferasi* catalizzata dal sito di legame del nucleotide. La tasca possiede tre conformazioni: libera legata a *CTP* e legata a *ATP* che controllano se viene aggiunta una *C* o una *A* alla terminazione 3' del tRNA.

4.3 Modifiche dei nucleotidi di rRNA e di tRNA

I nucleotidi negli rRNA e nei tRNA vengono modificati chimicamente dopo la trascrizione: vari enzimi modificano le basi:

- Al gruppo ammino: adenina e guanosina NH_2 .
- A un atomo di azoto: guanina *N7* e citosina *N3*.
- A un atomo di carbonio: citosina e uracile *C3*.

Al ribosio a 2'-OH. Le modifiche sono enzimatiche e reversibili. Possono essere piccole come metilazioni o grandi come l'aggiunta di un amminoacido come la treonina. Il sito, la quantità e la distribuzione delle modifiche variano tra molecole di RNA, organismi e compartimenti intracellulari. Molte modifiche sono essenziali per la crescita e la sopravvivenza. Le modifiche più comuni al rRNA sono la metilazione del ribosio 2'-OH e la pseudo-uridilazione della base. Molte modifiche degli rRNA si trovano in basi importanti per la funzione dei ribosomi. Sono state identificate più di 100 modifiche negli RNA attraverso epitranscriptome sequencing e il 75% di esse si trova nei rRNA. Queste aumentano il numero di dimensioni e struttura e la stabilità della molecola di tRNA. Anche gli mRNA sono modificati con il capping m^7G aggiungendo una guanosina metilata a n^7 . Gli RNA spliceosomali, RNA piccoli nucleari *snRNA*, piccoli RNA nucleolari *snoRNA*, *miRNA* e *siRNA* sono modificati. Molte modifiche sono collegate al piegamento, attività e stabilità degli RNA, differenziazione della cellula, determinazione del sesso e risposta allo stress. Mutazioni negli enzimi modificanti il RNA causano malattie negli esseri umani.

4.3.1 Modifiche chimiche dei tRNA procarioti ed eucarioti

I tRNA subiscono delle modifiche a posizioni specifiche. Nella regione dell'anticodone aumentano l'efficienza della traduzione, mentre nel corpo centrale correggono la struttura secondaria e terziaria e aumentano la stabilità del tRNA in quanto l'assenza di certe di esse causa degradazione. Le modifiche combinate definiscono l'identità di ogni tRNA e tRNA che portano lo stesso amminoacido possono essere diversamente modificati.

4.3.2 Modifiche chimiche degli rRNA

Dopo la sintesi i pre-rRNA subiscono il processamento e due tipi di modifiche catalizzate dagli *snoRNP*:

- 2'-OH metilazione del ribosio in certi nucleotidi mediata dal *snoRNA C/D*.
- Pseudo-uridilazione dell'uracile creando uridina o pseudo-uridina Ψ mediata da *snoRNA H/ACA*.

Queste potrebbero contribuire alla stabilità e al piegamento del rRNA. L'interazione con le proteine ribosomali modula la biogenesi dei ribosomi e ne aumenta l'attività.

Enzimi coinvolti

Gli enzimi che catalizzano la metilazione del ribosio e la conversione dell'uridina in pseudo-uridina nel pre-rRNA sono conservati in tutti gli organismi, ma varia il meccanismo di riconoscimento. Negli eucarioti ed archea si utilizzano piccoli RNA guida nucleolari *snoRNA* che portano l'enzima modificatore al sito corretto. Questi si associano con un numero di proteine per formare un *RNP* attivo detto *snoRNP*. *snoRNP C/D* e *H/ACA* contengono 4 diverse proteine ognuna e i loro nomi derivano da box conservate presenti nella loro sequenza a RNA. Gli *snoRNA* sono lunghi tra i 60 e i 300 nucleotidi e la maggior parte sono fatti dagli introni dei pre-mRNA. Gli *snoRNA* guida si accoppiano con le basi di regioni specifiche dei pre-rRNA obiettivo e guidano gli enzimi come metiltransferasi frillarina *Nop1*, e la pseudo-uridina sintetasi discherina a quelle posizioni. Gli *snoRNA* che guidano la metilazione 2'-OH sono *C/D*, mentre quelli che guidano la pseudo-uridilazione delle basi sono *H/ACA*.

4.4 Capping e poliadenilazione di mRNA

Entrambe le terminazioni degli RNA eucarioti sono modificati durante la trascrizione. Queste modifiche proteggono gli mRNA dalla degradazione da parte delle esonucleasi e aiutano con le interazioni proteiche come i ribosomi. La terminazione 5' è incappucciata con la guanosina-P attraverso legame 5'-5' trifosfato. Questa guanina è poi metilata a N7. Il cap 5' è necessario per allungamento efficiente e terminazione del trascritto, per il processamento del RNA, esporto dal nucleo e direzionamento della traduzione. In eucarioti complessi il 2'-OH della prima, seconda e qualche volta terza base sono metilate.

4.4.1 Capping dei pre-mRNA eucarioti al 5'

Il cap 5' viene aggiunto in tre passi poco dopo che il mRNA emerge dalla RNA polimerasi II ed è lungo tra i 20 e i 30 nucleotidi e prima che avvenga una qualsiasi altra attività di processamento:

4.4. CAPPING E POLIADENILAZIONE DI MRNA

1. La RNA 5' trifosfatasi rimuove il γ -fosfato dalla terminazione 5'.
2. La guanil trasferasi attacca una guanosina monofosfato *GMP* alla terminazione $\beta\alpha$ -difosfato in un legame 5'-5' trifosfato.
3. La guanina-7-metil trasferasi metila la guanina in posizione *N7*.
4. La 2'-OH metiltrasferasi trasferisce un gruppo metile dalla *S-adenosilmetionina* al 2'-OH al ribosio dei primi due o tre nucleotidi alla terminazione 5'.

Nel lievito il processo viene svolto da tre enzimi diversi, mentre in *C. elegans* e nei mammiferi avvengono grazie a un singolo complesso enzimatico di capping che contiene le tre attività enzimatiche e ulteriori metilazioni di 2'-OH. Inoltre può avvenire un'iper-metilazione del cap 5' anche in piccoli RNA nucleari o nucleolari come *snRNA U1/U5/U3*.

4.4.2 Poli-adenilazione 3' dei pre-mRNA eucarioti

La terminazione 3' della maggior parte degli mRNA eucarioti contiene 200 adenosine aggiunte dette coda poli-A. Gli mRNA possiedono un sito di poli-adenilazione interno dove questi sono rotti e aggiunta la coda. Un mRNA può contenere multipli siti di poli-adenilazione e le sequenze in mezzo possono partecipare alla loro regolazione: la poli-adenilazione a un sito distale trattiene multiple regioni regolatorie, mentre ad un sito prossimale le rimuove. La coda 3' poli-A protegge il RNA dalle esonucleasi 3' nel nucleo e citoplasma, aumenta l'efficienza per il trasporto nel citoplasma e promuove la traduzione del trascritto.

Riconoscimento del sito di poli-adenilazione

Quando la RNA polimerasi II arriva alla terminazione 3' di un gene, trascrive il motivo *AATAAA* crea nel trascritto il segnale di rottura e poli-adenilazione *AAUAAA*, un motivo *CA* sito di rottura e poli-A e una regione ricca di *U* o *G* o *C*. Quest'ultima è importante in quanto molti introni sono ricchi in *A* e *T* e pertanto fornisce la specificità necessaria per impedire che vengano rotti questi erroneamente. I siti *AAUAAA* sono riconosciuti da *CPSF*: cleavage and polyadenylation specific factor. Il pre-mRNA è tagliato a *CA* da RNA endonucleasi complessi fattori di rottura *CFI* e *CFII* la cui attività è stimolata da *CStF* (fattore di stimolazione della rottura). Dopo la rottura sono aggiunte 200 adenosine dalla polimerasi poli-A. Viene utilizzata la *A* in quanto *ATP* è il nucleotide più abbondante. L'apparato di poli-adenilazione può riconoscere diversi siti *CA* rotti nel trascritto di pre-mRNA creando diversi mRNA da esso.

Il processo di poli-adenilazione

La rottura e poli-adenilazione avvengono durante la trascrizione: è un segnale per la RNA polimerasi II che deve terminare la trascrizione e dissociarsi dal DNA.

Iniziazione L'iniziazione consiste nell'aggiunta delle prime 10 adenosine: dipende dal segnale di rottura e poli-adenilazione e dalle sequenze segnale *GU* e *CA*. Richiede *CPSF*, *CStF*, *CFI* e *CFII* oltre alla polimerasi poli-A *PAP*.

Allungamento L'allungamento consiste nell'aggiunta di tutte le adenosine successive: richiede *PAP* e il *PABPN1* nucleare: la proteina legante poli-A nucleare 1.

Effetti della lunghezza della coda poli-A

La lunghezza della coda poli-A determina quanto a lungo il mRNA sopravvive: questa infatti aumenta l'efficienza dell'inizio della traduzione nel citoplasma. Il legame di *PABPC1*: proteina legante poli-A citoplasmatica 1 alla coda e i fattori di inizio della traduzione *eIF4E* e *eIF4G* circolarizzano mRNA con la terminazione 5' e permettono il legame del ribosoma.

Poli-adenilazioni alternative

Poli-adenilazione nei procarioti e negli organelli Nei procarioti, nei mitocondri e nei cloroplasti la coda poli-A è un segnale per la degradazione del mRNA: attraverso *PNPasi* polinucleotide fosforilasi 3'-5' RNAsi, dalla *ss RNAasi II* che compiono il ciclo di degradazione.

Istoni Gli mRNA che producono le proteine istoniche hanno il cap 5' ma non la coda poli-A: presentano uno stem-loop nella regione 3' UTR che marca la fine del trascritto dopo rottura 3' endonucleolitica.

Maturazione degli mRNA istonici Il processamento di questi mRNA inizia dal legame con la proteina legante lo stem-loop *SLBP* allo stem-loop. Il legame del *sRNP U7* all'elemento istonico *HDE* a valle crea contatti attraverso accoppiamenti di base con la terminazione 5' del *U7 snRNA*. Il legame al pre-mRNA è stabilizzato da interazioni con uno zinc finger *U7 snRNP* e *SLBP*. Il pre-mRNA è rotto dal complesso endonucleasi *CPSF73/100/symplekin*. Il mRNA poi si chiude attraverso la circolarizzazione della terminazione 3' e del cappuccio 5' e lega *eIF4G/4e* e *SLBP* per reclutare un ribosoma per la traduzione.

4.4.3 Legame con altre attività di polimerizzazione del RNA

Il capping 5' e la poli-adenilazione 3' sono legate con altre attività di polimerizzazione del RNA: il capping è necessario per l'allungamento della trascrizione e la poli-adenilazione 3' per terminazione efficiente. Il dominio C-terminale *CTD*= $[YS^2PTS^5PS]_{26-52}$ della subunità *Rbp1* della RNA polimerasi II media tutti i processamenti del RNA come una piattaforma di evento. La *CTD* recluta sequenzialmente diversi complessi di processamento:

- La *CTD* è fosforilata all'inizio della trascrizione e recluta il complesso o gli enzimi per il capping 5'.
- L'allungamento porta ad ulteriore fosforilazione di *CTD* che recluta il macchinario di splicing.
- Questo viene seguito dal reclutamento del complesso di rottura e poli-adenilazione.

4.5 RNA splicing

Il RNA finale è costituito da sequenze di esoni discreti, originariamente separati da introni rimossi dal pre-RNA. Ci sono quattro diverse classi di introni e tutte devono essere rimosse dal pre-RNA. La maggior parte degli introni non contengono geni e sono escissi e degradati. Eccezioni sono costituite da *snoRNA* e *miRNA*. Alcuni introni sono rimossi da enzimi come *tRNA*, altri da *RNP* come lo spliceosoma, mentre altri ancora si separano da soli. Gli introni sono prevalenti negli eucarioti, anche se alcuni virus ne posseggono alcuni che fanno self-splicing. Una rimozione alternativa degli introni permette la creazione di diversi trascritti e proteine isoforme dallo stesso gene, che tra gli umani

variano tra i 100 000 e i 1 000 000. Gli introni permettono inoltre un mescolamento degli introni a livello del DNA, dove gli esoni sono scambiati e riordinati attraverso ricombinazione, permettendo la produzione di diversi gene e proteine che codificano. Lo splicing avviene nel nucleo e negli organelli contenenti DNA genomico come mitocondri e cloroplasti.

4.5.1 Scoperta degli introni

Nel 1977 si nota come la sequenza di un RNA funzionale è diversa dalla sequenza dei geni: la seconda è più lunga e presenta sezioni assenti nel gruppo finale: si chiamano tali zone introni in quanto non sono incluse nel trascritto maturo. I geni pertanto sono discontinui. Questi vengono scoperti attraverso la tecnica del *R-loop mapping*. Dopo un incubazione ad alte temperature avviene l'ibridazione tra pre-mRNA e la sequenza genica e mRNA e la sequenza. Il prodotto viene macchiato con uranil-acetato reso scuro con platino/palladio e visualizzato attraverso microscopia elettronica. Gli introni sono stati successivamente confermati sequenziando il DNA del gene, il cDNA derivato dal pre-mRNA e dal mRNA maturo. Originariamente si fecero due ipotesi: nella prima si supposeva che la RNA polimerasi fosse discontinua e saltasse delle sequenze, mentre nella seconda il trascritto è completo ed è successivamente elaborato. Si nota come la corretta è la seconda attraverso northern hybridization blots: isolamento degli RNA, ibridazione con una sequenza ss radioattivamente etichettata: analizzando il campione nel tempo si nota la scomparsa dei pre-mRNA di lunghezza intera e l'apparizione di mRNA intermedio e maturo.

Caratteristiche degli introni

Il numero di introni in un gene varia tra le specie. In *S. cerevisiae* il 5% dei geni possiede introni e tipicamente 1 per gene, mentre per gli umani li possiedono il 85% con 11 introni per proteina in media. La connettina ne possiede 363. Gli introni sono presenti nei geni codificanti mRNA, rRNA e tRNA. Negli umani il 95% della sequenza di un pre-mRNA è formata da introni con sequenze molto più lunghe rispetto agli esoni.

4.5.2 Tipi di splicing del RNA

Ci sono quattro tipi di RNA splicing:

- Splicing nucleare fatto dallo spliceosoma, un grande complesso ribonucleoproteico, e il tipo più comune di splicing. Avviene per i prodotti della RNA polimerasi II.
- Gruppo di introni I self-splicing.
- Gruppo di introni II self-splicing.
- tRNA splicing.
- Trans-splicing da parte dello spliceosoma.

Splicing nucleare

Le sequenze degli introni non sono conservate tranne che per corte sequenze di segnale che corrispondono ai segnali di riconoscimento e rimozione. Queste sono il sito donatore 5' introne, il sito A di branching, la lunghezza riccadi *CT* poli-pirimidina. *AG* finale 3' o sito accettore. Lo spliceosoma riconosce specifici siti intronici nel pre-mRNA in quanto i siti 5' e 3' sono molto meno conservati. Altre sequenze aiutano a definire i confini tra gli introni e gli esoni e aiutano le cellule a produrre diversi mRNA maturi dallo stesso gene.

Meccanismo di splicing I due esoni di ogni parte sono uniti da un processo a due passi: entrambi coinvolgono una reazione di transesterificazione in cui un legame fosfo-estere è rotto ma un altro è formato. L'energia simile tra i due legami indica che la reazione non richiede *ATP*. L'introne *lariat* viene degradato e pertanto la reazione non è reversibile. Si nota come è necessario *ATP* per l'assemblaggio dello spliceosoma.

Lo spliceosoma Lo spliceosoma è un macchinario formato da 60 proteine e 4 molecole di RNA nel lievito e 5 negli umani. Le dimensioni sono simili a quelle della piccola subunità ribosomiale e la composizione proteica differisce leggermente tra le specie. Lo splicing è mediato dal RNA in quanto lo spliceosoma è un ribozima. Non è coinvolta l'attività di RNAasi. Contiene 5 corti RNA o short nuclear RNA *snRNA* ricchi in uracile *U1*, *U2*, *U4*, *U5* e *U6*. Il numero indica l'ordine di attività e *U3* manca per motivi storici in quanto necessario nella maturazione del rRNA. I 5 *snRNA* formano accoppiamenti specifici con le basi di sequenze introniche conservate nel pre-mRNA. Ognuno di essi si lega a un insieme specifico di proteine formando 5 *snRNP*: small nuclear ribonucleoprotein. Ognuno di essi lega sempre a proteine *7Sm* che formano una ciambella che si lega a una sequenza conservata di 9nt nel *snRNA*: 5'-*AUUUGUG*-3'. L'accoppiamento di basi avviene tra *U1* e la sequenza donatrice 5' dell'introne e 3' dell'esone e tra *U2* e il branch point *A* e le basi intorno. Insieme decidono dove avviene lo splicing.

Formazione dello spliceosoma

1. *U1* riconosce il sito donatore 5' dell'introne.
2. *BBP* branching point binding protein riconosce il branch point, recluta *U2*.
3. *U2AF* *U2* associated factor si lega al sito di splice 3' includendo la sequenza accettrice *AG*.
4. *U2* riconosce il sito accettore 3' dell'introne.
5. *U4* recluta *U5* e *U6* al sito di splice 5' includendo la sequenza *GU*. Avviene la prima reazione di trans-esterificazione da parte di *U6*/
6. *U1* e *U4* lasciano il RNA.
7. *U2*, *U5* e *U6* mediano il riordinamento unendo i due esoni nella seconda reazione di transesterificazione da parte di *U2* e l'introne è rimosso.

snRNA

<i>snRNA</i>	Lunghezza (nt)	Funzione
<i>U1</i>	164	Riconosce il sito di splicing al 5' mediante l'appaiamento di una regione complementare.
<i>U2</i>	187	Riconosce il sito di ramificazione mediante l'appaiamento con una regione complementare.
<i>U4</i>	144	Forma un duplex con <i>U6</i> .
<i>U5</i>	116	Funzione sconosciuta, si lega a esone 1 e 2.
<i>U6</i>	106	Forma un duplex con <i>U4</i> , scalza <i>U1</i> nell'appaiamento con il sito di splicing al 5'.

Exon junction complex Il exon junction complex *EJC* è lasciato alla giunzione di splice dopo lo splicing in modo da marcare il trascritto come localmente processato.

Splicing errato Errori nello splicing portano a malattia come la distrofia muscolare di Duchenne e derivano da siti donatori, accettori o di branch mutati o mutazioni dei regolatori, proteine o *snRNA* spliceosomali.

Gruppi di introni I e II self-splicing

Nel 1982 viene scoperto un introne nel pre-rRNA 26S nel *Tetrahymena thermophila* ha fatto splicing di sé stesso dal proprio trascritto senza intervento di altre proteine o enzimi, co-fattori o *ATP*. Si nota come clonando il gene di rDNA in un plasmide, purificando la RNA polimerasi I batterica si produce un trascritto di rRNA che aggiungendo Mg^{2+} fa splicing. Il RNA viene pertanto considerato come un'entità simile a un enzima, un ribozima. Gli introni autocatalitici si dividono in quelli di gruppo I e di gruppo II con due meccanismi di self-splicing attraverso trans-esterificazione.

- Nel gruppo I avviene attraverso il 3'-OH di una guanosina al di fuori della sequenza di mRNA.
- Nel gruppo II avviene attraverso il 2'-OH di un'adenosina interna branchpoint simile all'attività dello spliceosoma.

Introni di gruppo I self-splicing Questi introni si escindono da soli al trascritto primario. Hanno una lunghezza che varia tra i 250 e i 500nt e si trovano in microorganismi eucarioti, piante, batteri e virus eucarioti nel DNA nucleare, mitocondriale e cloroplasto. Contengono una struttura secondaria conservata che contiene 9 regioni a stem-loop *P1-P9*. I siti di splice sono definiti dalla struttura tridimensionale dell'introne e dal riconoscimento di una *G* conservata in *P1* che forma una coppia wobble con *U* ultimo nucleotide dell'esone 1. Il loop *P7* lega *GTP*, *GDP*, *GMP* e guanina come il nucleofilo per la reazione. Gli introni codificano una maturasi, proteina senza attività catalitica che svolge il ruolo di RNA chaperone aiutando la reazione di splicing stabilizzando la struttura del RNA e l'attività di auto-splicing dell'introne. Gli introni potrebbero o no codificare una DNA endonucleasi che se presente rende l'introne mobile, ovvero può retro-trasporsi nel proprio allele nel meccanismo di homing.

Introni di gruppo I immobili Negli introni di gruppo I immobili il gruppo 3'-OH della guanosina libera *GTP* si localizza nella tasca *P7* e attacca il 5'-P al primo nucleotide dell'introne. Avviene la trans-esterificazione 1: si stacca la terminazione 5' dall'esone 1. Successivamente la terminazione rilasciata dell'esone 1 *U* 3'-OH attacca la giunzione introne-esone 2: 3'-P dell'ultimo nucleotide dell'introne e avviene la seconda reazione di trans-esterificazione. L'introne lineare non forma un lariat.

Introni di gruppo I mobili Negli introni di gruppo I mobili la trascrizione e lo splicing avviene come in quelli immobili, ma l'introne lineare viene esportato nel citoplasma dove viene tradotto. Si produce un'endonucleasi che ritorna nel nucleo. L'allele omologo subisce una rottura a doppio filamento e l'allele originale agisce come donatore per la riparazione. Avviene una riparazione *DSB* o *SDSA* (synthesis-dependent strand annealing). Il processo di homing avviene attraverso intermedi del DNA.

Ruoli evolutivi secondari delle endonucleasi homing Membri di varie famiglie di endonucleasi homing presentano omologia strutturale e relazioni funzionali con una grande varietà di proteine da vari organismi. Le funzioni biologiche includono enzimi di degradazione del DNA non specifici, endonucleasi di restrizione, enzimi di riparazione del DNA, resolvasi e fattori di trascrizione anche auto-repressivi. Queste relazioni suggeriscono che queste endonucleasi homing condividono antenati comuni con proteine coinvolte nella fedeltà dei genomi, loro mantenimento ed espressione genica. Pertanto quando sono espresse le endonucleasi homing possono contenere attività enzimatiche addizionali che agiscono nella cellula. Il homing amplifica il loro numero genico e il suo effetto nella biologia della cellula.

Introni di gruppo II self splicing Questi introni sono lunghi tra i 400 e i 1000nt. Si trovano in eucarioti, piante, archea e batteri. Possiedono una struttura secondaria conservata con 6 domini a stem-loop *D1-D6*. Il dominio *D4* potrebbero codificare per maturasi o maturasi, endonucleasi e trascrittasi inversa rendendo l'introne mobile che fa splicing invertito. La struttura terziaria degli esoni e il branching point *A* in *D6* si trovano vicini. Il meccanismo di splicing è simile a quello dello spliceosoma e guidato dagli ioni magnesio. Non è necessario il co-fattore nucleotidico: avviene l'attacco nucleofilo dal 2'-*PH* da *A* al 5'-*P* dell'esone 1. Successivamente avviene l'attacco nucleofilo dal 3'-*OH* dall'esone 1 al 3'-*P* dell'introne. L'introne rimosso forma un lariat. Si nota come questi introni sono antenati dello spliceosoma eucariote.

Introni mobili di gruppo II: retrohoming Se *D6* codifica per un endonucleasi questa taglia il filamento basso di DNA e il ss lariat liberato si re-integra nel filamento superiore invertendo i passi di trans-esterificazione della reazione in avanti. Dopo che la maturasi promuove lo splicing dell'introne questo individua grazie a regioni l'allele omologo e crea una rottura doppio filamento. Il lariat si inserisce e una trascrittasi inversa sintetizza il ssDNA in cui il taglio basso agisce come primer. Successivamente una DNA polimerasi sintetizza il filamento complementare.

Splicing del tRNA

I trascritti di tRNA eucarioti sono mono-cistronici e sono creati dalla RNA polimerasi III. Sono tipicamente formati da una metà 5' seguita da un introne e infine una metà 3'. Dopo la trascrizione avviene l'escissione dell'introne dal macchinario di splicing del tRNA. Successivamente una RNasi P e una tRNAasi Z tagliano le estremità e una tRNA nucleotidil trasferasi aggiunge *CCA* formando il tRNA maturo. La maturazione dei pre-tRNA in archea ed eucarioti non coinvolge reazioni di trans-esterificazione e necessita pertanto di un numero di enzimi: un'endonucleasi rimuove l'introne. Successivamente nella metà 5' si forma 2'-3' fosfato ciclico. Nella metà 3' una chinasi dipendente da *ATP* fosforilazione la terminazione 5'-*OH* e una ligasi tRNA-specifica e *ATP* aggiunge alla 5'-*P* un *AMP*. A questo punto nella metà 5' una fosfodiesterasi *PDE* apre l'anello di fosfato. Si forma il legame tra 3'-*OH* nella metà 5' e il 5'-*OH* dell'altra con il rilascio di *AMP*. Una fosfatasi rimuove il gruppo fosfato in 2'-*OH* della prima metà e si forma il legame fosfodiesterico. Si nota come lo splicing avviene nel citoplasma: il tRNA subisce delle modifiche, viene esportato nel citoplasma, reimportato nel nucleo e infine re-esportato.

Trans-splicing

Si dice cis-splicing quando gli esoni nello stesso pre-RNA sono uniti insieme. Si nota come è possibile come esoni di due diversi pre-RNA si possano unire nel trans-splicing. Questo processo avviene in archea, eucarioti unicellulari, piante e nematodi ma non negli umani. Si riconosce un pre-mRNA

accettore con un *A* branch point e un corto *SL RNA* donatore (spliced leader). Il secondo è più corto di 150nt, possiede un cap 5' e una sequenza non codificante leader di 20nt e un introne ed è contenuto in un *Sm snRNP*. Il *SL RNA* sostituisce il *U1 snRNP* e interagisce con altri *snRNP* al sito di splicing 3'. Il processo coinvolge lo spliceosoma. Si nota come i siti *GU* donatore e *AG* accettore risiedono in due diverse molecole di RNA. Avvengono due reazioni di trans-esterificazione. *SL RNA* viene trascritto dalla RNA polimerasi II.

4.6 Definizione degli esoni e splicing alternativo

Lo splicing può portare a più di un mRNA maturo: la maggior parte dei geni nei eucarioti complessi subisce splicing alternativo, dove sono usate diverse combinazioni di esoni. La maggior parte di essi sono costitutivi e sempre inclusi, mentre alcuni sono regolati e possono essere esclusi. Possono essere anche usati siti di splicing alternativi alla terminazione 5' o 3'. Si possono usare inizi di trascrizione alternativi e diversi di terminazione. Lo splicing alternativo è importante per la diversità genetica come nel gene *dscam* in *Drosophila* che può creare 38000 diversi trascritti maturi. Questo gene infatti contiene 24 geni e 4 che sono cluster che a loro volta contengono 98 esoni alternativi. Questo codifica per un axal guidance reporter per lo sviluppo neurale. In diversi tessuti vengono espressi diversi isoformi della proteina e lo splicing è regolato in tempo e spazio durante lo sviluppo.

4.6.1 Splicing alternativo: diversità e complessità delle specie

Negli eucarioti si nota un numero di geni simile ma una grande diversità nella complessità della specie. Questo avviene grazie allo splicing alternativo: da 1 pre-mRNA nel 90% dei geni umani si possono produrre diverse proteine. Si nota come gli introni sono sempre eliminati, mentre gli esoni potrebbero esserlo.

Riconoscimento dei veri siti di splicing

Le sequenze che definiscono le giunzioni tra introni ed esoni sono semplici, corte e degenerative: possono trovarsi da qualche altra parte e portare a splicing non voluto o i siti di splicing criptici. Essendo che lo splicing avviene con alta fedeltà lo spliceosoma deve poter riconoscere i veri siti di splicing. Ci sono due modelli che propongono il meccanismo di riconoscimento.

Definizione degli esoni Nei mammiferi le terminazioni 5' e 3' di un esone sono portate insieme da interazioni tra i complessi *U1* e *U2*: gli introni che li affiancano subiscono splicing se è presente a monte *U1* o *U2* a valle. Mutazioni nel sito di splice risultano in un'esclusione di un esone a causa della mancanza di legame di *RNP U1*.

Definizione degli introni In invertebrati, funghi e piante gli introni sono definiti da interazioni tra *U1* e *U2* legati ai confini 5' e 3' dell'introne. Mutazioni al sito di splice 5' risulterebbero in assenza di splicing e inclusione dell'introne nel trascritto maturo.

Conclusione Si nota come in entrambi i modelli i siti di splice 3' e 5' sono marcati mentre sono trascritti: gli esoni da proteine *SR* e gli introni da *hnRNP*.

4.6.2 Elementi di sequenze di RNA addizionali

Elementi di sequenze di RNA addizionali hanno un effetto sulla funzione dello spliceosoma. L'attività di questi elementi dipende dal contesto locale: lo stesso motivo può agire come repressore o attivatore se si trova in un esone o in un introne.

Slicing enhancers

Si dicono splicing enhancers *SE* sequenze che promuovono lo splicing. Possono essere intronici *ISE* o esonici *ESE*. Si legano a un *SR* proteine (ricche di serina/arginina) e promuovono l'assemblaggio dello spliceosoma.

Splicing silencers

Si dicono splicing silencers *SS* sequenze che inibiscono lo splicing. Possono essere intronici *ISS* o esonici *ESS*. Sono legati da proteine *hnRNP* (ribonucleoproteine nucleari eterogenee) che mascherano siti di splicing critici in un introne e inibiscono l'interazione tra *RNP U1* e *RNP U2* dello spliceosoma.

4.6.3 Exon shuffling

Si definisce exon shuffling il processo di mescolamento degli esoni che avviene durante la meiosi. Il mescolamento di tali domini necessari per la funzione delle proteine permette la creazione di nuove combinazioni di proteine funzionali.

4.7 *miRNA* e *siRNA*

4.7.1 *miRNA*

Si dicono microRNA *miRNA* piccoli RNA non codificanti lunghi tra i 20 e i 22nt. Sono stati scoperti in *C. elegans* nel 1993 nello sviluppo larvale. Si legano a mRNA attraverso accoppiamento di basi complementari nella region 3' *UTR*. In questo modo impediscono la traduzione e causano rottura e degradazione dei mRNA. Possono inoltre essere reclutati alla cromatina per silenziare la trascrizione. A causa della loro corta sequenza non sono molto specifici e possono silenziare diversi geni.

Formazione

I miRNA sono prodotti da un precursore unico a singolo filamento *pri-miRNA*. Sono endogeni e derivano da RNA codificanti o non codificanti, introni od esoni. Sono trascritti pertanto dalla RNA polimerasi II e presentano capping 5', splicing e poliadenilazione 3'. Il *pri-miRNA* presenta vari stem-loop. La formazione del *miRNA* dal *pri-mRNA* segue varie fasi:

1. Nel nucleo avviene cropping da parte di *Drosha*, un complesso RNAasi III, una endonucleasi per dsRNA che separa gli stem-loop.
2. Gli stem loop ora lunghi tra i 60 e i 100nt sono traslocati nel citoplasma da un complesso formato da esportina 5 e da *GTP* attraverso i pori nucleari.
3. Nel citoplasma avviene il dicing da parte del *Dicer*, un complesso RNasi III che rimuove il loop lasciando un *miRNA* duplex lungo tra i 19 e i 24nt.

4.8. RIBOZIMI AUTO-CATALITICI

4. Un filamento del *miRNA* guida o attivo lega alla subunità 3 del complesso *RISC*: la subunità argonauta.
5. *RISC-miRNA* si lega alla regione 3' *UTR* del mRNA obiettivo.

Interferenza a RNA

L'interferenza a RNA causa un silenziamento traduzionale bloccando il legame o l'attività del ribosoma, promuove la rottura e degradazione del mRNA. Infine il complesso *RISC-miRNA* torna nel nucleo per legare code istoniche o DNA reclutando enzimi di silenziamento.

4.7.2 *siRNA*

Si dicono *siRNA* gli small interfering RNA che silenziano il RNA. Sono formati nello stesso modo dei miRNA: emergono da un lungo precursore di dsRNA. Non coinvolgono Drosha e sono processati dal Dicer e diventano parte del complesso *RISC*. I *siRNA* si originano da dsRNA endogeno o esogeno. Si legano perfettamente al mRNA target con alta specificità causando la sua degradazione. *RISC-siRNA* può legarsi alla coda istonica o a sequenze di DNA *CpG* per reclutare enzimi di silenziamento epigenetico.

4.8 Ribozimi auto-catalitici

Esistono piccoli RNA enzimatici lunghi tra i 200 e i 400nt che rompono un legame fosfodiesterico a un sito specifico in una molecola di RNA come gli enzimi endonucleasi. Nel 1967 si nota come RNA forma strutture stabili e complesse di vario tipo con attività catalitica come di processamento e regolazione di piccoli genomi di RNA. Si possono sintetizzare in laboratorio ma non hanno applicazioni terapeutiche.

4.8.1 Ribozimi hammerhead

I ribozimi hammerhead sono così nominati a causa della loro caratteristica forma a martello. Sono altamente specifici per il substrato con due siti di legame posti a 90° tra di loro: a 3' si trova Stem III con il legame specifico per il substrato, mentre a 5' Stem I. Tra di loro si trova Stem I, il dominio catalitico in cui si trova il sito di rottura endonucleolitica.

Meccanismo di rottura

????????????????????????????

4.9 RNA editing

Sono diffusi due modifiche del RNA a causa della deaminazione della base. Si nota come modifiche nel primo o secondo amminoacido porta più facilmente a modifiche traduzionali.

- Deaminazione dell'adenosina porta alla formazione dell'inosina, edit comune negli umani. Questo avviene da parte di enzimi della famiglia *ADAR*: adenosine deaminase acting on RNA. L'inosina è interpretata come guanosina, pertanto cambi nella regione codificante possono cambiare la sequenza proteica si porta come aberrazioni in questo possono portare ad epilessia, depressione, schizofrenia, sclerosi laterale amiotrofica o cancro.

- Deaminazione della citosina in uracile. Questo avviene da parte di enzimi *PPR*: pentatricon peptide repeat negli mRNA mitocondriali e dei cloroplasti. Avviene costitutivamente nell'apolipoproteina B nell'essere umano: nel fegato il mRNA rimane invariato e rimane una proteina in forma lunga che trasporta colesterolo nel fegato. Nell'intestino tenue invece avviene un edit che porta alla creazione di un codone di stop e una forma corta richiesta per l'assorbimento di lipidi dal cibo.

4.9.1 Editing del trascritto attraverso inserzione o delezione di nucleotidi

L'inserzione o delezione di nucleotidi viene individuata nel 1986 in *Trypanosoma brucei*, che causa malaria. Avviene tipicamente nel RNA mitocondriale: possiede infatti un grande mitocondrio in cui il DNA esiste o come 20-50 grandi cerchi o come 5000-10 000 piccoli cerchi. I pre-mRNA prodotti dai grandi cerchi non sono funzionali e necessitano di inserzione o delezione di uracile fino a centinaia come avviene nella subunità 6 della ATPasi.

Meccanismo di editing

I geni di *T. brucei* sono presenti in forma non riconoscibile o criptogeni. Migliaia di RNA guida *gRNA* sono codificati dai piccoli cerchi e sono piccoli RNA anti-senso che si legano a valle dopo la posizione di editing. Questi sono necessari affinché l'editing avvenga nella posizione corretta e per determinare il numero corretto di *U*.

Editosoma L'editosoma è il complesso che compie l'editing e contiene:

- Una *ssRNAasi* endonucleasi per inserzione o rimozione di *U*.
- Una trasferasi uridile terminale *TUTasi* per l'inserzione di *U*.
- Una *ssRNA* esonucleasi per la delezione di *U*.
- Una RNA ligasi.
- Elicasi per lo stacco dei *gRNA*.

La sua attività è progressiva: inizia alla terminazione 3' e sua attività sequenziale lo porta verso la terminazione 5'. L'ultimo *gRNA* potrebbe inserire una *U* e creare un sito di inizio di traduzione.

4.10 Traslocazione nucleo-citoplasmatica di RNA processati

Dopo il processamento del RNA questo deve uscire il nucleo per agire o essere tradotto attraverso complessi *snRNP* o *snoRNP*. Si rendono pertanto necessari sistemi inclusi nella membrana per l'importo e l'esporto: il complesso dei pori nucleari e le carioferine, importine ed esportine insieme a *GTP* che riconoscono un sistema di localizzazione nucleare o di uscita, rispettivamente *NLS* e *NES* non presenti nel RNA. Il complesso dei pori nucleari *NPC* svolge un processo di controllo della qualità del RNA prima che esca dal nucleo e quello che non lo passa viene degradato dall'esosoma.

4.11 Degradazione di RNA endogeni

Gli RNA devono essere degradati ad un certo punto per rimuovere quelli non più necessari e svolgere un riciclo dei nucleotidi. Alcuni rRNA sono necessari e devono essere stabili, altri come gli mRNA sono richiesti per corti periodi e sono degradati rapidamente. Gli RNA endogeni sono degradati in maniera diversa rispetto a quelli stranieri o difettivi.

4.11.1 Stabilità dei mRNA

La stabilità dei mRNA è determinata da vari fattori.

Strutture alle terminazioni

Il cap 5' e la coda poli-A 3' proteggono contro la digestione di esonucleasi di mRNA eucarioti. Gli RNA batterici con un trifosfato 5' sono più stabili rispetto a quelli con un monofosfato. Stem loop 3' in batteri proteggono contro attività esonucleasica. Nei batteri la coda poli-A 3' diminuisce la stabilità.

Sequenze regolatorie interne

Le differenze nel tasso di turnover possono essere determinate nella sequenza e struttura stessa del RNA. Gli elementi destabilizzanti *ARE* sono elementi ricchi in *AU* con una ripetizione consenso $[AUUUU]_n$ nella regione 3' *UTR* di molti mRNA. In risposta a vari segnali *ARE* reclutano proteine leganti RNA che reclutano esonucleasi o endonucleasi che rimuovono la coda poli-A 3' causando la degradazione del mRNA:

- *AUF1* interagisce con l'esosoma e de-adenilasi *PARN* e *CCR4-NOT* causando la degradazione della coda poli-A.
- *HuR* compete con *AUF1* per il legame con *ARE* garantendo la stabilità del mRNA e l'inizio della traduzione.

Attività del RNA

Attività del RNA come splicing, trasporto e traduzione possono avere un impatto sulla half-life bloccando o permettendo l'accesso a enzimi di degradazione.

4.11.2 Degradazione del RNA nei procarioti

Nei procarioti l'attività enonucleasica inizia la degradazione degli RNA. In *E. coli* 12 RNAasi agiscono anche durante la traduzione. La degradazione di mRNA batterici è iniziata da un idrolasi pirofosfato e un endonucleasi come *RNAasi E*. I prodotti di questa digestione interna sono ulteriormente degradati da una esonucleasi 3'-5' come *RNAasi II* e da RNAasi 5'-3' come *RNAasi J1*. La *RNAasi E* è parte del grande complesso del degradosoma che contiene RNA elicasi e l'esonucleasi *PNP*. La RNA elicasi è necessaria per eliminare gli stem-loop che inibiscono l'attività di esonucleasi. La coda poli-A è sensibile all'attività di esonucleasi 3'-5' in quanto non strutturata e non protetta da proteine.

4.11.3 Degradazione del mRNA negli eucarioti

La coda poli-A impedisce la degradazione negli eucarioti e pertanto il primo passo è il suo accorciamento da una deadenilasi. Dopo la deadenilazione possono avvenire due eventi:

- Enzimi di decapping rimuovono il cap 5', seguiti da attività esonucleasica 5'-3' come *Rat1*.
- L'esosoma catalizza la degradazione 3'-5'.

Esosoma

??????????????

4.12 Degradazione di RNA esogeni *siRNA CRISPR*

Alcuni tipi di molecole possono essere dannose alla cellula come RNA invadente esogeno o RNA da virus o RNA endogeno difettivo. Il RNA difettivo è rimosso da interferenza a RNA *siRNA* negli eucarioti e da *CRISPR* nei batteri. Nel *siRNA* il dsRNA straniero è tagliato in migliaia di frammenti lunghi tra i 20 e i 30nt *siRNA*. Dei corti frammenti una guida attiva ss è caricata sulla nucleasi argonauta parte del complesso *RISC* e lo guida a mRNA specifico derivato dal dsRNA invadente per degradarlo.

4.12.1 Interferenza *CRISPR* nei batteri

Nei batteri il locus *CRISPR*: clustered interspaced palindromic repeats è coinvolto nella degradazione del DNA di fagi stranieri dopo il processamento del RNA e una guida a RNA. Il locus agisce come un sistema immunitario adattivo che fornisce una difesa contro infezioni ripete dello stesso fago. Questo sistema è basato su un meccanismo:

- Il batterio acquisisce una sequenza campione del fago dopo l'infezione.
- Questa sequenza viene integrata con altri campioni di DNA da altri fagi nel locus *CRISPR*.
- In caso avviene una seconda infezione il fago invadente è identificato comparando il DNA con i frammenti conservati nel locus *CRISPR*.
- Se si trova un match il DNA del fago invadente è distrutto.

CRISPR in *Streptococcus pyogenes*

Dopo che un fago ha infettato il suo DNA nel citoplasma di *S. pyogenes* il complesso DNA endonucleasi-ricombinasi *Cas1-Cas2* si forma. Un campione protospacer è tagliato dal DNA del fago e viene inserito nel genoma batterico al locus *CRISPR* e diventa uno spacer. Nel loco si trovano diversi spacer da precedenti contatti con fagi. Le sequenze spacer sono poi trascritte come un lungo pre-crRNA circolare. *S. pyogenes* produce le DNA endonucleasi *Cas9* come monitor che si legano al RNA lungo su un frammento spacer rilasciato da RNAasi III. La sequenza crRNA spacer guida *Cas9* alla sequenza complementare al genoma virale iniettato e marca tagli locali nel DNA.

Il trascritto del locus *CRISPR* Il locus *CRISPR* di *S. pyogenes* contiene un operone *cas*, un array *CRISPR* e geni codificanti per *crRNA* attivanti in trans o tracker. Le sequenze spacer sono affiancate da ripetizioni palindromiche e il RNA tracker forma una struttura terziaria con 3 loop. Il gene *cas9* produce la proteina *Cas9*.

Maturazione del *pre-crRNA* e formazione del complesso di interferenza Il *CRISPR* RNA lungo creato come un trascritto continuo matura. Il *tracrRNA* si accoppia con i palindromi e recluta *Cas9*. Si forma un complesso *spacer-tracrRNA-Cas9* localmente e viene rilasciato da RNAasi III. Il complesso di interferenza è poi guidato dallo *spacer* al DNA virale.

Utilizzo di *CRISPR-Cas9* come uno strumento di ingegneria genetica

Unendo sequenze derivate da *tracrRNA* e *crRNA* con un loop linker si può creare uno strumento programmabile per la manipolazione del DNA. La loro unione forma un RNA guida *sgRNA* che può essere creato sinteticamente e *Cas9* da *E. coli* attraverso ricombinazione. Il complesso di interferenza *Cas9-sgRNA* viene elettroporato in una cellula e scanna il genoma per un protospacer adjacent motif *PAM*. Quando questa viene trovata il DNA complementare è comparato con il *sgRNA*: se si trova un match il DNA è rotto localmente 3bp distante da *PAM* in entrambi i filamenti. Il lobo *NUC* rompe un filamento attraverso *HNH* e l'altro attraverso *RuvC*. Quando avviene una rottura a doppio filamento sono possibili due cammini:

- *NHEJ*: non-homologous end joining che potrebbe produrre inserzioni o delezioni, mutazioni indel che possono creare un frame-shift o un codone di stop prematuro con perdita della funzione del gene.
- *HDR*: homology directed repair: viene utilizzato uno stampo a DNA per riparare il sito dove avviene la rottura: quando DNA esogeno viene aggiunto alla cellula la sequenza potrebbe servire come stampo ed è poi integrata nel processo di riparazione. Questo potrebbe causare all'introduzione di una nuova sequenza genetica.

Capitolo 5

RNA regolatori

Capitolo 6

Traduzione

Capitolo 7

Modifica e targeting delle proteine

Capitolo 8

DNA mobile

Capitolo 9

Strumenti e tecniche della biologia molecolare