

Biologia molecolare della cellula 2

Giacomo Fantoni

Telegram: @GiacomoFantoni

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

22 gennaio 2021

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 <i>X</i> 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	55
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

5	RNA regolatori	73
5.1	Panoramica degli RNA regolatori	73
5.1.1	Principi fondamentali	73
5.1.2	Iterazioni tra le basi	73
5.1.3	Codifica	73
5.2	Piccoli RNA batterici	74
5.2.1	Funzioni	74
5.2.2	Traduzione del trascritto regolata dal ferro	74
5.2.3	Livelli di complementarietà	74
5.3	<i>sRNA</i> eucarioti: <i>miRNA</i> , <i>siRNA</i> e <i>piRNA</i>	75
5.3.1	Classi	75
5.4	La famiglia di proteine <i>ArgonAUTA</i>	75
5.4.1	Classi	75
5.5	Processamento di <i>sRNA</i> eucarioti	76
5.5.1	Il pathway di microRNA	76
5.5.2	Il pathway di piwi interacting RNA	77
5.5.3	Gli enzimi Drosha e Dicer RNAasi III	77
5.6	Silenziamento genico da parte di RNA eucarioti	78
5.7	Ruolo della difesa virale di <i>sRNA</i> batterici, eucarioti e di archea	78
5.8	Regolazione mediata da RNA in <i>cis</i>	78
5.8.1	Sintesi della amminoacil tRNA sintetasi	78
5.8.2	Riboswitches	79
5.9	RNA regolatori leganti proteine	79
5.10	RNA non codificanti lunghi intergenici	79
6	Traduzione	80
6.1	Panoramica della traduzione	80
6.1.1	Ribosomi	80
6.1.2	Produzione e traduzione di mRNA	80
6.1.3	Confronto tra trascrizione e traduzione	81
6.2	Ribosoma	81
6.2.1	Scoperta	81
6.2.2	Composizione	81
6.2.3	Il ribosoma come un ribozima	82
6.2.4	Assemblaggio dei ribosomi negli eucarioti	83
6.2.5	Rappresentazione funzionale del ribosoma	83
6.3	tRNA e codice genetico	83
6.3.1	Struttura	83
6.3.2	Sintesi	84
6.3.3	Codoni	84
6.3.4	Legame con il ribosoma	84
6.4	Amminoacil-tRNA sintetasi	84
6.4.1	Struttura	85
6.4.2	Caricamento	85
6.4.3	Correzione degli errori	85
6.5	mRNA	86
6.5.1	Struttura nei procarioti	86
6.5.2	Struttura negli eucarioti	86

6.6	Ciclo di traduzione	86
6.6.1	Iniziazione	87
6.6.2	Allungamento	87
6.6.3	Terminazione e riciclo dei ribosomi	87
6.7	Iniziazione della traduzione - caratteristiche comuni a batteri ed eucarioti	87
6.7.1	Il tRNA iniziale nei procarioti	87
6.8	Iniziazione della traduzione batterica	88
6.8.1	Riconoscimento della sequenza Shine-Dalgarno	88
6.8.2	Assemblaggio del complesso di inizio	88
6.9	Iniziazione della traduzione eucariotica	89
6.9.1	Processo di iniziazione	89
6.9.2	Metodi di scansione	89
6.9.3	Presenza della metionina all'inizio della proteina prodotta	90
6.10	Allungamento della traduzione	90
6.10.1	Decodifica, legame di un amminoacil-tRNA ^{aa} al sito A	90
6.10.2	Formazione del legame peptide tra amminoacidi nel centro di trasferimento del peptidil	91
6.10.3	Traslocazione del peptidil-tRNA dal sito A al sito P	92
6.10.4	Allungamento negli eucarioti	92
6.11	Terminazione e reinizio della traduzione	92
6.11.1	Fattori di rilascio	92
6.11.2	Riconoscimento del codone di stop	93
6.11.3	Riciclo del ribosoma	93
6.12	Energia richiesta per la traduzione	93
6.12.1	Iniziazione	93
6.12.2	Allungamento	94
6.12.3	Terminazione	94
6.13	Velocità ed accuratezza della traduzione	94
6.14	Salvataggio dei ribosomi e controllo qualità degli mRNA	95
6.14.1	mRNA senza stop codone	95
6.14.2	Il ribosoma stalla a un ostacolo	95
6.14.3	Stop codone prematuro	96
6.15	Ricodifica - codone di stop programmato read-through e frame-shifting	96
6.15.1	Frame-shifting	96
6.15.2	Soppressione non senso	97
6.16	Antibiotici che hanno come obiettivo l'attività del ribosoma	97
6.16.1	Esempi	98
6.16.2	Resistenza batterica agli antibiotici	98
6.17	Regolazione globale dell'iniziazione della traduzione in batteri ed eucarioti	98
6.17.1	Stringent response	99
6.17.2	Eucarioti	99
6.18	Regolazione dell'iniziazione attraverso sequenze agenti in <i>cis</i> nella 5'-UTR in batteri ed eucarioti	99
6.18.1	Cambi di conformazione degli mRNA	100
6.18.2	Riboswitches	100
6.18.3	Regolazione dell'attività di 5'-UTR da parte di sRNA nei procarioti	101
6.19	Regolazione della traduzione attraverso sequenze agenti in <i>cis</i> nella 3'-UTR negli eucarioti	101

6.19.1	Xenopus laevis	102
6.20	Trasporto e localizzazione degli mRNA	102
6.20.1	Localizzazione del mRNA di <i>ASH1</i> nel lievito	102
6.21	Granuli citoplasmatici di RNA e P-bodies	102
6.21.1	P-bodies	102
6.21.2	Granuli di stress	103
7	Modifica e targeting delle proteine	104
7.1	Piegamento delle proteine assistito da chaperones	104
7.1.1	Processo di piegamento	104
7.1.2	Ponti disolfuro	104
7.2	Targeting di cellule attraverso la cellula	105
7.2.1	Ordinamento delle proteine	105
7.2.2	Entrata ed uscita delle proteine dal nucleo	105
7.3	Rottura post-traduzionale della catena polipeptidica	105
7.3.1	Insulina	105
7.3.2	Rottura post traduzionale delle proteine e Alzheimer	106
7.3.3	Splicing di proteine: rimozione di inteine	106
7.4	Regolazione e modifica post-traduzionale di proteine	107
7.4.1	Ubiquitinaizone delle proteine	107
7.4.2	<i>SUMOilazione</i> delle proteine	109
7.4.3	Regolazione dell'attività di <i>PCNA</i> attraverso ubiquitinazione e <i>SUMOilazione</i>	110
7.4.4	Sindrome di Alzheimer	110
7.5	Fosforilazione	110
7.5.1	Chinasi	110
7.5.2	Effetti del gruppo fosfato	111
7.5.3	Tirosina chinasi	112
7.5.4	Chinasi dipendenti dalla ciclina <i>CDK</i>	112
7.6	Acetilazione	113
7.6.1	Meccanismo di acetilazione	113
7.6.2	Acetilazione da parte di istone acetil trasferasi	114
7.6.3	Deacetilazione da parte di istone deacetilasi	114
7.6.4	Aberrazioni dei processi	114
7.7	Metilazione	114
7.7.1	Protein methyltransferases <i>PRMT</i>	114
7.7.2	Demetilasi	114
7.7.3	Metilazione aberrante	115
7.8	I domini lettori	115
7.9	Glicosilazione	115
7.9.1	Glicosilazione in diversi organismi	115
7.9.2	Ruoli della glicosilazione	115
7.9.3	Glicani	116
7.9.4	Tipi di glicosilazione	116
7.9.5	Ulteriori modifiche	118
7.9.6	Glicosilazione specifica-specifica di proteine ricombinanti	118
7.10	Modifiche lipidiche	118
7.10.1	Tipologie	119
7.10.2	Modifiche multiple	119

7.11	<i>ADP</i> -ribosilazione	119
7.11.1	Sintesi di <i>ADP</i> -ribosio	119
7.11.2	Mono <i>ADP</i> -ribosilazione	119
7.11.3	Poli <i>ADP</i> -ribosilazione	120
7.11.4	<i>ADP</i> -ribosilazioni possibili	120
7.11.5	<i>Corinebacterium dipgterium</i>	120
7.12	Modifica chimica diretta	120
7.12.1	Specie reattive all'ossigeno <i>ROS</i>	120
7.12.2	Specie reattive all'azoto	120
8	DNA mobile	121
8.1	Panoramica degli elementi trasponibili	121
8.1.1	Trasposoni e malattie umane	121
8.1.2	Eventi di trasposizione	121
8.1.3	Scoperta dei trasposoni	123
8.1.4	Caratteristiche della trasposizione	123
8.1.5	Tipi di trasposoni	123
8.1.6	Elementi trasponibili nei procarioti	124
8.2	Panoramica dei trasposoni a DNA	125
8.2.1	Trasposizione DNA-only non replicativa cut-and-paste	125
8.2.2	Trasposizione DNA-only replicativa nick-and-paste	126
8.3	Retrotrasposoni	127
8.3.1	Trasposoni umani	128
8.3.2	Retrotrasposoni <i>LTR</i>	128
8.3.3	Retrotrasposoni non- <i>LTR</i>	129
8.4	Controllo della trasposizione	130
8.4.1	Meccanismi di controllo	130
8.5	Panoramica di <i>CSSR</i>	131
8.5.1	Ricombinasi sito-specifiche	131
8.5.2	Conversione <i>CSSR</i> di dimeri di DNA in monomeri	132
8.6	Integrazione ed escissione del batteriofago λ	132
8.6.1	Integrasi λ <i>Int</i>	132
8.6.2	Integrazione del fago λ	132
8.6.3	Escissione del fago λ	133
9	Strumenti e tecniche della biologia molecolare	134
9.1	Separazione di molecole biologiche	134
9.1.1	Separazione per elettroforesi su gel	134
9.1.2	Processo	134
9.1.3	Colorazione delle molecole	135
9.1.4	Stima della taglia del DNA	135
9.1.5	<i>PFGE</i>	135
9.1.6	Sodium dodcyl polyacrylamide gel electrophoresis	135
9.1.7	Concentrazione di gel	135
9.2	Amplificazione di sequenze di RNA e DNA	135
9.2.1	Panoramica	135
9.2.2	Polymerase chain reaction	136
9.2.3	Temperatura di annealing	136

9.2.4	PCR mutagenica	136
9.2.5	Amplificazione basata su PCR di RNA attraverso cDNA	136
9.3	Clonaggio di DNA	136
9.3.1	Panoramica	136
9.3.2	Isolamento dei plasmidi	137
9.3.3	Enzimi di restrizione	137
9.3.4	Strategie per il clonaggio genico	137
9.3.5	Mutagenesi sito-diretta	138
9.3.6	Libreria genica	138
9.4	Manipolazione genomica	138
9.4.1	Panoramica	138
9.4.2	Inserimento di un trasposone	139
9.4.3	Genomi delle piante	139
9.4.4	Ricombinazione omologa	139
9.5	Identificare la composizione di molecole biologiche	139
9.5.1	Sequenziamento del DNA	139
9.5.2	Sequenziamento delle proteine	140
9.5.3	BLAST	141
9.6	Identificazione di specifiche molecole di DNA	141
9.6.1	Panoramica	141
9.6.2	Ibridazione southern blot	141
9.6.3	DNA fingerprint	141
9.6.4	Ibridazione di colonie	142
9.6.5	Cariotipo	143
9.6.6	Fluorescent in situ hybridization	143
9.6.7	Spectral karyotyping	143
9.6.8	Array comparative genomic hybridization	143
9.7	Identificazione di specifiche molecole di RNA	143
9.7.1	Panoramica	143
9.7.2	Ibridazione northern blot	144
9.7.3	Microarray per profilo trascrittomico	144
9.7.4	Trascrizione di un gene reporter	144
9.8	Identificazione di specifiche proteine	144
9.8.1	Ibridazione western blot	144
9.8.2	Stable isotope labelling with amino acid in culture	145
9.9	Identificazione di interazioni tra molecole	145
9.9.1	Panoramica	145
9.9.2	Co-purificazione	145
9.9.3	Co-immunopurificazione	145
9.9.4	Analisi CLIP	146
9.9.5	Electrophoretic mobility shift assay	146
9.9.6	DNA footprinting	146
9.9.7	Colocalizzazione proteica attraverso microscopia a fluorescenza	146
9.10	Sequenziamento di molecole di DNA e RNA attraverso nanopori	147
9.10.1	Panoramica	147
9.10.2	Nanopori	147
9.10.3	Sequenziamento a nanopori	147

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.

1.1.1.1 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.

1.1.1.2 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

1.1.2.1 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.

1.1.2.2 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.

1.1.2.3 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 nucleoide

Nei procarioti, in assenza di nucleo il DNA si organizza in un nucleoide. 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 nucleoide 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.

1.3.1.1 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.

1.3.1.2 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.

1.3.1.3 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.

1.3.2.1 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.

1.3.2.2 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.

1.3.2.3 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

1.3.2.4 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.

1.3.2.5 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 dell'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*.

1.5.2.1 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.

1.5.2.2 Imitation switch

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

1.5.2.3 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.

1.5.2.4 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

1.7.2.1 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.

1.7.2.2 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.

1.7.2.2.1 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 house-keeping, 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.

1.7.3.1 Non corretta disattivazione di X durante la gastrulazione

1.7.3.1.1 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.

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

1.7.4.1.1 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.

1.7.4.1.2 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.

1.9.2.1 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.

1.9.2.2 Esempi di centromeri

1.9.2.2.1 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.

1.9.2.2.2 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.

1.9.2.2.3 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.

1.9.3.1 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 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.

2.1.1.1 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.

2.1.1.1.1 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.

2.1.1.1.2 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.

2.1.1.1.3 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.

2.1.1.1.4 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 iniziatrix 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.

2.2.1.1 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

2.2.2.1 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.

2.2.2.2 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.

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

2.3.2.2 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.

2.3.2.3 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

2.4.1.1 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.

2.4.1.2 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 $5' - 3'$.

2.4.1.3 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 a 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.

2.5.1.1 Regolazione dello svolgimento di *OriC*

2.5.1.1.1 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.

2.5.1.1.2 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*.

2.5.1.1.3 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.

2.5.2.1 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 *G₁* e attivate in *S*. Le origini non possono essere riutilizzate fino alla rifezione nella fase *G₁* successiva.

2.5.2.1.1 Selezione Durante la fase *G₁* *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*.

2.5.2.1.2 Attivazione Durante la fase *S* 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 stallone delle forcelle viene attivata la risposta al danno del DNA e l'entrata in mitosi è bloccata fino alla correzione degli errori.

2.5.2.1.2.1 *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*.

2.5.2.2 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.

2.5.3.1 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.

2.5.3.2 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.

2.6.1.1 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

2.6.2.1 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.

2.6.2.2 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 del nucleotide in arrivo, il pollice mantiene il dsDNA allungato ma non contribuisce alla reazione di polimerizzazione.

2.6.3.1 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 Mg^{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.

2.6.4.1 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.

2.6.4.2 Cause di errori

2.6.4.2.1 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: ($A \leftrightarrow C \vee T$, $G \leftrightarrow C \vee T$).

2.6.4.2.2 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.

2.6.5.1 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.

2.6.5.1.1 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 l'overhang in 3'OH o riempirlo in 5' per rendere blunt le terminazioni di DNA trattato con enzimi di restrizione.

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

2.6.5.2 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.

2.6.6.1 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.

2.7.2.1 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 all'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

2.11.2.1 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.

2.11.2.2 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 genoma 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 RNA 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'.

3.1.1.1 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.

3.1.1.2 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 close 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.

3.1.1.3 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.

3.2.1.1 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.

3.2.1.1.1 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.

3.2.1.2 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.

3.2.1.2.1 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.

3.3.1.1 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.

3.3.1.1.1 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.

3.3.4.1 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*.

3.3.4.2 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 fosfodiester 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 $\gamma\beta$ -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

3.5.4.1 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 ristrutturazione dell'organizzazione cromatinica corretta potrebbe prevenire l'occupazione di siti TATA critici nelle regioni codificanti da *TBP* e il reclutamento di RNA polimerasi II.

3.5.4.2 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.

3.6.1.1 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.

3.6.1.2 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.

3.6.1.2.1 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

3.6.2.1 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.

3.6.2.2 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

3.6.2.2.1 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.

3.6.2.2.2 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.

3.6.2.2.3 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 poliadenilazione.

3.6.2.3 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.

3.7.1.1 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.

3.7.1.1.1 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.

3.7.1.1.2 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.

3.7.6.1 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.

3.8.1.1 Homeodomain

Il homeodomain è un comune dominio eucariote legante il DNA. È monomero. 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 attivatrici. 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.

3.8.6.1 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

3.9.2.1 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,

3.9.2.2 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.

3.10.1.1 Analisi dei mutanti complementari: crescita senza lattosio

3.10.1.1.1 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*.

3.10.1.1.2 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*.

3.10.2.1 Mutazioni di *LacI*

Si riconoscono 3 categorie di mutanti di *LacI*.

3.11. L'OPERONE TRIPTOFANO *TRP* IN *E. COLI*

3.10.2.1.1 *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.

3.10.2.1.2 *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.

3.10.2.1.3 *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.

3.10.3.1 *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

3.12. REGOLAZIONE DELLA TRASCRIZIONE DA PARTE DI RIBOSWITCHES DEL TRASCRITTO

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 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.

3.11.2.1 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.

3.11.2.1.1 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.

3.11.2.1.2 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.

3.12.1.1 Bassi livelli di adenina

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

3.12.1.2 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

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

3.13.3.1 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.

3.13.3.2 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.

3.13.4.1 Controllo temporale dell'espressione dei geni λ

3.13.4.1.1 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 .

3.13.4.1.2 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 .

3.13.4.1.3 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.

3.13.4.2 Regolazione della trascrizione di *N* e *Q*

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

3.13.4.2.1 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*.

3.13.4.2.2 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.

3.13.4.2.2.1 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.

3.13.4.2.2.2 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 .

3.13.5.1 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.

3.13.5.1.1 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.

3.13.5.1.1.1 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 . 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.

3.13.5.1.2 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} .

3.13.5.1.2.1 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 .

3.13.5.1.2.2 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.

3.13.5.1.2.3 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.

3.13.5.1.2.4 Coinvolgimento di O_L La formazione di loop tra O_R e O_L da parte dell'ot-tamero 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.

3.13.6.1 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.

3.13.6.1.1 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.

3.13.6.1.2 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.

3.13.6.1.3 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

3.14. REGOLAZIONE DELLA TRASCRIZIONE DA SISTEMI DI TRASDUZIONE DEL SEGNALE A DUE COMPONENTI

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.

3.13.7.1 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.

3.15.1.1 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.

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

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

3.15.1.1.2 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.

3.15.1.2 Risposta ai livelli di azoto e carbonio

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

3.15.1.2.1 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.

3.15.1.2.2 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.

3.15.2.0.1 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.

3.15.2.0.2 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.

3.16.1.0.1 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.

3.16.1.0.2 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 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

4.2.1.1 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

4.2.1.1.1 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.

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

4.2.1.1.2.1 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.

4.2.1.1.2.2 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.

4.2.1.2 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.

4.2.1.2.1 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.

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

4.2.1.2.1.2 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

4.2.2.1 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.

4.2.2.2 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 guasina 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.

4.3.2.1 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.

4.4.2.1 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.

4.4.2.2 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.

4.4.2.2.1 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*.

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

4.4.2.3 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.

4.4.2.4 Poli-adenilazioni alternative

4.4.2.4.1 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.

4.4.2.4.2 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.

4.4.2.4.2.1 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.

4.5.1.1 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 congettura 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.

4.5.2.1 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.

4.5. RNA SPLICING

4.5.2.1.1 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.

4.5.2.1.2 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 *7 Sm* 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.

4.5.2.1.2.1 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.

4.5.2.1.2.2 *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'.

4.5.2.1.3 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.

4.5.2.1.4 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.

4.5.2.2 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.

4.5.2.2.1 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-trasporarsi nel proprio allele nel meccanismo di homing.

4.5.2.2.1.1 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.

4.5.2.2.1.2 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.

4.5.2.2.1.3 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.

4.5.2.2.2 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.

4.5.2.2.2.1 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.

4.5.2.3 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.

4.5.2.4 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.

4.6.1.1 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.

4.6.1.1.1 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*.

4.6.1.1.2 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.

4.6.1.1.3 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.

4.6.2.1 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.

4.6.2.2 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.

4.7.1.1 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.

4.7.1.2 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.

4.8.1.1 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.

- Deamminazione 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.

4.9.1.1 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*.

4.9.1.1.1 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 la 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 dal 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.

4.11.1.1 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.

4.11.1.2 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.

4.11.1.3 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'.

4.11.3.1 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.

4.12.1.1 *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.

4.12.1.1.1 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*.

4.12.1.1.2 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.

4.12.1.2 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

5.1 Panoramica degli RNA regolatori

Gli RNA regolatori controllano molti processi biologici in tutti i regni della vita.

5.1.1 Principi fondamentali

Per tutti gli RNA regolatori:

- Il trascritto primario è processato per formare la molecola funzionante.
- Utilizzano accoppiamento di basi con i target di RNA o DNA.
- Interagiscono spesso con altre componenti con altre proteine per permettere la loro funzione. Lo fanno aumentando la loro funzionalità o trasportandole.

5.1.2 Iterazioni tra le basi

Le interazioni tra le basi con altre molecole di RNA o DNA con gli RNA regolatori possono avere diversi effetti:

- Possono distruggere il legame della molecola legata con una proteina modificandone la struttura tridimensionale.
- Possono cambiare la struttura del RNA.
- Possono promuovere il legame con proteine.

5.1.3 Codifica

Gli RNA regolatori possono essere codificati in diversi modi rispetto al RNA target.

- Nella maggior parte dei casi si trovano codificati su una regione di DNA separata: *trans-encoded*.
- Si possono trovare sullo stesso filamento di DNA ma in anti-senso rispetto al target.
- Si possono trovare come parte della stessa sequenza target *cis-encoded*.

5.2 Piccoli RNA batterici

Gli RNA regolatori batterici sono sintetizzati come trascritti lunghi tra i 100 e i 300 nt e si dicono small RNA *sRNA*. Alcuni sono ulteriormente processati in frammenti più piccoli, ma la maggior parte agiscono come molecole intatte a differenza di quanto succede negli eucarioti in cui sono molto più processati e piccoli. La maggior parte degli sRNA che fa accoppiamento di basi sono codificati *in trans* e sono prodotti in risposta all'ambiente.

5.2.1 Funzioni

Molti si legano al mRNA vicino al motivo di Shine-Dalgarno, il luogo di legame dei ribosomi e inizio della traduzione: operano un suo controllo negativo. Alcune volte la aumentano promuovendo il legame del ribosoma distruggendo strutture sul mRNA che lo impedivano. Alcuni sRNA influenzano la degradazione del RNA target reclutando ribonucleasi. Si nota come la flessibilità strutturale degli RNA regolatori permette il legame di altre molecole rispetto a RNA e DNA come proteine *RNP* e metaboliti per la formazione di riboswitches, molecole che legano il RNA per impedire la traduzione.

5.2.2 Traduzione del trascritto regolata dal ferro

Il sRNA *RyhB* è lungo 90 nt in *E. coli* e reprime la traduzione di enzimi conservatori e utilizzatori del ferro quando i livelli di ferro cellulare sono bassi.

5.2.2.1 Bassi livelli di ferro

Questo RNA interagisce con il trascritto in diversi modi: impedisce il legame del ribosoma con certi mRNA e promuove il legame con quelli codificanti prodotti necessari in condizioni di basso ferro. Recluta inoltre la ss endonucleasi *RNAasi E* degradando il mRNA obiettivo.

5.2.2.2 Alti livelli di ferro

Con alti livelli di ferro i trascritti degli enzimi conservatori e utilizzatori del ferro sono tradotti o stabili e la traduzione di prodotti necessari in condizione di basso livello di ferro è impedita dalla struttura del mRNA secondaria.

5.2.3 Livelli di complementarietà

Il sRNA può interagire con multipli RNA target a diversi livelli in base al livello di complementarietà. Si nota come la complementarietà di sRNA *trans-encoded* è limitata a 10-20 bp e tipicamente una corta regione di coppie è critica e questo filamento è conservato per un certo livello.

5.2.3.1 Proteina chaperon del RNA

La proteina chaperone del RNA *Hfq* aiuta il sRNA a trovare il proprio target all'interno dei migliaia di trascritti della cellula. È un anello proteico esamerico e il legame con *Hfq* al motivo 5'-AAYAAYAA-3' porta insieme le molecole di sRNA e mRNA per promuovere l'accoppiamento tra le basi. Le due molecole spesso si legano a parti diverse di *Hfq*. Quando i livelli di *Hfq* sono più bassi del affinché avvengano tutte le interazioni sRNA compete per *Hfq* e diventa parte del sistema regolatorio.

5.2.3.2 mRNA esca

mRNA decoy o esca può antagonizzare il sRNA in quanto si lega a una molecola di sRNA impedendo la sua azione al mRNA target.

5.2.3.2.1 Chitobiosi Si nota come alti livelli di chitobiosi, un dimero di glucosammine legate β -1,4 induce la trascrizione di un RNA esca *chbB* che lega al sRNA *ChiX* che libera il mRNA obiettivo e causa la sua degradazione permettendo la traduzione dell'operone policistronico del chitobiosi.

5.3 *sRNA* eucarioti: *miRNA*, *siRNA* e *piRNA*

Gli RNA piccoli negli eucarioti sono lunghi circa 22nt e sono derivati da trascritti più lunghi. Si associano con la famiglia *Argonauto* di elicasi-RNAasi che facilitano l'interazione con l'obiettivo.

5.3.1 Classi

5.3.1.1 MicroRNA

I *miRNA* sono derivati da filamenti primari endogeni creati da RNA polimerasi II e sotto-regolano gli RNA citoplasmatici attraverso degradazione degli mRNA e repressione traduzionale. Sono più di 1000 quelli codificati dal genoma umano e regolano $\frac{1}{3}$ dei geni che codificano proteine. Reclutano una proteina argonauto che in caso di perfect hit porta alla degradazione dell'obiettivo, altrimenti blocca la traduzione e lo trasloca nei P-bodies dove viene degradato. Successivamente avviene un rientro nucleare che porta alla formazione di eterocromatina nel DNA dove è trascritto il RNA bersaglio.

5.3.1.2 RNA piccoli interferenti

I *siRNA* sono derivati da RNA a doppio filamento esogeni come RNA virali e hanno come target RNA per la loro degradazione come meccanismo di difesa cellulare.

5.3.1.3 RNA interagenti con Piwi

I *piRNA* sono derivati da regioni ripetitive del genoma o trasposoni codificanti troncati e sotto-regolano la trascrizione degli elementi trasponibili.

5.4 La famiglia di proteine *Argonauto*

Tutti gli sRNA eucarioti svolgono la loro funzione associati a proteine Argonauto che si lega ad essi per facilitare la loro interazione con gli mRNA obiettivo.

5.4.1 Classi

5.4.1.1 Proteine Argonauto

Le proteine argonauto sono coinvolte nei meccanismi di miRNA e siRNA in animali, piante e funghi.

5.4.1.2 Proteine Piwi

Le proteine *Piwi* sono coinvolte nei meccanismi di piRNA.

5.4.1.3 Proteine *WAGO*

Le proteine *WAGO* si trovano in *C. elegans*.

5.5 Processamento di sRNA eucarioti

5.5.1 Il pathway di microRNA

5.5.1.1 Processamento

Il trascritto *pri-miRNA* a singolo filamento si piega in una struttura a stem-loop. Questi a ds sono rimossi dal *pri-miRNA* dal "complesso microprocessore" nucleare *Drosha+DGCR8* che produce un pre-miRNA lungo tra i 60 e i 70nt a hairpin con un 3'-OH e un 5' monofosfato. Il hairpin è esportato dal nucleo attraverso esportine ed è caricato nel complesso *RISC* contenente *Dicer-Argonata* citoplasmatico. Il Dicer taglia il pre-miRNA per produrre un 22bp *miR* : *miR**.

5.5.1.2 Effetti

Si nota come un complesso *miRNA-RISC* possa avere come target centinaia di copie di mRNA. Inoltre se il contatto tra *miRNA-RISC* è debole e il target mRNA non è tagliato può subire un blocco della traduzione anche quando iniziata. Il ribosoma e il mRNA viene incluso nei corpi di processamento citoplasmatici *P-bodies* dove il ribosoma si dissocia e il mRNA è ultimamente degradato. Il miRNA si può trovare in introni, esoni e sequenze codificanti o non-codificanti. Da un singolo *pri-miRNA* possono essere prodotti diversi miRNA: in *C. elegans* ne vengono prodotti 7.

5.5.1.3 Caricamento su Argonata

Una volta che è generato il miRNA viene caricato sull'Argonata del complesso *RISC* che include Dicer e una proteina legante dsRNA *TRBP* (TAR RNA binding protein). Un filamento guida o attivo *miR* è mantenuto, mentre l'altro, il filamento passeggero viene separato e tagliato dall'attività elicastica ed RNAasica dell'Argonata nel processo di sorting. Il filamento guida è coinvolto nel silenziamento o degradazione del RNA target.

5.5.1.4 il cammino dei piccoli RNA interferenti

5.5.1.5 Processamento

I *siRNA* sono derivati da dsRNA da fonti esogene e sono coinvolti nella difesa cellulare attraverso RNA interference. Non si trova un passaggio di rottura nel nucleo e *Drosha* non è coinvolta: Dicer taglia il dsRNA attraverso tagli sequenziali ogni 22bp. Sono prodotte le molecole di *siR* : *siR**.

5.5.1.6 Caricamento su Argonata

Una volta generato la molecola di siRNA viene caricata sulla proteina Argonata: un filamento è mantenuto mentre l'altro rimosso dall'attività elicastica e RNAasica di Argonata. Il filamento mantenuto è coinvolto nel silenziamento del RNA target.

5.5.1.7 Confronto tra *miR* : *miR** e *siR* : *siR**

Si nota come entrambe siano simili: entrambe possiedono 3'-OH e 5' monofosfati e possono subire modifiche post-trascrizionali come metilazioni che aumentano la loro stabilità intracellulare. *siR* : *siR** sono tipicamente completamente complementari e formano dimeri perfetti per regolare un mRNA target specifico, mentre le altre non lo sono e possono regolare target diversi.

5.5.2 Il pathway di piwi interacting RNA

Gli elementi trasponibili *TE* sono componenti strutturali principali dei genomi eucarioti, ma la loro mobilitazione ha tipicamente effetti negativi sul genoma ospite. Per contrastarli le cellule hanno sviluppato meccanismi genetici ed epigenetici per tenerli silenziati. Uno di questi coinvolge il complesso *Piwi-piRNA* che reprime i *TE* rompendo il loro trascritto nel citoplasma e direzionando specifici rimodellatori della cromatina ai loci *TE* nel nucleo. La maggior parte degli RNA interagenti con Piei sono derivati da cluster di piRNA genomici.

5.5.2.1 Processamento

I piRNA sono trascritti dai cluster e processati fino a raggiungere la lunghezza di 24-30nt attraverso un meccanismo a ping pong o amplificazione. Ogni trasposone inserito in orientamento inverso nel cluster a piRNA può dare origine a piRNA anti-senso. I piRNA anti-senso sono incorporati in una proteina Argonauta Piwi e direzionano la sua attività RNAasica attraverso il trascritto trasposone di senso. Il prodotto della rottura è legato da un'altra proteina Piwi e accorciato a dimensione di piRNA. Il piRNA senso è utilizzato per rompere cluster di piRNA trascritti per generare più piRNA anti-senso. Alla fine il complesso di piRNA anti-senso e Piei si muovono nel nucleo per reprimere i trasposoni attraverso metilazione del DNA e modifica degli isoni delle sequenze promotrici o codificanti.

5.5.3 Gli enzimi Drosha e Dicer RNAasi III

Gli enzimi della famiglia delle ds RNAasi III sono coinvolti nel processamento di miRNA e siRNA in quanto tagliano il dsRNA. Si dividono in tre classi:

- Enzimi di classe I come RNAasi III con un dominio catalitico e agiscono come dimeri.
- Enzimi di classe II come Drosha con due domini catalitici e agiscono come monomeri.
- Enzimi di classe III come Dicer con due domini catalitici e agiscono come monomeri.

5.5.3.1 Drosha

Drosha determina la lunghezza dei frammenti tagliati attraverso la proteina accessorio *DGCR8* che si lega a 11bp dalla base e forma con Drosha il complesso microprocessore.

5.5.3.2 Dicer

Dicer determina la lunghezza dei frammenti tagliati assicurando la terminazione 3' del RNA con il dominio *PAZ* e tagliando a una certa distanza da esso.

5.5.3.2.1 Dominio *PAZ* Questo dominio determina la lunghezza del dsRNA tagliato a 22nt.

5.5.3.3 Argonauta

Le proteine Argonauta possiedono 4 domini:

- *PAZ* si lega alla terminazione 3' del dsRNA legato.
- *Mid* interagisce con la terminazione 5' del dsRNA.
- *Piwi* lega 2 Mg^{2+} e taglia il mRNA target ss.
- *N*.

I primi tre insieme orientano il RNA guida legato per facilitare la scansione di mRNA target.

5.6 Silenziamento genico da parte di RNA eucarioti

I siti di legame per miRNA si trovano tipicamente in 3' *UTR* del mRNA target e possono essere l'obiettivo di diversi *miRISC*. Questi tipicamente si accoppiano inizialmente attraverso una sequenza 5' di 2-8 nucleotidi "seed" mentre la terminazione 3' del miRNA rimane strettamente legata al dominio *PAZ*. L'accoppiamento tra miRNA e il target è tipicamente imperfetto con mismatch a posizioni 10 e 11, il dominio *PIWI* rimane inattivo e non li taglia: questo legame causa repressione traduzionale. Quando l'accoppiamento è completo avviene una rottura del mRNA, enzimi di deadenilazione 3' e decapping 5' vengono reclutati e comincia la degradazione del mRNA. *siRISC* si può legare lungo la sua intera lunghezza, rilascia il dominio *PAZ* inducendo un cambio conformazionale che attiva l'attività RNAasica del dominio *PIWI* che rompe e degrada il mRNA.

5.7 Ruolo della difesa virale di *sRNA* batterici, eucarioti e di archea

I virus hanno sviluppato meccanismi per contrastare il silenziamento del RNA: il virus carnation ringspot possiede una proteina *p19* che lega a una molecola di 21 nucleotidi di siRNA e impedisce che venga incorporata nel complesso *RISC*.

5.8 Regolazione mediata da RNA in *cis*

Nella cellula sono attivi RNA regolatori più grandi e complessi degli sRNA. Per esempio gli RNA possono essere repressivi dei livelli di amminoacidi e reprimere o attivare la traduzione in accordo a tale informazione.

5.8.1 Sintesi della amminoacil tRNA sintetasi

I tRNA controllano la sintesi della propria amminoacil tRNA sintetasi *aaRS*: quando abbastanza di essa è presente i tRNA carichi si legano in *cis* al mRNA del proprio enzima *aaRS* e promuovono la terminazione della sua trascrizione. Quando non è abbastanza presente i tRNA scarichi si legano al mRNA dei propri *aaRS* cambiando la sua conformazione e permettendo alla RNA polimerasi di leggerlo.

5.8.2 Riboswitches

I riboswitches si trovano principalmente nei batteri e possono controllare trascrizione, traduzione e splicing. Possiedono due domini principi: l'aptamero e l'effettore. Il metabolita si lega all'aptamero e induce un cambio conformazionale all'effettore che ha un effetto sull'espressione genica. L'effetto varia per promuovere o impedire la trascrizione o traduzione. I metaboliti sono vitamine, ioni, zuccheri, purine e altri, oltre a diversi stati come la fosforilazione.

5.9 RNA regolatori leganti proteine

Gli RNA leganti proteine possono avere un gran numero di effetti biologici attraverso molti meccanismi, uno dei quali è la titolazione di una proteina lontano dal RNA target. Per esempio il *CsrB* RNA non codificante batterico (carbon storage regulator B) possiede multipli motivi leganti *GGA* che possono legarsi e sequestrare il repressore traduzionale *CsrA* che quando libero si lega al mRNA e blocca la traduzione durante il metabolismo del glicogeno. Un altro esempio è il RNA *6S* che mimica il DNA in una bolla di trascrizione aperta: quando i livelli di *NTP* sono bassi il *6S*RNA compete con il promotore del DNA per la RNA polimerasi e il suo fattore σ sequestrandola. Quando i livelli sono alti il *6S* viene trascritto causando il rilascio della RNA polimerasi sequestrata.

5.10 RNA non codificanti lunghi intergenici

Può avvenire anche la trascrizione di regioni non codificanti proteine, creando trascritti molto lunghi o lunghi RNA non codificanti intergenici *lincRNA*. Un esempio è il RNA *Xist* coinvolto nella disattivazione del cromosoma *X*: il trascritto processato copre il cromosoma *X* che lo ha prodotto e recluta il complesso polycomb che silenzia il cromosoma *X* attraverso metilazione di *H3K9* e di *H3K27*.

Capitolo 6

Traduzione

6.1 Panoramica della traduzione

Si intende per traduzione del RNA la produzione della proteina dall'informazione contenuta in un mRNA, mentre per sintesi proteica la polimerizzazione degli amminoacidi a formare una proteina. Questi processi sono un processo complesso ed altamente conservato che coinvolge diverse componenti: mRNA, amminoacidi, tRNA, amminoacil-tRNA sintetasi, ribosomi e rRNA. Questo processo è impegnativo per la cellula in quanto richiede molta energia e risorse nutrizionali: in una cellula di lievito che si divide ogni $90min$ le proteine devono essere sintetizzate per la crescita e divisione, pertanto vengono prodotti 33 ribosomi al secondo.

6.1.1 Ribosomi

Il ribosoma è formato da:

- 4 rRNA sintetizzati negli eucarioti da RNA polimerasi I e III.
- 80 proteine strutturali i cui geni sono parti del *RP regulon* (ribosomal protein).
- 50% dell'attività della RNA polimerasi II è dedicata alla produzione di proteine ribosomali.
- 236 proteine sono necessarie per costruire un ribosoma attivo e questi geni fanno parte del *Ribi regulon*.

La sintesi delle proteine varia negli eucarioti da 2 (crescita lenta) a 6 (crescita veloce) amminoacidi al secondo, mentre nei procarioti tra i 10 e i 20. La velocità può essere aumentata dall'attività di multipli ribosomi (da 2 a 100) su un mRNA singolo che formano il polisoma. Si definisce pertanto ribosoma il macchinario che opera la traduzione, il monosoma 1 ribosoma attaccato a 1 mRNA traduzionalmente attivo e polisoma un numero di ribosomi legati a 1 mRNA traduzionalmente attivi.

6.1.2 Produzione e traduzione di mRNA

Si nota come sia la produzione che la traduzione di mRNA avvengono in direzione 5'-3'. Inoltre avvenendo la traduzione nel citoplasma si rende necessario negli eucarioti un esporto nucleare degli mRNA.

6.1.3 Confronto tra trascrizione e traduzione

6.1.3.1 Procarioti

La trascrizione totale del RNA in una cellula di *E. coli* viene svolta da 1000 fino a 10 000 RNA polimerasi e procede a velocità massimale di $40-80 \frac{nt}{sec}$. La traduzione di *E. coli* ha velocità di $20 \frac{aa}{sec}$ si nota come i due tassi sono quasi uguali. Infatti se la traduzione fosse più veloce della trascrizione i ribosomi colliderebbero con la RNA polimerasi nei procarioti, dove i processi possono avvenire simultaneamente. Inoltre i ribosomi sono importanti per trascrizione veloce: sembrano impedire alla RNA polimerasi di fare backtrack e pause, creando pertanto un accoppiamento inverso tra traduzione e trascrizione.

6.1.3.2 Eucarioti

La trascrizione in cellule mammifere ha tassi di allungamento simili a quelli misurati in *E. coli* tra i 50 e i $100 \frac{nt}{sec}$. Viene suggerito che queste lunghezze di trascrizione rapida siano intervallate con lunghe pause portando a un tasso medio un ordine di grandezza inferiore rispetto a *E. coli*. L'allungamento della traduzione in cellule mammifere è anch'esso più lento: in condizioni ottime $6 \frac{nt}{sec}$.

6.2 Ribosoma

6.2.1 Scoperta

Nel 1941 viene scoperta la correlazione tra RNA e livelli di proteine. Nel 1943 attraverso centrifugazione ad alta velocità viene identificata una frazione che contiene la maggior parte del RNA nel citoplasma e poi identificata come il ER ruvido. Nel 1955, attraverso microscopia elettronica vengono osservate granuli densi liberi nel citoplasma o legati al ER che contengono il RNA. Nel 1958 viene scoperto il ribosoma. Si nota come i ribosomi legati alla membrana ER sintetizzano proteine per l'esporto cellulare o inserzione in membrane, mentre quelli liberi nel citoplasma le sintetizzano per tutte le attività citosoliche e nucleari.

6.2.2 Composizione

I ribosomi sono grandi da 2.5 fino a $4mDa$ e sono formati per $\frac{2}{3}$ da rRNA e per $\frac{1}{3}$ da proteine. Si dividono nella subunità minore che decifra il mRNA e media l'interazione tra mRNA e tRNA e la subunità maggiore che catalizza la formazione di legami peptide tra gli amminoacidi e possiede un tunnel attraverso cui esce il peptide di crescita. L'interfaccia tra le subunità è importante per i movimenti di tRNA e mRNA nel ribosoma. Un errore avviene una volta ogni 1000-10 000 monomeri. I fattori di traduzione, spesso *GTPasi* si associano con i ribosomi per aiutare la traduzione, composta da quattro fasi principali: iniezione, allungamento, terminazione, separazione delle subunità ribosomiali e riciclo.

6.2.2.1 Confronto tra ribosomi eucarioti e procarioti

I ribosomi eucarioti e batterici sono funzionalmente e strutturalmente conservati ma differiscono nella composizione proteica.

Ribosoma batterico (70S):

- Subunità maggiore (50S):
 - rRNA 5S lungo 120nt.
 - rRNA 23S lungo 2900nt.
 - Circa 34 proteine *L1-L34*.
- Subunità minore (30S):
 - rRNA 16S lungo 1540nt.
 - Circa 21 proteine *S1-S21*.

Ribosoma eucariote (80S):

- Subunità maggiore (60S):
 - rRNA 5S lungo 120nt.
 - rRNA 5.8S lungo 160nt.
 - rRNA 28S lungo 4700nt.
 - Circa 49 proteine.
- Subunità minore (40S):
 - rRNA 18S lungo 1900nt.
 - Circa 33 proteine.

6.2.2.2 Ruolo del rRNA

La struttura del ribosoma è determinata da quella del proprio rRNA. In una subunità le proteine estendono braccia nelle regioni di rRNA, solitamente altamente basiche e si pensa aiutino con l'impacchettamento dei backbone fosfati degli rRNA negativamente carichi. A causa dell'alto bisogno dei ribosomi i geni degli rRNA sono i più intensamente trascritti: il 90% di tutto il RNA è rRNA e viene organizzato in vettori. Gli rRNA processati da pre-rRNA vengono divisi in domini: il 16S (18S negli eucarioti) possiede tre domini principali e uno minore, mentre il 23S (28S negli eucarioti) possiede 6 domini. Questi domini sono distinti nella piccola subunità, mentre nella grande sono intrecciati. La subunità maggiore possiede inoltre sia in batteri che eucarioti il 5S rRNA e solo negli eucarioti il 5.8S rRNA. Gli RNA sono cruciali per l'attività dei ribosomi in quanto la regione decodificante del rRNA 16S media l'interazione tra il tRNA e mRNA e tra mRNA e il ribosoma, mentre il rRNA 23S nel centro della peptidil-trasferasi interagisce con il tRNA.

6.2.3 Il ribosoma come un ribozima

Gli rRNA catalizzano la sintesi proteica. Si nota come le componenti dei ribosomi a RNA erano presenti prima di quelle proteiche che sono state reclutate al ribosoma più tardi nell'evoluzione. Si può pertanto considerare il ribosoma come un ribozima coperto da proteine. Questo si deduce sia funzionalmente: il rRNA 23S estratto dalla subunità maggiore 50S dei procarioti senza proteine attaccate ad esso può catalizzare il legame tra i peptidi con bassa efficienza che strutturalmente: i siti rRNA 16S e peptidil-trasferasi (rRNA 23S) sono formati da rRNA e non si trova vicino alcuna proteina ribosomiale.

6.2.3.1 Modifiche ai nucleosidi degli rRNA

Gli rRNA eucarioti e procarioti contengono nucleotidi alternativi a causa dei modifiche post-trascrizionali enzimatiche dei quattro nucleotidi canonici.

6.2.3.2 Proteine ribosomiali

Le proteine nel ribosoma supportano il piegamento del rRNA garantendo ad esso e al ribosoma la struttura corretta per l'attività, aumentano l'efficienza del processo di traduzione. Per questi motivi ogni proteina ribosomiale è essenziale per la vitalità della cellula. Sono pertanto altamente conservate. Le proteine sono grandi, globulari e basiche. Si localizzano esternamente ma con code

che protrudono nella struttura del rRNA. I ribosomi eucariotici possiedono 82 proteine, 27 in più rispetto a quelli procarioti. Sono identificate come *rpS*“numero” per quelle presenti nella subunità minore e *rpL*“numero” per quelle nella maggiore.

6.2.4 Assemblaggio dei ribosomi negli eucarioti

Negli eucarioti il primo passo per l'assemblaggio dei ribosomi inizia nel nucleolo, dove rRNA 35S, rRNA 5S, proteine ribosomiali, fattori proteici non ribosomiali e snoRNA vanno a formare il pre-90S. Questo complesso esce dal nucleolo e si divide in pre-60S con gli rRNA 27S e 5S e nel pre-40S con il rRNA 20S. Questi due sono i pre-ribosomi. I pre-ribosomi escono dal nucleo attraverso due complessi del poro nucleare ed entrano nel citoplasma, dove maturano nel 60S con rRNA 25S, 5.8S e 5S e nel 40S con rRNA 18S. DA APPROFONDIRE.

6.2.4.1 Nucleoli

I nucleoli sono ripetizioni di loci di rDNA eucariote clustered trascritto da RNA polimerasi I e III.

6.2.5 Rappresentazione funzionale del ribosoma

Il ribosoma possiede nella subunità minore un canale per il passaggio del mRNA. Dal tunnel poi si formano due fori: il sito A e il sito P. Nel sito A viene decodificato il codice genetico e caricato l'amminoacido corretto. Il sito A finisce nella subunità minore dove comunica con il tunnel di uscita della catena peptidica nascente. Dopo che è stato caricato nel sito A il tRNA si sposta nel sito P dove avviene il legame peptide con la catena peptidica in crescita. Il tRNA perde l'amminoacido e passa al sito E dove viene rilasciato nel citoplasma.

6.3 tRNA e codice genetico

Negli anni 50 Crick ipotizza l'esistenza di una molecola adattatrice, un link fisico tra un codone nel mRNA e l'amminoacido a lui corrispondente. Gli adattatori sono piccoli RNA lunghi tra i 74 e i 94nt o tRNA. Questi decifrano il mRNA e trasportano un amminoacido. Si trova almeno 1 tRNA per ogni amminoacido, per 20 se ne trovano più di 40.

6.3.1 Struttura

La struttura del tRNA possiede quattro regioni di RNA a doppio filamento e tre stem-loop *T*, *A* e *D*. Il tRNA possiede nucleotidi modificati il loop *D* *DHU* prende il nome dalla di-idro-uridina, mentre il loop *T* *TpsiC* possiede ribo-timidina e pseudo-uridina ϕ . Le terminazioni 5' e 3' si accoppiano e formano lo stem accettore, con una coda 3' conservata *CCA* che lega l'amminoacido. Il loop dell'anticodone possiede 3 nucleotidi che si accoppiano con il codone nel mRNA. I tRNA hanno sequenze abbastanza conservate, mentre la struttura lo è latamente. Possiedono molti nucleotidi modificati a causa di modifiche post-trascrizionali enzimatiche che gli danno un identificativo univoco. In particolare l'inosina e le sue varianti modificate e le purine *A* o *G* in posizione 37 che affiancano l'anticodone sono ipermodificate per impedire che la base 37 si accoppi con il codone nel mRNA e per produrre un loop stabile *AC*.

6.3.2 Sintesi

Nel nucleo si forma il trascritto iniziale dei tRNA che negli eucarioti possiede un introne. Successivamente vengono processate le terminazioni e aggiunta la coda 3' *CCA*. Viene trasportato nel citoplasma dove si trova nella forma matura.

6.3.3 Codoni

Un codone tripletta che specifica per un singolo amminoacido si dice sense codon (61), mentre uno che non lo fa è detto stop codone o nonsense codone (3). Si noti come il codice a triplette è il più semplice che può specificare tutti e 20 gli amminoacidi. Viene utilizzato ogni codone possibile, pertanto alcuni amminoacidi sono codificati da più di un codone: in molti casi i primi due nucleotidi sono lo stesso e differisce lo stesso, ma non sempre. Solo i tRNA per metionina e triptofano riconoscono un singolo codone. I diversi tRNA che trasportano lo stesso amminoacido sono detti iso-accettori.

6.3.3.1 Interazione codone anti-codone

Il codone nel mRNA interagisce con l'anti-codone nel *U*-turn del tRNA. La posizione 1 e 2 sul mRNA (da 5' a 3') sono letti da accoppiamenti rigidi con le posizioni 2 e 3 sull'anticodone. Per la posizione 3 del mRNA sono permesse deviazioni o wobble pairing che permette interazioni non canoniche come *G-U*. In alcuni casi con inosina. Pertanto ogni codone non necessita del proprio tRNA: ce ne sono circa 40 per i 61 codoni di senso. I più comuni sono: *U - G*, *I - U*, *I - C*, *I - A*.

6.3.3.2 Codoni rari

Alcuni codoni sono utilizzati meno frequentemente di altri e sono detti codoni rari, codificati da tRNA rari. Il codice genetico è più o meno lo stesso in tutti gli organismi e l'evoluzione ha conservato i codoni in modo che le mutazioni che cambiano l'amminoacido codificato risultano in un amminoacido simile che lo modifica. Ci sono eccezioni come nei mitocondri e in mycoplasma.

6.3.4 Legame con il ribosoma

Il tRNA si lega in sequenza a tre siti nel ribosoma: il sito amminoacile, il sito peptide e il sito di uscita. Il centro di decodifica è l'area tra il basso di A e il canale del mRNA e il centro di trasferimento peptide è all'alto del sito P.

6.4 Amminoacil-tRNA sintetasi

L'amminoacil tRNA sintetasi attacca un amminoacido sul tRNA a 3' *AAC*, ne esistono di 24 tipi per i 20 amminocidi. Una *aaRS* per un amminoacido, come la glicina è definita come *GlyRS*. Il tRNA su cui un certo amminoacido deve essere taricato è identificato da certi marcatori a modifiche nucleosidiche. Si nota come ogni amminoacido possieda una amminoacil-tRNA sintetasi specifica e quello che viene caricato su un tRNA specifico determina la nomenclatura del tRNA con *etRNA^{Met}*. Si dice affine ("cognate") l'amminoacido corretto. I tRNA vengono riconosciuti grazie a caratteristiche di struttura e di sequenza o elementi di identità.

6.4.1 Struttura

La maggior parte delle proteine *aaRS* contengono un sito di amminocilazione e uno di editing. Esistono due classi dell'enzima con siti attivi differenti strutturalmente e che si legano a tRNA diversi strutturalmente.

6.4.1.1 Classe I

Le *aaRS* di classe *I* legano l'amminoacido al 2'-OH del ribosio dell'adenosina, agiscono come monomeri o dimeri e si legano alla fessura minore dello stem accettore.

6.4.1.2 Classe II

Le *aaRS* di classe *II* legano l'amminoacido al 3'-OH del ribosio dell'adenosina, agiscono come dimeri o tetrameri e si legano alla fessura maggiore dello stem accettore.

6.4.1.3 Sito catalitico o di attivazione

In questo dominio avviene il caricamento dell'amminoacido a due passaggi e dipende dall'ATP.

6.4.1.4 Sito di legame per l'anticodone

Questo dominio interagisce con l'anticodone del tRNA e garantisce il legame del tRNA corretto all'amminoacido.

6.4.1.5 Sito di editing

Questo dominio rimuove un amminoacido aggiunto scorrettamente.

6.4.2 Caricamento

Il caricamento dell'amminoacido è accurato, con meno di un errore ogni 10^4 eventi di amminocilazione grazie al controllo qualità del modello a doppio setaccio. L'amminoacido corretto viene scelto in un processo a due passi.

6.4.2.1 Attivazione amminoacilica

Quando deve essere caricato l'amminoacido viene attivato attaccando *AMP* che rilascia pirofosfato e fornisce l'energia. L'amminoacil-adenilato attivato rimane attaccato all'enzima *aaRS*.

6.4.2.2 Trasferimento dell'amminoacido

Dopo l'attivazione l'enzima trasferisce l'amminoacido al 2' o al 3' del ribosio dell'adenosina sulla coda *CCA* del tRNA.

6.4.3 Correzione degli errori

Alcuni amminoacidi sono molto simili strutturalmente come i tRNA ed essendo il ribosoma incapace di distinguere tra i tRNA caricati correttamente o scorrettamente avviene un processo di correzione degli errori da parte della *aaRS*: 10 su 24 compiono editing nei loro siti.

6.4.3.1 Selezione degli amminoacidi - il modello a doppio setaccio

6.4.3.1.1 Primo setaccio Il primo setaccio si trova al sito di attivazione e svolge un'esclusione per dimensione, è grossolano ed esclude amminoacidi troppo grandi.

6.4.3.1.2 Secondo setaccio Il secondo setaccio si trova al sito di editing o di proofreading o idrolitico, è fine e idrolizza gli *aa-AMP* troppo piccoli nel pre-transfer editing.

6.4.3.1.3 Post-transfer editing Se un *aa-AMP* viene incluso nonostante il doppio setaccio avviene post-transfer editing in trans. In questo processo il tRNA mal-etichettato lascia il *aaRS* e fattori idrolitici indipendenti come *YbaK* o *AlaXP* rimuovono l'amminoacido.

6.4.3.2 *aaRS* senza capacità di editing

Solo 10 delle 24 *aaRS* possono editare e quelle di classe II sono le più efficienti. La presenza di siti di editing garantisce un vantaggio della sopravvivenza con migliori condizioni con una più alta proporzione di proteine correttamente sintetizzate. Gli amminoacidi più usati correlano con le *aaRS* con un sito di editing.

6.4.3.2.1 Mal-traduzioni Le *aaRS* senza capacità di editing creano un rischio di mal-traduzione che possono causare la risposta delle proteine non piegate che causa la loro degradazione. Altre cellule tollerano la mal-traduzione con un effetto sulla vitalità visibile solo in condizioni di stress. La mal-traduzione può avere un beneficio adattivo con sopravvivenza in condizioni negative.

6.5 mRNA

Il rRNA è molto stabile e non permette alla cellula di rispondere a condizioni variabili abbastanza velocemente. Si necessita pertanto di un RNA come portatore delle informazioni genetiche con una vita corta e che può essere creato e degradato velocemente: il RNA messaggero. Ha un'abbondanza percentuale bassa rispetto a rRNA 80%, tRNA 15% del 5% e viene rapidamente turned over, con un tempo vitale di minuti nei batteri e dai minuti alle ore negli eucarioti.

6.5.1 Struttura nei procarioti

Nei procarioti il mRNA si presenta principalmente policistronico con una corta 5'-UTR e varie regioni codificanti separate da regioni intercistroniche. Alla terminazione presenta una corta sequenza 3'-UTR.

6.5.2 Struttura negli eucarioti

Negli eucarioti il mRNA si presenta principalmente monocistronico con un cap 5', una lunga 5'-UTR, una regione codificante, una lunga 3'-UTR e una coda poli-A. La 3'-UTR contiene elementi stabilizzatori e di regolazione.

6.6 Ciclo di traduzione

La traduzione coinvolge quattro passaggi.

6.6.1 Iniziazione

Il *AUG* all'inizio del reading frame aperto viene identificato da fattori di iniziazione, dal ribosoma e da un tRNA iniziatore metionina speciale. Iniziazione precoce coinvolge il legame della subunità minore ribosomiale al mRNA e la subunità maggiore alla minore. Questo risulta in un ribosoma legato al mRNA con un tRNA caricato a metionina legato al sito P. Il ribosoma è ora pronto per muoversi lungo il mRNA.

6.6.2 Allungamento

Il fattore di allungamento *Tu* nei batteri, *eEF1A* negli eucarioti carica il successivo tRNA carico nel sito A in base al codone nel mRNA. La formazione del legame peptidico è catalizzata tra l'amminoacido nel sito P e quello nel sito A. La reazione trasferisce il polipeptide in crescita al tRNA nel sito A. Successivamente *EFG* nei batteri e *eEF2* negli eucarioti promuove il movimento del ribosoma sul codone successivo attraverso traslocazione. Questo muove il peptidil-tRNA che era nel sito A nel sito P e porta un nuovo codone nel sito A. Il tRNA che si trova ora nel sito E lascia il ribosoma.

6.6.3 Terminazione e riciclo dei ribosomi

La terminazione avviene quando il ribosoma raggiunge un codone di stop *UAG*, *UAA* o *UGA* che vengono riconosciuti da fattori di rilascio di classe I. Il fattore di rilascio batterico *RF1* riconosce *UAA* e *UAG*, *RF2* *UAA* e *UGA*. Negli eucarioti *eRF1* riconosce tutti i codoni di stop. L'interazione tra codone di stop e *RF1/2* promuove il rilascio del polipeptide. Infine la subunità maggiore e minore si dissociano e rilasciano il rimanente tRNA e mRNA. Il fattore di riciclo *RRF* e *EFG* aiutano i ribosomi a dissociarsi nei batteri.

6.7 Iniziazione della traduzione - caratteristiche comuni a batteri ed eucarioti

Ci sono tre passaggi per l'iniziazione della traduzione raggiunti diversamente in eucarioti e batteri:

- La subunità ribosomiale minore identifica il codone di inizio nel mRNA.
- Un *metionil-tRNA^{Met}* è caricato nel sito P del ribosoma e si accoppia con le basi del codone di inizio.
- La subunità ribosomiale maggiore unisce il complesso.

Il codone di inizio è tipicamente *AUG* e viene decodificato dall'iniziatore *tRNA^{Met}* che differisce in eucarioti *tRNA_I^{Met}* e batteri *tRNA_f^{Met}*, dove *f* denota un gruppo formile. Inoltre diversi fattori di iniziazione *IF* (*GTPasi*) sono coinvolti nel legame del *metionil-tRNA^{Met}* o di *fmetionil-tRNA^{Met}* al sito P di eucarioti e batteri. L'iniziatore batterico ha un mis-match *C-A* nello stem accettore, mentre l'iniziatore *tRNA* eucariotico ha un match *A-U*. Entrambi possiedono tre coppie *G-C* nello stem anticodone, importanti per legare il fattore di allungamento *EF-Tu* (batteri) o *eIF1A* (eucarioti).

6.7.1 Il tRNA iniziale nei procarioti

Sia che in procarioti che in eucarioti la sintesi della proteina inizia con l'inclusione di una metionina *AUG*, *GUG*, *UUG*, ma solo nei procarioti la *Met* iniziale è una *formil-Met*. Dopo il legame al proprio

$tRNA_I$ la metionina viene modificata enzimaticamente con un gruppo formile alla terminazione N formando $fMet-tRNA_f^{Met}$. L'enzima coinvolto è la metionil-tRNA trasformilasi e $fMet$ non può essere utilizzata per l'allungamento delle proteine in quanto il gruppo NH_2 è bloccato.

6.8 Iniziazione della traduzione batterica

Gli mRNA batterici sono tipicamente policistronici: contengono diversi reading frame aperti, ognuno con il proprio codone di start e stop. I codoni di iniziazione possiedono tipicamente una sequenza di Shine-Dalgarno, un tratto a poli-purine che si trova tra i 7 e i 13nt a monte del codone di inizio AUG . Shine-Dalgarno si accoppia con una regione a poli-pirimidine nella terminazione 3' del rRNA 16S batterico, la sequenza anzi-Shine-Dalgarno. L'interazione tra le basi guida l'iniziatore AUG nel sito P del ribosoma.

6.8.1 Riconoscimento della sequenza Shine-Dalgarno

Il complesso di pre-iniziazione formato dalla subunità 30S, mRNA, $fMet-tRNA_f^{Met}$ e IF deve assemblarsi sul codone di inizio AUG . Si deve pertanto distinguere tra i diversi codoni AUG quello di inizio. Mettendo in vitro i ribosomi, mRNA, tRNA e GTP questi formano un complesso che non può allungarsi. Successivamente si tratta con RNA endonucleasi ed esonucleasi e si isola mRNA protetto dai ribosomi. Si sequenzia il frammento trovato e si nota come si trova una sequenza conservata di 7nt $AGGAGGU$ dai 7 ai 13nt a monte di AUG che è il sito di legame del ribosoma o sequenza di Shine-Dalgarno SD . Si nota come il consenso della sequenza anti-Shine-Dalgarno con essa controlla la forza della traduzione.

6.8.2 Assemblaggio del complesso di inizio

La subunità minore del ribosoma 30S si posiziona sul mRNA grazie alla terminazione 3' del rRNA 16S che riconosce la sequenza di SD . Successivamente viene reclutata la subunità maggiore 50S. Si nota come essendo il mRNA procariote poli-cistronico sequenze di Shine-Dalgarno e AUG sono presenti all'inizio di ogni cistrone.

6.8.2.1 Fattori di inizio

Nei procarioti tre fattori di inizio guidano $fMet-tRNA_f^{Met}$ al sito P : $IF1$, $IF2$ e $IF3$.

6.8.2.1.1 $IF1$ e $IF3$ Questi due fattori legano i siti A ed E nella subunità minore in assenza di mRNA o di $fMet-tRNA_f^{Met}$, direzionando il tRNA iniziatore al sito P e impedendo un legame inappropriato con la subunità maggiore.

6.8.2.1.2 $IF2$ Questo fattore di inizio è una $GTPasi$ che idrolizza GTP per fornire l'energia necessaria all'unione delle subunità del ribosoma. L'idrolisi del GTP causa inoltre il suo rilascio.

6.8.2.1.3 $IF1$ e $IF3$ Questi due poi vengono rilasciati quando le subunità si combinano completando l'iniziazione.

6.9 Iniziazione della traduzione eucariotica

Gli mRNA eucarioti codificano tipicamente per una proteina (sono monocistronici). Non si trovano equivalenti eucariotici per la sequenza di Shine-Dalgarno e la subunità minore non si lega direttamente al mRNA e si necessitano di più fattori di inizio. L'iniziazione avviene tipicamente al primo *AUG* del mRNA. Questo può essere inefficiente e causare l'iniziazione al secondo o terzo *AUG*. Il riconoscimento di *AUG* è sensibile al contesto di sequenza: sequenze di Kozak con consenso *A/GXXAUGG*. Si nota come il primo metile non è formilato e la subunità minore si lega al cap 5' del mRNA che diventa il sito di legame per il ribosoma. Successivamente si muove lungo il mRNA al primo *AUG* facendo scanning in una sequenza di Kozak. La sequenza di Kozak scansionata ha un consenso debole e nei vertebrati è *GCC CCAUGG* e a differenza dei procarioti non è il sito di legame del ribosoma *RBS*. Il sito di legame del ribosoma è il cap 5' o *IRES* (internal ribosome entry site).

6.9.1 Processo di iniziazione

6.9.1.1 Complesso del loop chiuso

Il cap 5' e la coda poli-A del mRNA sono coinvolti nell'iniziazione in quanto preparano il mRNA affinché venga scansionato dal ribosoma. Il cap 5' viene legato da *eIF4E* e la coda 3' da *PABP* (polyA binding protein) che interagiscono tra di loro attraverso un complesso di altri fattori di inizio (*4G*, *4A*, *4B*) formando un complesso a loop chiuso. Questo complesso protegge il mRNA dall'attività esonucleasica e potrebbe funzionare anche come un sistema di controllo di qualità per eliminare mRNA non processati o danneggiati impedendo che vengano tradotti in proteine.

6.9.1.2 Reclutamento del complesso di iniziazione

eEF1A e *eEF1* si legano nei siti *A* ed *E* del ribosoma e *met-tRNA_I^{Met}* al sito *P* della subunità minore del ribosoma 40S. Il complesso viene poi reclutato al complesso del loop chiuso attraverso interazioni tra *eIF3* e *eIF4G*, rispettivamente sul complesso di iniziazione e del loop chiuso. La subunità ribosomiale minore, legata a un numero di fattori di inizio forma il complesso di pre-iniziazione 43S e può legarsi al mRNA formando il fattore di pre-iniziazione 48S. *eIF4A*, un'elicasi *eIF4B*, suo attivatore svolgono la struttura secondaria del cap 5' del mRNA e permettono al *Met-tRNA_I^{Met}* nel complesso di pre-iniziazione 48S di fare una scansione per la sequenza di Kozak. *eIF5B* catalizza l'unione della subunità maggiore 60S e tutti i fattori di iniziazione si dissociano. Il complesso di iniziazione 80S formatosi è ora competente per l'allungamento.

6.9.2 Metodi di scansione

6.9.2.1 Scansione dipendente dal cap 5' e da *eIF4F*

Il complesso *eIF4* formato da *eIF4E* + *G* + *A* + *B* si lega al cap 5' attraverso *eIF4E*. Questo recluta il complesso di pre-inizio 43S formando il 48S che scansiona il mRNA per *AUG*. La 5' *UTR* inizia a formare un anello uscendo durante la scansione e quando la subunità 40S raggiunge la sequenza di Kozak e il *AUG* nella sequenza vengono rilasciati i fattori di inizio ad esclusione di *eIF4E-G* che rimangono legati al cap. Viene rilasciato il loop formato dal cap e viene reclutata la subunità maggiore 60S sulla cima di *AUG*.

6.9.2.2 Iniziazione indipendente dal cap 5' e dipendente da *IRES*

L'iniziazione indipendente dal cap 5' viene identificata negli RNA dei virus eucariotici e per certi mRNA. Necessita di una sequenza *IRES* nel 5'-UTR del mRNA. Si trova spesso negli mRNA eucarioti policistronici.

6.9.3 Presenza della metionina all'inizio della proteina prodotta

6.9.3.1 Procarioti

Nei procarioti il gruppo formile viene sempre rimosso dalla prima metionina e spesso è l'intera metionina ad essere rimossa.

6.9.3.2 Eucarioti

Negli eucarioti la metionina iniziale viene rimossa da aminopeptidasi *MAP1*, *MAP2* essenziali per la vitalità, in $\frac{2}{3}$ delle proteine. Viene rimossa quando il secondo amminoacido è di piccola o media dimensione come alanina, cisteina, glicina, prolina, serina, treonina e valina. Non viene rimossa quando il secondo amminoacido è grande come arginina, asparagina, acido aspartico, glutammina, acido glutammico, isoleucina, leucina, lisina o metionina.

6.9.3.3 Rimozione della metionina

La rimozione della metionina permette il suo riciclo in quanto è l'amminoacido più laborioso da sintetizzare. Potrebbe inoltre ridurre l'emivita di una proteina secondo la regola della terminazione *N* o permettere modifiche *N* terminali del secondo amminoacido.

6.10 Allungamento della traduzione

L'allungamento della traduzione è un processo conservato. Si nota come alla fine dell'iniziazione il ribosoma è posizionato al primo *AUG*. *Met-tRNA_f^{Met}* of *Met-tRNA_f^{Met}* localizza al sito *P* e si lega al *AUG*. Ogni amminoacido si attacca alla catena polipeptidica in crescita attraverso un ciclo di:

- Decodifica: entrata di un amminoacido *aa-tRNA^{aa}* nel sito *A*.
- Formazione del legame peptidico tra amminoacidi nel sito *P* grazie al rRNA 23S peptidiltrasferasi.
- Traslocazione: del mRNA:peptidil-tRNA da *A* a *G* lasciando ancora il sito *A* aperto.

6.10.1 Decodifica, legame di un amminoacil-tRNA^{aa} al sito A

Questo processo richiede per i procarioti i fattori di allungamento *EF*: *EF-Tu* (thermo unstable), *Ef-Ts* (thermo stable) e *EF-G* (translocation factor) e *GTP*. Al sito *P* si trova *fMet-tRNA_f^{Met}* che si lega a *AUG* portata da *IF2-GTP*. Nel sito *A* l'entrata di *aa-tRNA^{aa}* è mediata dalla *GTPasi EF-Tu* legata a *GTP* formando il complesso *aa-tRNA^{aa}:EF-TU-GTP*. Il legame tra l'anticodone del complesso e il codone del mRNA avviene nel centro di decodifica del sito *A*, la tasca di rRNA a 16S. Il corretto legame tra codone e anticodone avviene grazie a legami a idrogeno e il segno di un'associazione stabile nel sito *A* è un cambio strutturale nella tasca del rRNA 16S. Successivamente l'idrolisi del *GTP* da parte di *EF-TU* causa il rilascio di *ET-TU-GDP* e se il fit no è buono viene

rimosso anche $aa-tRNA^{aa}$. $Ef-Ts$, un fattore di scambio di GTP ricarica $EF-Tu$ ripristinandolo nello stato carico con GTP rendendolo pronto per caricare il prossimo $aa-tRNA^{aa}$.

6.10.1.1 Il complesso $aa-tRNA^{aa}:EF-Tu-GTP$

$EF-Tu$ è una delle proteine più abbondanti nei batteri, e costituisce il 5% di tutte le proteine. Protegge il legame labile tra il tRNA e il proprio amminoacido quando $aa-tRNA^{aa}$ lascia il proprio $aaRS$. Inoltre $EF-Tu-GTP$ carica $aa-tRNA^{aa}$ rapidamente e con alta fedeltà nel sito A del ribosoma. Infine aiuta a decidere se l'anticodone del $aa-tRNA^{aa}$ è un fit abbastanza buono quando legato al codone.

6.10.1.2 Fit corretto tra codone ed anti-codone

L'interazione tra codone e anti-codone attraverso legami a idrogeno può essere cognate quando è completamente accurata, near-cognate quando ha un mal-appaiamento su una singola base o non-cognate quando ci sono 2 o 3 mal-appaiamenti. Il riconoscimento del cognate stimola un cambio conformazionale nel ribosoma determinando l'accettazione del $aa-tRNA^{aa}$. L'alta fedeltà del passaggio di decodifica è ottenuta attraverso un processo a due passi di accettazione o rifiuto.

6.10.1.2.1 Selezione iniziale basata sul tempo $EF-Tu$ è una $GTPasi$ che usa l'energia fornita dall'idrolisi del GTP per valutare il fit tra codone e anticodone. Prende forma un equilibrio: il ribosoma decide quanto bene $aa-tRNA^{aa}$ portato da $EF-Tu-GTP$ interagisce con il codone nel sito A .

6.10.1.2.1.1 Buon legame Il $aa-tRNA^{aa}$ si lega più a lungo ed è più probabile che avvenga l'idrolisi del GTP .

6.10.1.2.1.2 Cattivo legame $aa-tRNA^{aa}+EF-Tu-GTP$ si separa dal ribosoma prima che possa avvenire l'idrolisi del GTP .

6.10.1.2.2 Proofreading basato sulla struttura La valutazione dell'accoppiamento delle basi tra codone e anti-codone avviene grazie al 16S rRNA attraverso movimenti interni delle due adenosine e della guanina conservate presenti nel centro di decodifica. Quando il 16S rRNA riconosce il legame come corretto avviene un cambio conformazionale nel ribosoma che stimola l'idrolisi del GTP da parte di $EF-Tu$. In questo modo $EF-Tu-GDP$ lascia il ribosoma e l'amminoacido viene incluso nella coda polipeptidica. Se il legame è scorretto $aa-tRNA^{aa}$ lascia il sito A .

6.10.2 Formazione del legame peptide tra amminoacidi nel centro di trasferimento del peptidil

Tra l'ultimo amminoacido alla terminazione 3' del peptidil-tRNA nel sito P e il nuovo amminoacido alla terminazione 3' del $aa-tRNA^{aa}$ nel sito A avviene un trasferimento del poli-peptide dal tRNA al sito P sull'amminoacido di $aa-tRNA^{aa}$ nel sito A . Questo avviene nel sito attivo del ribosoma grazie all'attività di peptidiltrasferasi, un ribozima rRNA 23S. Ora i tRNA si trovano in stati intermedi E/P e P/A .

6.10.2.1 Catalisi mediata da rRNA

Il 23S rRNA o 28S rRNA::5.8S rRNA negli eucarioti porta il *peptidil-tRNA* e *aa-tRNA^{aa}* insieme grazie loop conservati *A* e *P* presenti nei corrispettivi siti e ai nucleotidi conservati in entrambi. La reazione coinvolge un attacco nucleofilo :*N* su un centro deficiente di elettroni $C^{\delta+}$. Vengono utilizzati ioni di magnesio come co-fattori.

6.10.3 Traslocazione del peptidil-tRNA dal sito A al sito P

Si dice traslocazione il movimento di *mRNA:peptidil-tRNA* dal sito *A* al sito *P*, che causa il movimento del ribosoma di 3 nucleotidi lungo il mRNA. Il *tRNA^{aa}* lascia il sito *P* per entrare nel sito *E* da cui esce dal ribosoma. Il prossimo codone viene esposto nel sito *A* vuoto per il prossimo *aa-tRNA^{aa}*. La traslocazione richiede il fattore di traslocazione *GTPasi EF-G+GTP* che si lega al sito *A* vuoto nello stato legato al *GTP*. Il movimento del *mRNA:peptidil-tRNA* dal sito *A* al sito *P* richiede l'energia fornita dall'idrolisi del *GTP* legato a *EF-G*. A questo punto *EF-G-GDP* lascia il sito *A* e viene ricaricato con *GTP* dal fattore di scambio del nucleotide guanina *GEF*. Il *aa-tRNA^{aa}* è reclutato e l'allungamento può continuare. *EF-G-GTP* sposta il peptidil-tRNA dal sito *A* al sito *P* e si nota come il passo di traslocazione è irreversibile dal momento in cui *GTP* viene idrolizzata da *EF-G*. *EF-G* è simile strutturalmente a *EF-Tu:aa-tRNA^{aa}* e causa il movimento in avanti spingendo il tRNA. Durante la traslocazione la testa della subunità 30S ruota di 17° attorno il 16S rRNA causando il movimento in avanti del ribosoma lungo il mRNA. Oltre a far passare i tRNA dagli stati ibridi *E/P* e *P/A* in *E/E* e *P/P* l'idrolisi del *GTP* rompe il contatto tra tRNA ed mRNA permettendo l'uscita dal sito *E*.

6.10.4 Allungamento negli eucarioti

Il processo di allungamento negli eucarioti è simile a quello dei procarioti. Cambiano i fattori di allungamento, che sono detti *eEF* (eukaryotic Elongation Factor):

- *eEF1A* corrisponde a *EF-Tu*.
- *eEF1B* corrisponde a *EF-Ts*.
- *eEF2* corrisponde a *EF-G*.

6.11 Terminazione e reinizio della traduzione

La traduzione continua fino a che il ribosoma incontra un codone di stop nel mRNA: *UAA*, *UAG*, *UGA*.

6.11.1 Fattori di rilascio

6.11.1.1 Fattori di rilascio di classe I

I fattori di rilascio di classe I *RF* non sono tRNA e riconoscono i codoni di stop. Per i batteri sono *RF1* per *UAA* e *UAG*, *RF2* per *UAA* e *UGA*. Quello eucariotico *eRF1* riconosce tutti e tre i codoni di stop. I *RF* entrano ed occupano il sito *A* e promuovono il rilascio idrolitico del peptide finito portando una molecola d'acqua che attacca il legame estere liberandolo. Si nota come *RF* e tRNA sono molto simili in struttura. Gli *RF* batterici e quelli eucariotici non sono imparentati ma hanno lo stesso motivo *GCQ* necessario per la catalisi della reazione del rilascio del polipeptide.

Nell'idrolisi del legame estere del peptide avviene la deprotonazione del 2'-OH dove l'acqua agisce come donatrice di protoni ed elettroni.

6.11.1.2 Fattori di rilascio di classe II

I fattori di rilascio di classe II sono *GTPasi*. Nei batteri *RF3-GTP* rimuove *RF1* o *RF2* dal sito A dopo il rilascio del peptide *RF1/RF2*: l'energia ottenuta dall'idrolisi del *GTP* è utilizzata per la loro rimozione. L'ultimo tRNA nel sito P prende lo stato ibrido *E/P* vicino al sito A vuoto. Negli eucarioti il fattore di rilascio di classe II *eRF3-GTP* si lega al fattore di rilascio di classe I *eRF1*. Entrambi si legano al sito A e questo si accoppia con l'idrolisi del *GTP* di *eRF3-GTP* con il rilascio del peptide. Successivamente la *ATPasi ABCE1* rimuove *eRF3-GDP* dal sito A legandosi a *eRF1*.

6.11.2 Riconoscimento del codone di stop

Il riconoscimento del codone di stop è molto accurato e il rilascio prematuro del peptide è molto raro. Nonostante questo gli errori di decodifica sono più probabili durante condizioni di stress. Un mismatch tra un codone::anti-codone viene mosso nel sito P dopo il movimento di un codone da parte del ribosoma che causa un cambio conformazionale che diminuisce la fedeltà nel sito A che diventa un sito A a bassa fedeltà e *RF3* viene reclutata a tale sito indipendentemente dal codone di stop. Il peptide difettoso viene riconosciuto e degradato da peptidasi e proteasi cellulari. Questo agisce come un addizionale controllo di qualità.

6.11.3 Riciclo del ribosoma

Dopo che la traduzione è terminata il ribosoma deve essere rilasciato in modo che sia riciclato. Nei batteri il fattore di riciclo del ribosoma *RRF* e la *GTPasi* di traslocazione *EF-G-GTP* localizza al sito A vuoto promuovendo il disassemblaggio del ribosoma con l'energia fornita dall'idrolisi del *GTP* di *EF-G*. tRNA e mRNA si dissociano dalla subunità minore stabilizzata dal legame di *IF3* al sito E che previene l'evento di ri-associazione tra le subunità del ribosoma. Negli eucarioti in assenza di un ortologo del batterico *RRF* *eRF1* rimane legato al peptide dopo il rilascio con *ABCE1* che divide il ribosoma idrolizzando *ATP*. I fattori di iniziazione principali *eIF1* *eIF1a* e *eIF3* successivamente si legano alla subunità ribosomiale minore per impedire che si riunisca con la maggiore.

6.12 Energia richiesta per la traduzione

Circa il 60% dell'energia di una cellula viene utilizzata per la sintesi delle proteine.

6.12.1 Iniziazione

6.12.1.1 Procarioti

- 1 *ATP* per caricare *aa* sul *tRNA^{aa}* da parte di *aaRS*.
- 1 *GTP* per l'iniziazione.

6.12.1.2 Eucarioti

- 3 *ATP*: 1 per *aa-tRNA^{aa}*, 1 per la RNA elicasi e 1 per la scansione di *AUG*.

- 2 *GTP*: 1 consumato da *eIF2* per il rilascio del fattore di iniziazione e 1 consumato da *eLF5B* per unire le subunità 40S e 60S.

6.12.2 Allungamento

- 1 *ATP* da parte di *aaRS* per creare l'amminoacido-tRNA legame acile ad alta energia. L'energia generata dalla rottura di questo legame viene utilizzata per catalizzare 1 legame peptide dalla peptidil trasferasi.
- 1 *GTP* da parte di *EF-Tu* per portare *aa-tRNA^{aa}* al sito *A*.
- 1 *GTP* per la traduzione di *EF-G*.

6.12.3 Terminazione

- 1 *GTP* da parte di *RF3/eRF3* per rilasciare *RF1/2/eRF1* dal sito *A*.

6.13 Velocità ed accuratezza della traduzione

La velocità della sintesi delle proteine è di $15 \frac{aa}{sec}$ nei procarioti e $4 \frac{aa}{sec}$ negli eucarioti. La maggior lentezza è dovuta al fatto che avviene in un ambiente diverso rispetto alla trascrizione e non sono coordinate come nei procarioti, coinvolge più proteine e alcuni passaggi sono più intricati: processing del mRNA e scansione per *AUG*. Si nota però come la traduzione negli eucarioti sia 3 volte più accurata rispetto ai procarioti: in *E. coli* il tasso di errore è 10^{-4} - 10^{-5} . Tipicamente gli errori di sostituzione di amminoacidi sono dovuti a: amminoacilazioni sbagliate del *tRNA^{aa}* da parte di *aaRS* o di selezione sbagliata di *aa-tRNA^{aa}* da parte di *EF-Tu-GTP* o di *eEF1A-GTP*. Altri errori possono essere selezione sbagliata del sito di inizio, frameshift. Si consideri ora:

- Sia r la frequenza di errore della traduzione.
- Sia N il numero di amminoacidi contenuti in una proteina.
- La probabilità che un amminoacido sia sbagliato $P_w = r$.
- La probabilità che un amminoacido sia corretto $P_c = 1 - r$.
- La probabilità che una proteina contenente N amminoacidi sia sintetizzata correttamente $P_{Pc} = (1 - r)^N$

$N = 300$	$r = 10^{-4}$	$r = 10^{-2}$
P_w	0.01%	1%
P_c	99.9%	99%
P_{Pc}	97%	5%

Si nota come la sintesi delle proteine sia più prona ad errori rispetto a trascrizione e replicazione in quanto un errore durante l'allungamento di una proteina potrebbe essere molto meno drammatico rispetto a una mutazione del DNA o nella produzione del RNA, che porterebbe alla produzione di molte più proteine errate da esso.

6.14 Salvataggio dei ribosomi e controllo qualità degli mRNA

I ribosomi possono stallare o arrestarsi e devono essere salvati: il mRNA associato viene degradato, la proteina sintetizzata viene targettata per la proteolisi e il ribosoma e i tRNA sono riciclati. I ribosomi possono leggere il mRNA e riconoscere errori in esso: il controllo qualità mediato dai ribosomi garantisce che mRNA difettivi o tradotti non completamente sono eliminati.

6.14.1 mRNA senza stop codone

6.14.1.1 La risposta tmRNA nei procarioti

In quanto trascrizione e traduzione non sono separati spazialmente nei batteri, la seconda può avvenire su mRNA troncati. Il ribosoma stalla quando raggiunge la terminazione di un mRNA senza uno stop codone. La situazione è risolta dal transfer mRNA o tmRNA, lungo circa 300nt simile sia a tRNA che a mRNA: si adatta strutturalmente come un tRNA: le terminazioni 3' e 5' si uniscono per formare un dominio simile a tRNA. Viene tipicamente caricato con un alanina: *Ala-tmRNA* (alcuni *Gly-tmRNA^{Gly}*). Si trova inoltre nella sequenza un *ORF* e un codone di stop. *Ala-tmRNA* viene caricato da *EF-Tu-GTP* nel sito *A* dove il legame viene stabilizzato da *smgB* (small protein B). *smgB* inserisce la sua coda nel canale a mRNA vuoto sotto il sito *A* confermando che il mRNA è troncato. *Ala-tmRNA* agisce come un *Ala-tRNA^{Ala}* e l'alanina è aggiunta al polipeptide in stallo. Si nota come il legame avviene grazie alla similitudine strutturale tra un tRNA e tmRNA. Dopo l'aggiunta dell'alanina il tmRNA crea una struttura simile a un codone disponibile nel sito E. Si trovano ora 27 codoni che codificano per 9 amminoacidi, in questo modo il ribosoma continua a leggere e la catena a crescere fino a quando il ribosoma incontra il codone di stop del tmRNA. Il tag di amminoacidi codificato dal tmRNA attaccato al peptide segnala per la sua degradazione. Il ribosoma si disassembla, il no-stop RNA è rilasciato e degradato dalla esonucleasi *RNAse R* da 3' a 5'.

6.14.1.1.1 Evoluzione del tmRNA Si nota come il tmRNA è simile a un introne che si trova negli eucarioti.

6.14.1.2 La risposta del non-stop decay negli eucarioti

Differisce dai batteri in quanto i ribosomi possono continuare a leggere il trascritto fino alla fine della coda poli-A. Il ribosoma deve pertanto rimuovere la *PABP* legato ad essa e allunga la catena peptidica crescente con lisine AAA. Il ribosoma stalla all'ultimo AAA. Il ribosoma recluta il complesso *Ski7/Ski* che recluta l'esosoma. *Ski7* è simile a *RF3* e rilascia la proteina, mentre l'esosoma degrada il mRNA mutante da 3' a 5'. La proteina taggata con poli-lisina anormale viene ubiquitinata e degradata dal proteasoma.

6.14.2 Il ribosoma stalla a un ostacolo

6.14.2.1 Risposta del no-go decay negli eucarioti

Il mRNA che circonda il ribosoma è tagliato da endonucleasi e degradato dall'esonucleasi 5'-3' *Xrn1* e da 3' a 5' dall'esosoma. *Dom34* è simile strutturalmente a *eRF1* e a *Hbs1* simile a *eRF3*. Queste si legano al sito *A* e causano la dissociazione del ribosoma dal mRNA ed entrambe le subunità sono riciclate. Il frammento di mRNA coperto dal ribosoma, lungo circa 18nt viene degradato e la proteina immatura viene rilasciata e ubiquitinata causando la sua degradazione dal proteasoma.

6.14.2.2 Risposta dei procarioti

I ribosomi batterici contengono attività di RNA elicasi nelle proteine della piccola subunità: *rpS1*, *rpS3* e *rpS4*. *rpS1* viene utilizzata per l'iniziazione della traduzione regione *SD* in un hairpin che antagonizza un sRNA legato a una terminazione 5'. *rpS3* e *rpS4* vengono utilizzati per l'allungamento, è presente nell'entrata al mRNA e svolge gli hairpins. La differenza nella risposta, essendo i ribosomi conservati.

6.14.3 Stop codone prematuro

6.14.3.1 Risposta nonsense-mediated decay negli eucarioti

Gli mRNA con un codone di terminazione prematuro a causa di mutazioni del DNA, splicing errato o trascrizione errata produrrebbero una proteina troncata. Negli eucarioti gli stop codoni si trovano nell'esone finale, pertanto codoni di stop trovati da altre parti sono marcati come prematuri. Il nonsense-mediated decay viene pertanto causato dal ribosoma e dal exon junction complex. Lo splicing lascia un complesso proteico, l'exon junction complex che localmente marca l'unione di due esoni. Codoni di stop che si trovano a monte di un *EJC* indicano il loro nonsenso. Veri codoni di stop si trovano dopo l'ultimo *EJC* e non sono seguiti da un altro. Mentre traduce il ribosoma elimina il *EJC* quando non trova un codone di stop di fronte ad esso. In combinazione di codone di stop e *EJC* il ribosoma stalla. Successivamente vengono reclutati *UPF* fattori di sorveglianza *eRF1-GTP* e *eRF3* al sito *A*. Il ribosoma si disassembla e il mRNA con il codone di stop prematuro viene degradato.

6.15 Ricodifica - codone di stop programmato read-through e frame-shifting

Nel recoding un codone viene interpretato diversamente in un mRNA specifico. Avviene molto raramente e può produrre diverse proteine da un singolo gene:

- Soppressione non senso: un codone di stop è mal letto e non avviene terminazione.
- Frame-shifting: il mRNA fa shift in modo che la sintesi della proteina proceda in un diverso reading frame.

6.15.1 Frame-shifting

I ribosomi si muovono lungo il mRNA e leggono i codoni di 3 nucleotidi in modo sequenziale. Dopo la lettura del codone di inizio se il ribosoma si muove di un numero diverso di nucleotidi avviene un frame-shift del reading frame. In questo modo sono letti diversi codoni dal ribosoma producendo così una proteina diversa. Il frame-shifting è programmato e può essere coinvolto nella regolazione genica. Tipicamente avviene un nucleotide in avanti o indietro dopo che il codone *AUG* di inizio è letto correttamente e il codone di stop può così venire ignorato.

6.15.1.1 Fattore di terminazione batterico *RF2*

RF2 possiede 2 open reading frame: un corto *ORF1* seguito da uno più lungo seguito in un frame +1. Questo frameshift +1 accade alla sequenza *CUUUGAC*. Una sequenza *SD* in *ORF1* stimola lo shift. In abbondanza di *RF2* il ribosoma riconosce il codone di stop *CUUUGAC*, mentre in

deficienza il legame di esso a *UGA* è più lento e avviene un frameshift +1 in quanto *Leu-tRNA^{Leu}* riconosce il near-cognate *UUU* nel frame +1 e pertanto il prossimo codone è *GAC*.

6.15.2 Soppressione non senso

La soppressione non senso è rara ma più comune quando associata a specifici elementi di mRNA. UN motivo a esanucleotide *CARYYA* dopo lo stop codone in certi mRNA virali causa un aumento di read-through. Il precursore *GalPol* nel virus della leucemia murina è creato da una soppressione non senso: nella sintesi di *GalPol* si forma un pseudo-knot prossimale a 3'. Questo incoraggia la mal-lettura di *UAG* da *Glu-tRNA^{Glu}* che deve competere con i fattori di terminazione.

6.15.2.1 Retrovirus

Il frame-shifting programmato è anche necessario in molti retrovirus e retrotrasposoni *LTR* per produrre *Pol*, una proteina con attività di trascrittasi inversa, *RNAasi* e integrasi. La maggior parte dei ribosomi terminano al gene *gag* ma alcuni subiscono uno shift -1 risultando in una proteina di fusione *Gag-Pol*. Questo shift avviene a causa dello scivolamento del mRNA nel ribosoma facilitato da uno pseudoknot a valle. La sequenza è scivolosa in quanto il tRNA nel sito *P* può legarsi nel frame -1.

6.15.2.2 Inclusione di amminoacidi non standard

I codoni di stop possono permettere l'incorporazione di amminoacidi non standard.

6.15.2.2.1 Selenocisteina La selenocisteina *Sec* è simile a serina e cistina ma possiede selenio invece di un atomo di ossigeno o zolfo. Questa viene incorporata in diversi enzimi in siti catalitici dove può agire come un forte agente riduttore. Il Selenocisteina-tRNA è generato caricando *tRNA^{Sec}* con serina e convertendo la serina in selenocisteina. *tRNA^{Sec}* ha un anticodone che fa match con il codone di stop *UGA*. In *E. coli* la sequenza di inserzione della selenocisteina *SECIS* si trova a 3' del *UGA* che viene ricodificato. Per determinare il *UGA* che viene riconosciuto la proteina simile a *EF-Tu SelB* porta il tRNA carico al codone e agisce come guida per interagire con lo stem loop *SECIS*.

6.15.2.2.2 Pirrolisina La pirrolisina è una lisina modificata prodotta da due lisine che si trova nei siti catalitici delle metiltrasferasi. Si trova in archea metagenomici e batteri. La pirrolisina viene incorporata a un codone di stop *UGA* in modo simile alla selenocisteina. Viene utilizzato un *tRNA^{Pyl}* specializzato ma non si conosce se si trovano elementi di mRNA simili alla sequenza *SEICS*. *EF-Tu-GTP* viene coinvolto nel caricamento di *Pyl-tRNA^{Pyl}*.

6.16 Antibiotici che hanno come obiettivo l'attività del ribosoma

Gli antibiotici sono piccole molecole chimiche che uccidono o distruggono la crescita di organismi. Gli antibiotici più efficaci per uso terapeutico sono quelli che hanno come obiettivo un processo batterico o fungineo senza alterare quello dell'organismo ospite. Piccole differenze nella traduzione tra batteri ed eucarioti permette agli antibiotici di targettare selettivamente il ribosoma batterico e le proteine di traduzione, pertanto gli antibiotici tendono a legarsi a regioni critiche del ribosoma o a fattori di traduzione.

6.16.1 Esempi

6.16.1.1 Clorafenicolo

Si lega al 23S rRNA e inibisce la formazione del legame peptidico.

6.16.1.2 Eritromicina

Si lega alla porzione 50S e impedisce il movimento di traslocazione del ribosoma lungo il mRNA.

6.16.1.3 Tetraciclina

Interferisce con l'attacco di tRNA al complesso di mRNA-ribosoma.

6.16.1.4 Streptomicina

Cambia la forma della porzione 30S causando una lettura scorretta del codice sul mRNA. Sono incorporati gli amminoacidi sbagliati che producono proteine non funzionali causando la morte del microorganismo. Interagisce con la proteina ribosomale S12 causando cambi conformazionali nel centro di decodifica. Avvengono pertanto delle sostituzioni che rendono la proteina non funzionale.

6.16.1.5 Puromicina

Antibiotico che agisce tra procarioti ed eucarioti. Utilizzato unicamente nei laboratori per selezionare le cellule umane che esprimono il gene per la resistenza alla puromicina. È simile strutturalmente alla parte esterna del braccio accettore del tirosil-tRNA. Si lega al sito A della subunità maggiore e la peptidil trasferasi lo aggiunge alla catena poli-peptidica in crescita causando una terminazione prematura della catena.

6.16.1.6 Acido fusidico

Anti-batterico, blocca la traslocazione del ribosoma in quanto si lega a *EF-G-GDP* impedendo il suo rilascio in modo che il ribosoma non possa più caricare il nuovo *aa-tRNA^{aa}* bloccando l'allungamento della proteina.

6.16.2 Resistenza batterica agli antibiotici

La resistenza dei batteri agli antibiotici può essere dovuta a un influsso impedito, mutazioni dell'obiettivo, sue modifiche, sovra-produzione di un simile dell'obiettivo, protezione associata a fattori, modifica dell'antibiotico o sua degradazione.

6.17 Regolazione globale dell'iniziazione della traduzione in batteri ed eucarioti

La traduzione è regolata globalmente in risposta a cambi di sviluppo ed ambientali. Avviene tipicamente all'iniziazione e si può discriminare tra regolazione a livello globale o a livello locale. In condizioni sfavorevoli come pochi amminoacidi si deve ridurre la sintesi delle proteine per salvare risorse ed energia per sopravvivere portando a una riduzione globale di attività metaboliche e sintetiche. I tRNA sono etichettati con l'amminoacido cognate e i tRNA non carichi nella cellula sono

segno di bassa presenza di amminoacidi. Pertanto in caso di legame nel sito *A* di un tRNA scarico il batterio attiva la risposta stringente.

6.17.1 Stringent response

Quando arriva nel sito *A* un tRNA libero non avviene allungamento e il ribosoma si blocca. La *RelA* (*p*)*ppGpp* sintetasi o fattore stringent e *rpL11* sentono il tRNA scarico nel sito *A*. *RelA* produce (*p*)*ppGpp*, un pentaosfato guanina nucleotide da *GTP* utilizzando *ATP* e *AMP*. *ppGpp* o l'effettore di risposta stringent causando la risposta stringent. Questa molecola poi si lega a RNA Polimerasi abbassando l'affinità dei batteri per σ^{70} . Vengono inoltre attivati i geni in risposta allo stress favorendo l'interazione della RNA polimerasi con i fattori di stress σ^S , σ^E e σ^N e con il fattore di trascrizione *DksA*. In queste condizioni rimane invariata la produzione di rRNA e tRNA alter all'attività e produzione dei ribosomi. Diminuisce del 10% la sintesi delle proteine, mentre aumenta la sintesi di amminoacidi e di operoni legati allo stress.

6.17.1.1 Miglioramento delle condizioni

Quando le condizioni migliorano *SpoT* idrolizza *ppGpp* in *GDP*+*PPi* e la RNA polimerasi trascrive i target normali usando σ^{70} .

6.17.2 Eucarioti

6.17.2.1 Inibizione

Negli eucarioti a basse concentrazioni di amminoacidi i tRNA scarichi si legano alla chinasi *Gcn2* e il complesso fosforila il fattore di inizio della traduzione *eIF2-GDP* che porterebbe il tRNA iniziale al sito *P*, ma quando fosforilata si lega fortemente a *eIF2B* in modo che *eIF2-GDP* non possa essere ricaricato con *GTP* causando un arresto globale della traduzione. Le cellule successivamente operano una risposta allo stress in modo da indurre cambi trascrizionali e traduzionali per ripristinare i livelli di amminoacido.

6.17.2.2 Promozione

L'iniziazione eucariotica è stimolata dalla formazione di un mRNA circolarizzato, processo che coinvolge *eIF4E* al cap 5' e *PABP* alla coda poli-A che si associano attraverso interazioni con *eIF4G*. *eIF4E* viene regolato a livello trascrizionale e da fosforilazione e può essere sequestrato da un numero di *4E-BP* (proteine leganti *eIF4E*) in risposta a diverse condizioni di crescita diminuendo la circolarizzazione e attività di traduzione globali. La fosforilazione di uno di questi due attori inibisce la loro interazione permettendo la circolarizzazione del mRNA e il conseguente aumento della traduzione.

6.18 Regolazione dell'iniziazione attraverso sequenze agenti in *cis* nella 5'-UTR in batteri ed eucarioti

La regolazione della traduzione può anche avvenire al livello di un singolo trascritto oltre che a livello cellulare. La risposta più rapida è attraverso la regolazione dell'iniziazione della traduzione.

6.18.1 Cambi di conformazione degli mRNA

Gli mRNA assumono strutture diverse in condizioni diverse con effetti sui livelli di traduzione. Nei batteri la sequenza di Shine-Dalgarno viene spesso oscurata impedendo l'iniziazione.

6.18.1.1 Traduzione di *PrfA*

La traduzione del mRNA di *PrfA* che codifica un attivatore di trascrizione patogenico necessario per l'infezione di *Listeria monocytogene* è regolata da cambi di temperatura: a basse temperatura la sequenza *SD* è inaccessibile, mentre a 37° diventa accessibile.

6.18.2 Riboswitches

I riboswitches sono piccole molecole regolatorie che controllano la propria sintesi.

6.18.2.1 Auto-repressione traduzionale della sintesi delle proteine ribosomiali in *E. coli*

Il ribosoma di *E. coli* contiene 55 proteine e 3 rRNA. Le *RP* sono codificate in 19 operoni policistronici. Se il numero di *RP* è corretto per i propri rRNA viene prodotto il ribosoma, mentre se il numero di *RP* è maggiore dei corrispettivi rRNA la produzione del ribosoma è bloccata: 1 *RP* si lega alla 5'-UTR del mRNA policistronico per impedire il legame al ribosoma reprimendo la propria sintesi e quella delle altre *RP* da esso prodotte.

6.18.2.2 Regolazione del ferro

La regolazione 5'-UTR è meno comune negli eucarioti, ma viene utilizzata per regolare il ferro: ferro libero nella cellula è tossico, pertanto viene legato strettamente dalla ferritina. Maggiore la concentrazione di ferro, maggiore ferritina necessaria. La 5'-UTR del mRNA della ferritina contiene un elemento responsivo del ferro *IRE* che si lega alla proteina regolatrice del ferro *IRP*.

6.18.2.2.1 Ferro scarso *IRP* si lega a *IRE* impedendo l'accesso di *AUG* da parte del ribosoma.

6.18.2.2.2 Ferro abbondante Il ferro si lega a *IRP* che non possono legarsi a *IRE*, pertanto la traduzione continua a produrre ferritina.

6.18.2.3 Controllo globale della sintesi degli amminoacidi nel lievito regolando la traduzione del mRNA *GCN4* alla 5'-UTR

Gcn4 è un attivatore trascrizionale di più di 500 geni coinvolti nella biosintesi degli amminoacidi il cui livello è regolato traduzionalmente in base ai livelli di amminoacidi. Il suo mRNA nella 5'-UTR contiene 4 *ORF* a valle *uORF1-4* falsi prima della sequenza codificante. Gli *uORF* sono formati da un codone di start, un paio di codoni e un codone di stop.

6.18.2.3.1 Alti livelli di amminoacidi In questa situazione non si necessita di produrre *Gcn4*: dal cap 5' del mRNA il complesso di pre-iniziazione 48S scansiona il mRNA fino a trovare la sequenza di Kozak di *uORF1*: viene reclutata la subunità 60S, tradotti due amminoacidi, si incontra il codone di stop, si termina la traduzione e il ribosoma si disassembla. Nel 50% dei casi quando la subunità 60S lascia il complesso 48S di pre-inizio rimane legato al mRNA al codone di stop e a causa di

6.19. REGOLAZIONE DELLA TRADUZIONE ATTRAVERSO SEQUENZE AGENTI IN *CIS* NELLA 3'-UTR NEGLI EUCARIOTI

alti livelli di *eIF2-GTP* avviene un reinizio fino a *uORF2*. Il processo si ripete fino a *uORF4*, dove avviene la scarica completa del 40S e del 60S: il ribosoma non arriva al vero *AUG* e non avviene la traduzione del mRNA.

6.18.2.3.2 Bassi livelli di amminoacidi In questa situazione si necessita di produrre *Gcn4*: pochi tRNA sono etichettati con amminoacidi, pertanto non avviene un reinizio veloce e il complesso di pre-inizio 48S continua a scansionare con una bassa frequenza saltando degli *uORF* fino a trovare il *AUG* di *Gcn4*, pertanto con bassa frequenza viene reclutata la subunità 60S in modo da tradurre *Gcn4* e produrne bassi livelli.

6.18.2.4 Auto-repressione traduzionale della proteina legante poli-A citoplasmatica

PABP si lega alla coda poli-A di mRNA e permette insieme a *eIF4G* la sua circolarizzazione aiutando a stabilire il complesso di inizio 43S. *PABP* viene prodotta in eccesso e si lega nella propria coda poli-A impedendo il legame del complesso di pre-iniziazione nella 5'-UTR. Non avviene scansione né reclutamento della subunità maggiore.

6.18.3 Regolazione dell'attività di 5'-UTR da parte di sRNA nei procarioti

I piccoli RNA non codificanti aiutano i batteri ad adattarsi a condizioni di stress ossidativo variabili: l'ossidante H_2O_2 prodotto in un batterio in crescita aerobica in quanto può essere convertito in radicali superossidanti che possono danneggiare il DNA. Gli sRNA sono lunghi dai 50 ai 500nt, altamente strutturati con diversi stem-loop.

6.18.3.1 Repressione di *flhA* mRNA da *oxyS* sRNA

oxyS copre la sequenza *SD* e parte della sequenza codificante impedendo il legame del ribosoma.

6.18.3.2 Attivazione di *rpoS* mRNA da parte di *DsrA* sRNA

Il fattore sigma di stress in *E. coli* o $RpoS = \sigma^S = \sigma^{38}$.

6.18.3.2.1 Condizioni di crescita normali La sequenza *SD* e *AUG* sono seppellite nella struttura secondaria del 5'-UTR del mRNA di *rpoS* che non viene prodotto in condizioni normali con il proprio mRNA degradato.

6.18.3.2.2 Condizioni di stress *DsrA* a 87nt con un chaperone *Hfq* che si lega alla sua terminazione 5' si accoppiano con le basi della terminazione 5' del mRNA *rpoS* smascherando *SD* e *AUG*. La regione a valle è rimossa, *rpoS* viene tradotto e viene prodotto σ^S .

6.19 Regolazione della traduzione attraverso sequenze agenti in *cis* nella 3'-UTR negli eucarioti

Questo tipo di regolazione avviene negli mRNA degli eucarioti e la regola inibendo la circolarizzazione del mRNA o inibendo la traduzione dopo la sua circolarizzazione.

6.19.1 *Xenopus laevis*

Negli oociti di *Xenopus laevis* gli mRNA non sono tradotti inizialmente: si trovano in stato dormente. La loro traduzione dipende dalla lunghezza della coda di poli-A, inizialmente corta. L'elemento 3'-UTR di poli-adenilazione *CPE* citoplasmatico viene legato da *CPEB* che sequestra *eIF4E* attraverso Maskin bloccando la formazione del loop chiuso richiesto per l'iniziazione della traduzione. La chinasi *eg2* fosforila *CPEB* e *CPEB-P* recluta *CPSF* a una sequenza *AAUAAA* nel mRNA. Successivamente *CPSD* recluta la poli-A polimerasi *PAP* che estende la coda poli-A. Ora *PABP* può legarsi alla coda permettendo il legame di *eIF4G* e l'allontanamento di Maskin permettendo alla traduzione di procedere.

6.20 Trasporto e localizzazione degli mRNA

Negli eucarioti gli mRNA sono trasportati dal nucleo al citoplasma come mRNP. Gli mRNA possono muoversi a un sito specifico nel citoplasma per essere ancorati e tradotti solo là. Si nota come i gradienti di mRNA sono asimmetrici nella cellula e la localizzazione avviene grazie a proteine motrici legate agli RNA nei complessi mRNP. La traduzione degli mRNA viene inibita durante il trasporto. Pertanto gli mRNA contengono "zip codes" lunghi da 10 a 1000 basi nella regione 3'-UTR o metazoans o nella sequenza codificante.

6.20.1 Localizzazione del mRNA di *ASH1* nel lievito

Ash1 è un fattore di traduzione che agisce unicamente nella cellula figlia. Il suo mRNA viene espresso dalla cellula madre e trasportato nel citoplasma e poi nella gemma. Contiene uno "zip code" a stem-loop. Viene trasportato dal loop da proteine trasportatrici *She* lungo fili di actina. Il mRNA si accumula nella cellula figlia dove viene prodotta la proteina *Ash1*.

6.21 Granuli citoplasmatici di RNA e P-bodies

La conservazione degli mRNA non utilizzati nelle cellule mammiere avviene nei processing bodies o P-bodies o nei granuli di stress.

6.21.1 P-bodies

I P-bodies sono sempre presenti e contengono la maggior parte degli enzimi degradanti il RNA: enzimi di decapping 5', 3' deadenilasi e 5'-3' esonucleasi. Questi enzimi degradano gli RNA con codoni non senso, mRNA con *ARE* e agiscono nel silenziamento degli mRNA. Sono il vero sito di degradazione degli mRNA.

6.21.1.1 Componenti

- mRNA.
- 5'-3' esonucleasi *XRN1*.
- Fattori di deadenilazione *CCR4*, *CAF1*, *NOT1-4*.
- Fattori di rimozione del cap 5' *LSM1-7*, *DCP192*, *PAT1*, *DHH1*, *CPEB*.
- Fattori coinvolti nella degradazione non-senso *SMG5.7*, *UPF1*.
- Fattori di pathway del miRNA *miRNA*, *AGO1-4*, *GW182*.

6.21.2 Granuli di stress

I granuli di stress si formano unicamente durante condizioni di stress e sono zone di conservazione degli mRNA durante lo stress. Contengono mRNA, miRNA, subunità ribosomiali, fattori di iniziazione della traduzione, enzimi nucleolitici, elicasi proteine leganti il RNA e altro. Una cellula sotto stress riprogramma il proprio metabolismo di mRNA: che sono presenti come complessi di pre-iniziazione in stallo rilasciati dai granuli e tradotti quando necessario. Questi sono legati all'*ER*. Quando la condizione di stress termina il proteasoma 26S lo elimina.

6.21.2.1 Componenti

- mRNA con code poli-A.
- Fattori di iniziazione della traduzione *40S*, *eIF4E*, *eIF4G*, *eIF3*, *eIF2*.
- Fattori di controllo della traduzione *CPEB*, *PABP*, *DHH1*.
- Fattori di degradazione del mRNA *DHH1*, *Staufen*.
- Fattori di assemblaggio *snRNP*.
- Fattori di processamento del RNA.

Capitolo 7

Modifica e targeting delle proteine

7.1 Piegamento delle proteine assistito da chaperones

Le proteine devono essere piegate correttamente in modo da assumere la struttura tridimensionale corretta per funzionare correttamente. Tipicamente questo processo richiede degli aiutanti o chaperones. Il mal-piegamento delle proteine e la loro aggregazione può avere conseguenze patologiche come Alzheimer e Parkinson.

7.1.1 Processo di piegamento

La maggior parte dei peptidi emerge dal canale di uscita del ribosoma in forma estrusa, con struttura lineare. Il piegamento può cominciare appena la terminazione N esce dal ribosoma dove il peptide incontra dei chaperoni. Nei batteri i peptidi interagiscono con un chaperone associato al ribosoma detto “trigger factor” e successivamente possono piegarsi non assistite, legarsi a chaperones *Hsp70* come *DnaK* con co-chaperones *DnaJ* e *GrpE* o legarsi ai chaperon *Hsp60*. Un complesso di piegatura è formato da *GroES* e *GroEL*, che formano un cilindro. I chaperones lavorano interagendo con regione idrofobiche sul peptide che emerge impedendo che si associno scorrettamente con altre regioni idrofobiche. Il nome heat-shock proteins deriva dal fatto che le proteine formate da ribosomi ad alte temperature si mal-piegano causando la morte della cellula.

7.1.2 Ponti disolfuro

I legami disolfuro tra i residui di cisteina sono importanti per la stabilità e funzione delle proteine. Le cisteine infatti possono esistere in forma:

- Ridotta con il tiolo $-SH$, esiste più frequentemente in ambiente riducendo come il citosol. Subisce minimo stress ossidativo grazie all'attività della glutathionina reduttasi.
- Ossidata con il legame disolfuro $-S-S-$, esiste più frequentemente in ambiente ossidativo come ER e mitocondri.

Essendo i citosol di un batterio ridotto le proteine batteriche non possiedono legami disolfuro, mentre l'attività della sulfidril ossidasi permette alle proteine di essere discrete con legami disolfuro.

7.2 Targeting di cellule attraverso la cellula

Le proteine sono spesso trasportate a regioni cellulari distinte. Per direzionare la localizzazione sono presenti motivi specifici nella proteina che possono essere presenti alla terminazione N o C, con le prime spesso rimosse dopo che il loro lavoro si è svolto. I motivi di ordinamento possono essere presenti in diversi numeri per una localizzazione precisa. I motivi non sono rigidi ma si trova una certa flessibilità a livello di composizione.

7.2.1 Ordinamento delle proteine

L'ordinamento in compartimenti cellulari richiede il riconoscimento dei sorting motif, trasporto e traslocazione in o attraverso una membrana. I trasportatori, recettori o motori relocano le proteine e le passano a traslocatori. I trasportatori includono un peptide di riconoscimento del segnale *SRP* che riconosce e si lega al segnale sulla proteina che emerge dal ribosoma. Successivamente la proteina si muove attraverso un traslocatore nella membrana venendo o spinta o tirata utilizzando l'energia messa a disposizione dai chaperones.

7.2.2 Entrata ed uscita delle proteine dal nucleo

Il trasporto da e verso il nucleo avviene attraverso il complesso dei pori nucleari *NPC*, che sono traslocatori. Le proteine ribosomali sono create nel citoplasma e trasportate nel nucleo dove si assemblano con rRNA. La subunità maggiore e minore sono poi esportate nel citoplasma. *NPC* sono ricchi di amminoacidi carichi positivamente e sono riconosciuti dai trasportatori importine ed esportine che si legano a parti del *NPC*. La proteina legata può attraversare ora il poro.

7.2.2.1 Direzionalità

La *GTPasi Ran* governa la direzionalità del trasporto. La *RanGTP* si trova principalmente nel nucleo mentre la *RanGDP* si trova nel citoplasma. Il gradiente di *GDP/GTP Ran* promuove il legame e il rilascio delle proteine dentro e al di fuori del nucleo. Il cargo potrebbe includere *RNP*. Nel nucleo si trova *RCC1* o guanine-nucleotide exchange factor che cambia *RanGDP* in *RanGTP* causando la sua esportazione dal nucleo. Nel citoplasma invece si trova una *RanGAP* o *RanGTPasi* activating protein che stimola l'idrolisi di *RanGTP*.

7.3 Rottura post-traduzionale della catena polipeptidica

Molte proteine devono essere rotte da proteasi prima che possano diventare funzionali. Questo in quanto si può dover rimuovere la prima metionina da parte di metionina aminopeptidasi o rottura proteolitica di una sequenza leader del polipeptide. La rottura proteolitica può garantire una sua attività temporizzata. La chimotripsinogina è un precursore inattivo della chimotripsina, un enzima digestivo dell'intestino e sintetizzata nel pancreas. Il precursore è inattivo quando sintetizzato per impedire la digestione di cellule pancreatiche. Successivamente si muove nell'intestino dove viene tagliato e diventa attiva.

7.3.1 Insulina

L'insulina viene prodotta dal pancreas e subisce un processamento da parte di endopeptidasi. Il precursore preproinsulina creato dal ribosoma nel ER ruvido possiede una sequenza segnale che la

direziona al ER. Il segnale viene poi rotto nel ER formando ponti disolfuro (ER è un ambiente ossidativo) tra le catene *A* e *B*. Ora la proinsulina si muove nel Golgi. Ulteriori rotture della proinsulina nel GOLgi rimuove la catena *C* creando così l'insulina attiva. I livelli della catena *C* o antigene sono misurati per il diabete: quando i livelli di glucosio nel sangue sono alti attraverso *ELISA*. La catena *C* infatti possiede un'emivita molto più alta (20-30min) rispetto all'insulina (3-5min), in quanto la seconda viene degradata da enzimi di fegato e reni. L'insulina è un ormone che aiuta le cellule nell'uptake del glucosio. Esistono due tipi di diabete: di tipo 1 in cui il pancreas non produce abbastanza insulina o di tipo 2 in cui l'insulina non stimola l'uptake di glucosio nella cellula.

7.3.2 Rottura post traduzionale delle proteine e Alzheimer

La proteina *APP* amyloid precursor proteina, errore di rottura, placche amiloidi (aggiungere poi) i neuroni poi non sono più in grado di funzionare causando problemi al cervello.

7.3.3 Splicing di proteine: rimozione di inteine

Alcune proteine nei batteri, archea e eucarioti unicellulari possiedono inteine, lunghe tra i 138 e 844 sono parti di una proteina rimosse attraverso un processo di splicing delle proteine. Le inteine catalizzano la propria rimozione e legano le eseine che le affiancano. Il risultato sono due proteine: le eseine unite e l'intaina libera. L'intaina deve essere scissa in quanto altrimenti la proteina che la contiene rimane inattiva. L'intaina agisce tipicamente come un elemento genetico, muovendo la propria sequenza di DNA nel genoma.

7.3.3.1 Splicing di inteine

1. Un gruppo *HX* un un residuo conservato nella inteina: cisteina, serina o treonina attacca il gruppo carbossile del legame peptide nell'esteina N terminale.
2. Lo stesso carbossile viene attaccato da un residuo simile nella terminazione C dell'esteina.
3. Il gruppo ammino dell'ultimo residuo dell'intaina attacca il proprio gruppo carbossile causando il rilascio dell'intaina.
4. Le due eseine unite subiscono ulteriore attacco di un gruppo ammino sul carbossile per creare un legame peptidico.

Si nota come non viene coinvolto Mg^{2+} . Lo sono le istidine cariche positivamente nella proteina.

7.3.3.2 Gene di endonucleasi homing

Le inteine possono includere un homing endonuclease gene *HEG* dominio oltre ai domini di splicing. Questo dominio è responsabile per la diffusione dell'intaina rompendo il DNA sull'allele libero dell'intaina sul cromosoma omologo, attivando il sistema di riparazione del danno che copia il DNA codificante l'intaina in un sito prima privo di essa. Sono simili agli introni mobili di gruppo I. Il dominio *HEEG* non è necessario per lo splicing della proteina.

7.4 Regolazione e modifica post-traduzionale di proteine

L'attività e stabilità delle proteine è regolata da modifiche post-traduzionali, la via più veloce per regolare l'attività della cellula ed è regolata da enzimi. Può essere reversibile o irreversibile, fisiologica o patologica. Tutte le varie modifiche a livello di trascrizione, traduzione e modifiche post-traduzionale porta a un grado di complessità estremamente elevato. L'alto livello di complessità delle informazioni è dovuto a modifiche post-traduzionali delle proteine, che permettono per cambi veloci in risposte e adattamenti cellulari: il 5% del proteoma umano è composto da enzimi che catalizzano più di 200 diversi tipi di modifiche post-traduzionali di proteine. Che possono esistere di vari tipi in parallelo sulla proteina. Le modifiche sono enzimatiche, pertanto molto dinamiche e reversibili.

7.4.1 Ubiquitinazione delle proteine

L'ubiquitinazione è una grande modificazione post-traduzionale e consiste nell'aggiunta di una proteina in un'altra. L'ubiquitina è un peptide eucariote di 76 amminoacidi che marca la proteina per degradazione o ha un ruolo nel processo regolatorio come la risposta al danno del DNA. Viene covalentemente attaccata al gruppo amminoacido della lisina o anche non canonicamente a metionina, cisteina, serina o treonina. La molecola di ubiquitina aggiunge 8.5kDa. Oltre al proteosoma può portare proteine anche a proteasi nei lisozimi. Il legame tra Ubiquitina e lisina avviene tra gruppo carbossilico e amminico. Un processo enzimatico complesso che coinvolge tre enzimi *E1-3* con *ATP*. Il gruppo di ubiquitina è legato da lisina con gruppo carbossilico dell'ultimo amminoacido della coda: glicina 76.

7.4.1.1 Poli-ubiquitinazione

Nella proteina si trovano anche lisine distribuite su tutta la lunghezza e che possono essere substrati per l'ubiquitina: può pertanto avvenire poli-ubiquitinazione, importante in quanto fa parte di pathway di segnalazione. In base al punto di poli-ubiquitinazione si attiva un pathway diverso. Sulla lisina della proteina viene ubiquitinato attraverso il legame tra glicina 76 e lisina. Nell'ubiquitina si osserva una lisina 48, substrato per il prossimo gruppo ubiquitina. La lisina 48 si lega pertanto alla glicina di un'altra ubiquitina. Diverse ubiquitine inoltre si possono legare a più lisine diverse di una proteina oltre che a un'ubiquitina già aggiunta. Si formano pertanto catene di poli-ubiquitina lineari, ramificate, eterogenee od omogenee che danno vita a diverse conformazioni strutturali in base al linkage. Ubiquitinazioni multiple sono classificate dalla lisina nella prima ubiquitina a cui si attacca la successiva. Si possono trovare nella cellula catene di ubiquitina libere: questo avviene in quanto essa si stacca dalla proteina in modo da poter essere riciclata. Esistono inoltre anche altre modifiche simili all'ubiquitinazione che possono essere incluse nelle catene.

7.4.1.1.1 Mono-ubiquitinazione La mono-ubiquitinazione causa alla proteina del trafficking, endocitosi o espressione-silenziamento genico.

7.4.1.1.2 Ubiquitinazione *K11* L'ubiquitinazione *K11* porta a degradazione della proteina.

7.4.1.1.3 Ubiquitinazione *K48* L'ubiquitinazione *K48* porta a degradazione della proteina nel proteosoma.

7.4.1.1.4 Ubiquitinazione *K63* L'ubiquitinazione *K63* porta a tolleranza del danno al DNA, attivazione delle chinasi, trafficking e traduzione non-proteolitica.

7.4.1.1.5 Ubiquitinazione *K6* L'ubiquitinazione *K6* porta a una risposta infiammatoria.

7.4.1.2 Pathway di ubiquitinazione

Il processo di ubiquitinazione coinvolge tre gruppi di enzimi *E1-3* e dipende dalla presenza di *ATP*. La grande varietà di enzimi diversi garantisce grande specificità per l'ubiquitinazione.

7.4.1.2.1 *E1* - attivazione dell'ubiquitina L'ubiquitina viene ricevuta da un ubiquitin-activating enzyme creando un legame tra il gruppo carbossilico sulla glicina 76 e il gruppo sulfidrilico sulla cisteina di *E1*. Il processo richiede *ATP*, che funziona da sistema di controllo del processo. *E1* in presenza di *ATP* si lega all'ubiquitina e rilascia *AMP-PP_i* creando un legame tioestere.

7.4.1.2.2 *E2* - coniugazione dell'ubiquitina Avviene una trans-tioesterificazione: *E1* rilascia ubiquitina e si forma un nuovo legame tioestere con *E2*.

7.4.1.2.3 *E3* - trasferimento dell'ubiquitina L'enzima ubiquitina ligasi *E3* quando riceve l'ubiquitina da *E2* si trova già legato al substrato e catalizza il trasferimento di ubiquitina da *E2* alla proteina.

7.4.1.2.3.1 *E3 RING* ubiquitina ligasi Il trasferimento avviene da *E2* al substrato direttamente. Un dominio RING finger coordina Zn^{2+} attraverso cisteine e residui di istidina a intervalli regolari. Il RING unisce *E2* e il substrato insieme e media il trasferimento dell'ubiquitina.

7.4.1.2.3.2 *E3 HECT* ubiquitina ligasi Il trasferimento avviene da *E2* a un legame estere intermedio con *E3* sul gruppo sulfidrilico della cisteina prima che avvenga il trasferimento sul substrato. Il dominio HECT consiste di un lobo N terminale che interagisce con *E2* e un lobo C terminale che contiene la cisteina del sito attivo. Il movimento di un hinge loop flessibile porta vicini il lobo C e *E2*.

7.4.1.3 Degradazione della proteina nel proteosoma

La degradazione di proteine poli-ubiquitinate *K11* e *K40* avviene attraverso il proteosoma. Il proteosoma è un complesso proteico 26S che viene diviso in una parte centrale o particella nucleare 20S che contiene 30 proteine e due complessi regolatori 19S che contengono 19 proteine e sono formati da cap e lid.

7.4.1.3.1 Funzioni delle componenti del proteosoma

7.4.1.3.1.1 Proteine del lid Queste proteine si legano alle proteine del substrato e de-ubiquitinasi.

7.4.1.3.1.2 Proteine del cap Queste proteine sono *AAA-proteasi*, denaturano e spiegano la struttura terziaria della proteina substrato usando *ATP* e le spingono nella parte centrale.

7.4.1.3.1.3 Proteine del nuclep Queste proteine tagliano la proteina in frammenti lunghi tra i 3 e i 15 amminoacidi che possono essere poi degradati ulteriormente da proteasi. Il nucleo si divide in quattro subunità: due α e due β simmetriche rispetto al centro. Per ogni subunità β si trovano tre peptidasi.

7.4.1.3.2 Localizzazione del proteosoma Il proteosoma localizza sia nel citosol, dove si associa con i centromeri, la rete del citoscheletro e le membrane esterne del ER o nel nucleoplasma, dove si associa con i corpi *PML* ma non con i nucleoli.

7.4.1.3.3 Processo di degradazione La coda di poli-ubiquitina si lega al lid e l'anello di *ATPasi* la trasporta nel nucleo, dove le peptidasi la tagliano. A questo punto i peptidi escono attraverso l'anello di *ATPasi* sull'altro canale. La direzionalità del processo è data da?

7.4.2 SUMOilazione delle proteine

Altre piccole proteine simili all'ubiquitina (ubiquitin-like proteins) possono essere usate per modificare le proteine. In particolare *SUMO* si attacca alle catene laterali della lisina durante la sumoilazione. Il processo viene catalizzato da enzimi simili a quelli dell'ubiquitina ed è importante per l'espressione genica e il targeting delle proteine, oltre a controllare la degradazione competendo con l'ubiquitina per i siti di legame.

7.4.2.1 Caratteristiche

SUMO o small ubiquitin-like modifier contiene 100 amminoacidi. Inizia con una metionina e termina con una glicina e pesa $12kDa$. Negli eucarioti è conservata: se ne trovano tre negli umani e una nel lievito. Si coniuga a proteine nel substrato a lisine, con consenso *yKxE*, dove *y* è un amminoacido idrofobico, *x* un amminoacido e *E* acido glutammico. Si attacca attraverso l'ultimo amminoacido che viene esposto solo dopo il processamento. Usa lo stesso pathway di ubiquitina e viene tipicamente utilizzata per targettare proteine nel nucleo. Il processo di attacco di *SUMO* è reversibile attraverso *de-SUMOilasi*, che sono delle isopeptidasi. Può formare catene eterogenee con altre proteine *SUMO*, con ubiquitina e *NEDD8* (altra proteina simile a ubiquitina).

7.4.2.2 Substrati e funzioni

7.4.2.2.1 Proteine dei pori nucleari *SUMO* permette il traffico nucleare.

7.4.2.2.2 Fattori di trascrizione *SUMO* influenza l'espressione genica.

7.4.2.2.3 Proteine della riparazione e replicazione del DNA *SUMO* influenza la stabilità genomica.

7.4.2.2.4 Proteine dei cinetocori e dei centromeri *SUMO* influenza l'integrità cromosomica.

7.4.2.2.5 SUMOilazione bilanciata *SUMOilazione* bilanciata porta alla progressione ed arresto del ciclo cellulare, sopravvivenza cellulare o apoptosi, divisione e proliferazione cellulare, differenziamento ed invecchiamento.

7.4.2.2.6 SUMOilazione deregolata *SUMOilazione* deregolata porta a tumori, malattie neurodegenerative, diabete, infezioni virali e difetti nello sviluppo.

7.4.3 Regolazione dell'attività di *PCNA* attraverso ubiquitinazione e *SUMOilazione*

Questo esempio viene preso dal lievito e *PCNA* è l'antigene per le cellule nucleari proliferanti. Equivale alla pinza scorrevole nei batteri: recluta la DNA polimerasi al DNA ed è coinvolta nella replicazione e riparazione del danno al DNA.

7.4.3.1 Replicazione del DNA

Mono-SU di *PCNA*^{K₁₂₇,K₁₆₄} con *E2 Ubc9* causa una promozione della replicazione da parte di *PCNA*.

7.4.3.2 Danno al DNA

Mono-Ub di *PCNA*^{K₁₆₄} attraverso *E2 Rad6* ed *E3 Rad18*. *Rad5* si lega a *PCNA*^{K₁₆₄} e recluta *UBC13*, un *E2* e *MMS2*, un *E3* causando una catena *Poly-Ub Ub*^{K₆₃} in modo che *PCNA* promuova la riparazione del danno al DNA.

7.4.4 Sindrome di Alzheimer

La sindrome di Alzheimer viene causata una rottura errata di *APP* (amiloid precursor protein), che contiene una regione che può essere tagliata da tre enzimi diversi a posizioni diverse: $\alpha\beta$. In caso di rottura corretta interviene la α -secretasi che forma *sAPP*, *APP* α , una proteina solubile che riduce la concentrazione di Ca^{2+} aumentando la neuroprotezione e la neuroplasticità. In caso di rottura sbagliata da parte di β -secretasi e γ -secretasi si forma la proteina amiloide β che forma degli aggregati essendo insolubile portando a un aumento di Ca^{2+} , neurotossicità e una dimensione anormale dei neuriti, portando infine alla sindrome di Alzheimer. La *SUMOilazione* della β -secretasi le impedisce di creare rotture non volute.

7.5 Fosforilazione

La fosforilazione è la modifica post-traduzionale più frequente. È reversibile e più del 30% delle proteine in una cellula sono fosforilate da circ 500 chinasi, enzimi che fosforilano le proteine. L'insieme completo delle chinasi codificato in una cellula forma il chinoma. Esistono inoltre anche circa 150 proteine fosfatasi, enzimi che defosforilano le proteine fosforilate.

7.5.1 Chinasi

Le proteine chinasi (esistono anche le chinasi del DNA) sono classificate in famiglie in base a struttura e funzione formando un albero genealogico. Trasferiscono il γ -fosfato dal *ATP* su una serina, treonina, tirosina, istidina o aspartato. *ATP* e proteina formano una fosfo-proteina e *ADP*.

7.5.1.1 Classificazione

Le chinasi vengono classificate in base al residuo che fosforilano:

- Chinasi a specificità singola: fosforilano un amminoacido specifico: serina *S*, treonina *T* o tirosina *Y*.
- Chinasi a specificità doppia: fosforilano due specifici amminoacidi: *S* e *T* o *S* e *Y*.

- Chinasi a specificità duale: agiscono come tirosina chinasi e serine-treonina chinasi e possono fosforilare S , T e Y .

Le più abbondanti sono le thinas S e T , riflettendo la frequenza con cui i tre residui sono fosforilati:

$$pS : pT : pY \rightarrow 1800 : 200 : 1$$

Anche l'istidina si può fosforilare, ma questo evento avviene raramente. Nei batteri è più frequente la fosforilazione su istidina e aspartato.

7.5.1.2 Struttura

Le chinasi contengono domini di legame per ATP altamente conservati. Alcune chinasi sono anche $ATPasi$ con diverse attività nel dominio chinasi. Il sito di riconoscimento del substrato invece varia molto tra le varie chinasi.

7.5.2 Effetti del gruppo fosfato

A pH fisiologico il gruppo fosfato è di-anionico: possiede infatti due atomi di ossigeno carichi negativamente. Questo ossigeno può formare legami a idrogeno e mediare le interazioni elettrostatiche che possono stabilizzare una proteina internamente per esempio con amminoacidi carichi positivamente come arginine o lisine. 1 O^- in 2 gruppi fosfato può legarsi a Mg^{2+} , il cofattore enzimatico per gli attacchi nucleofili. La fosforilazione pertanto può:

- Cambiare il riconoscimento proteina-proteina. Essendo le interazioni tra proteine le basi per la segnalazione inter/intracellulare esistono reti di segnalazione guidati da chinasi.
- Cambiare l'attività di una proteina: la fosforilazione aggiunge massa e carica negativa.

La fosforilazione regola ogni aspetto della vita della cellula:

- Piegatura delle proteine.
- Localizzazione delle proteine.
- Attività delle proteine.
- Degradazione delle proteine.
- Interazioni tra proteine.

Inoltre la fosforilazione può causare eventi di modifiche seguenti di ogni tipo di proteina nel priming. Essendo un processo molto veloce anche le sue conseguenze lo sono. Infine le proteine con N siti di fosforilazione possono esistere in 2^N possibili stati di fosforilazione, espandendo significativamente l'attività delle proteine e formando un fosfo-proteoma.

7.5.2.1 Risposta alle infezioni - pathway $JAK-STAT$

Il pathway del fattore di trascrizione JAK -chinasi $STAT$ illustra l'importanza della fosforilazione. Quando il ligando α -interferon viene secreto in risposta a un'infezione virale questo si lega ai recettori interferon sulla superficie cellulare. La JAK chinasi è associata con il dominio citoplasmatico del recettore. Il legame di α -interferon unisce due recettori e i domini di JAK chinasi che si fosforilano tra di loro oltre alla coda del recettore. $STAT$ viene reclutato alla coda del recettore, viene fosforilata e rilasciata. Fosfo- $STAT$ entra nel nucleo e attiva i fattori di trascrizione in modo da rispondere all'infezione.

7.5.3 Tirosina chinasi

Esistono 91 tirosina chinasi nelle cellule umane. 59 sono chinasi recettori transmembrana, mentre 32 sono citoplasmatiche. Sono classificate in un numero di famiglie in base alla struttura e ricevono un segnale esterno (ligando) che viene trasdotto nel citoplasma e nel nucleo. Il ligando causa un'oligomerizzazione del recettore chinasi che porta a un'auto-attivazione attraverso fosforilazione. Il recettore fosforilato a sua volta fosforila proteine target causando una cascata di segnali. Queste chinasi sono mutate o sovra-prodotte in molti tumori, specialmente il caso in cui diventano attive costituzionalmente rendendo non necessario un segnale di input per la loro attivazione.

7.5.4 Chinasi dipendenti dalla ciclina *CDK*

Le chinasi dipendenti dalla ciclina sono regolatori chiave del ciclo cellulare, della trascrizione (fosforilano infatti *RNA Pol II*), del metabolismo del DNA e della proliferazione. La loro attività dipende da una subunità regolatoria detta ciclina che aumenta l'attività di 40 000 volte legando *ATP* e Mg^{2+} . Esistono diversi tipi di chinasi e diverse cicline, differenziate attraverso l'evoluzione:

- Nel lievito gemmante si trovano 6 *CDK*, 2 per il ciclo cellulare e 4 per la trascrizione.
- Negli umani se ne trovano 20, 11 per il ciclo cellulare e 9 per la trascrizione.

7.5.4.1 *CDK* del lievito

- *Cdc28* si lega a 9 cicline per regolare la progressione del ciclo cellulare.
- *Pho85* si lega a 10 cicline per fare sensing dei nutrienti, integrare la trascrizione della cellula e la progressione del ciclo cellulare.

Si nota come i livelli di *CDK* rimangono costanti nel lievito: la loro attività viene modulata da cambi nella presenza delle cicline.

7.5.4.2 Controllo del ciclo cellulare del lievito sotto condizioni di privazione di fosfato e azoto

La ciclina *Cln3* di G_1 è una proteina instabile che si accumula per iniziare il ciclo cellulare formando un complesso con *Cdc28*. Il complesso inibisce il repressore di trascrizione G_1 *Whi5* fosforilandolo.

7.5.4.2.1 Nutrienti sufficienti Con nutrienti sufficienti la chinasi *Pho85* e la ciclina *Pho80* si attivano e due siti di consenso per *Pho85* in *Cln3* sono fosforilati per stabilizzare *Cln3* causando l'inizio del ciclo cellulare.

7.5.4.2.2 Mancanza di fosfato *Pho85-Pho80* sono inibite da *Pho81* impedendo a *Cln3* di fosforilarsi. *Cln3* viene distrutta causando un arresto del ciclo cellulare in G_1 .

7.5.4.2.3 Mancanza di azoto *Pho85* e la ciclina *Pcl2* o *Clg1* attivano attraverso fosforilazione *Ssa1* per sequestrare *Cdk1-Cln3* portando alla loro degradazione. Viene causato in questo modo un arresto del ciclo cellulare in G_1 .

7.5.4.3 Antagonizzare l'attività delle *CDK*

Esistono diversi modi per antagonizzare l'attività delle *CDK*:

- Degradazione delle cicline.
- Inibitori di *CDK*, proteine *CKI*: si legano a *CDK* o al complesso *CDK-ciclina*. Se si considerano *CDK* come i motori del ciclo cellulare *CKI* sono dei freni al motore. Disattivando *CKI* si ha nella cellula proliferazione incontrollata e tumorigenesi.
- Degradazione o inibizione dell'enzima attivante di *CDK*.
- Rimozione o degradazione del ligando che causa la trascrizione del gene codificante *CDK* come idrolisi di *cAMP* da *cAMP* fosfodiesterasi.
- Fosfatasi che defosforilano la *CDK* invertendo la loro attività indirettamente defosforilando i substrati fosforilati da *CDK*. Queste proteine sono meno specifiche rispetto alle chinasi e loro mutazioni non sono fortemente associate con malattie o cancro.

7.6 Acetilazione

7.6.1 Meccanismo di acetilazione

Acetil-coA viene usato come donatore del gruppo acetile per etichettare covalentemente la terminazione *N* di una proteina o lisine interne. Più del 80% delle proteine eucariotiche è acetilato e la N-acetiltrasferasi *NAT* agisce durante la sintesi del ribosoma, quando la proteina in uscita è lunga tra i 20 e i 50 amminoacidi.

7.6.1.1 Composizione di *Acetil-coA*

Acetil-coA è formato da:

- Gruppo acetile.
- Acido pantotenico (vitamina *B5*).
- β -mercapto-etilamina.
- *ADP* fosforilato.

7.6.1.2 Acetilazione delle proteine alla terminazione *N* *AcNt*

La terminazione *Nt* presenta un gruppo ammino NH_3 libero a cui *NAT* attacca il gruppo acetile, liberando *CoA* (da *Ac-CoA*) e formando un *AcNt*. Il processo inverso viene svolto da *NDAC* (N-terminal deacetilasi).

7.6.1.3 Acetilazione di catene laterali di lisine interne alla proteina *AcK*

La catena laterale della lisina *K* presenta un gruppo ammino libero a cui *KAT* (lisina-acetiltrasferasi) attacca il gruppo acetile formando *N ϵ -acetil-lisina* *AcK* e liberando *CoA* (da *Ac-CoA*). Il processo inverso viene svolto da *KDAC* (lisina-deacetilasi).

7.6.2 Acetilazione da parte di istone acetil trasferasi

L'istone acetil trasferasi acetila anche altre proteine oltre agli istoni. L'acetilazione ha effetto sulla carica neutralizzando i gruppi ammino positivi. Apre la struttura cromatina per la trascrizione, con effetto sull'ereditarietà epigenetica. I fattori di trascrizione sono complessi che spesso contengono enzimi (de)acetilanti.

7.6.2.1 Esempio - istone acetil trasferasi *Gcn5*

L'istone acetil trasferasi *HAT Gcn5* viene identificata nel lievito ma è presente in tutti gli eucarioti. Attiva la trascrizione e interagisce con le proteine substrato attraverso un bromodominio.

7.6.3 Deacetilazione da parte di istone deacetilasi

La deacetilazione di proteine acetilate avviene grazie a istone deacetilasi *HDAC*. La coda istonica ipo-acetilata causa una condensazione del cromosoma e pertanto un silenziamento trascrizionale. Si notano due famiglie di *HDAC*:

- *HDAC* classiche con Zn^{2+} come cofattori.
- Sirtuine con NAD^+ come cofattori. Sono importanti nell'ipo-acetilazione di *H3* e *H4*, nella stabilità genomica (stabilità del vettore di *rDNA*) e nell'invecchiamento (influiscono sulla stabilità dei telomeri).

7.6.4 Aberrazioni dei processi

Acetilazione e deacetilazione non regolate contribuiscono ai cancro.

7.7 Metilazione

La metilazione è catalizzata da protein methyltransferases *PRMT*, che utilizzano il *S-adenosil metionina* come fonte del gruppo metile. Avviene sulle catene laterali di azoto di arginina e lisina. A differenza dell'acetilazione possono essere aggiunti multipli gruppi ammino (fino a 3). Non cambia la carica cationica ma aumenta l'idrofobia e l'ingombro sterico dell'amminoacido influenzando le interazioni proteina-proteina. Il processo inverso della metilazione della lisina avviene grazie a ammino-ossidasi o demetilasi specifiche alla lisina.

7.7.1 Protein methyltransferases *PRMT*

Il substrato preferito delle *PRMT* sono le code *N* terminali degli istoni *H3* e *H4*. La coda che trasporta un gran numero delle altre modifiche è quella di *H3*, che può contenere fino a 110 592 modifiche.

7.7.2 Demetilasi

Le demetilasi spesso coesistono in un complesso del fattore di trascrizione con *PRMT* e definiscono lo stato di trascrizione: istoni metilati impongono repressione.

7.7.3 Metilazione aberrante

Metilazioni e demetilazioni non regolate contribuiscono al cancro.

7.8 I domini lettori

Fosforilazioni, acetilazioni e metilazioni sono modifiche proteiche comuni reversibili che hanno effetto sulla conformazione e funzione della proteina. Pertanto molte proteine possiedono domini speciali che riconoscono catene laterali modificate:

- Domini *SH2*: riconoscono le tirosine fosforilate.
- Bromodomini: riconoscono le lisine acetilate.
- Cromosomini: riconoscono le lisine metilate.

7.9 Glicosilazione

Catene di carboidrati o *glicani* sono aggiunte su una proteina nella glicosilazione, che produce una glicoproteina. Circa metà delle proteine sono glicosilate da una grande varietà di oligosaccaridi, formati da vari monosaccaridi in varie combinazioni, spesso ramificati. La struttura e tipo di oligosaccaride aggiunto alla proteina decide cosa succede alla proteina. Gli enzimi che aggiungono gli oligosaccaridi sono *glicosiltrasferasi*, mentre quelli che li rimuovono sono *glicosidasi*. La glicosilazione avviene nel ER e nel Golgi.

7.9.1 Glicosilazione in diversi organismi

La glicosilazione avviene in archea, procarioti ed eucarioti.

7.9.1.1 Eucarioti

Negli eucarioti determina:

- Localizzazione cellulare per membrane o secrezione extracellulare.
- Aiuta nel piegamento delle proteine.
- Aiuta le proteine ad essere stabili e ad assumere la struttura funzionale.
- Partecipano nella risposta immunitaria.

7.9.1.2 Procarioti

Nei procarioti:

- Aiuta nella patogenesi.
- Invasione delle cellule ospite di proteine di superficie batterica.

7.9.2 Ruoli della glicosilazione

La glicosilazione:

- Aumenta la solubilità della proteina.
- Impedisce l'aggregazione delle proteine: i carboidrati tendono a non interagire.
- Piccoli monosaccaridi sono coinvolti nella segnalazione.

- Aiutano con il riconoscimento di proteine sulla superficie cellulare.
- Aggiungono volume alla proteina in quanto tendono a distendersi invece che a piegarsi.

7.9.3 Glicani

Essendo i carboidrati idrofilici si trovano tipicamente al di fuori della molecola. I glicani possono formare da semplici monosaccaridi fino a centinaia di saccaridi e possono formare fino al 90% della massa di una glicoproteina.

7.9.3.1 Monosaccaridi

La maggior parte dei glicani sono fatti da monomeri di esosi:

- D-glucosio *Glc*.
- D-galattosio *Gal*.
- D-mannosio *Man*.
- N-acetil-D-glucosammina *GlcNAc*.
- N-acetil-D-galattosammina *GalNAc*.

Ognuno di essi viene aggiunto o rimosso da un enzima specifico: per esempio l'attacco di una N-acetilglucosammina a serina o treonina *O-GlcNAc* è reversibile:

- Viene aggiunto da O-acetilglucosammina trasferasi.
- Viene rimosso da O-acetilglucosamminidasi.

7.9.4 Tipi di glicosilazione

I carboidrati sono aggiunti attraverso legami glicosidici *N-linked* od *O-linked*. I primi sono più comuni e i tipicamente i più complessi. Le glicosilazioni si dividono in base a su quale amminoacido viene attaccato l'oligosaccaride:

- *N*-glicosilazione: il glicano si lega a una asparagina nel reticolo endoplasmatico: il legame avviene su un gruppo amminico.
- *O*-glicosilazione: il glicano si lega a una serina o una treonina nel reticolo endoplasmatico, nel Golgi, nel citoplasma o nel nucleo: il legame avviene sul gruppo idrossilico.
- Glipiazione: il glicano collega un fosfolipide alla parte *C*-terminale di una proteina permettendone l'ancoraggio alla membrana.
- *C*-glicosilazione: il mannosio si lega all'anello indolico di un triptofano.
- Fosfoglicosilazione: un glicano si lega a una serina mediante un legame fosfodiesterico.

7.9.4.1 Glicosilazione *N-linked*

La catena di carboidrati ramificata forma un precursore a 14 monosaccaridi viene legata covalentemente a NH_2 di un'asparagina. L'asparagina è parte di una sequenza consenso *Asn-X-Ser/Thr*, specialmente per le asparagine N-terminali e dove *X* è un qualsiasi amminoacido tranne la prolina. È l'unica reazione di glicosilazione che avviene co-traduzionalmente mentre il peptide in crescita è attaccato al ribosoma e sta venendo importato nel lume del ER.

7.9.4.1.1 Sintesi del precursore La sintesi del precursore avviene in due fasi nella membrana del ER e richiede più di 23 passaggi enzimatici:

1. Al gruppo fosfato di un dolicoles mono-fosforilato (poli-isoprenoide) presente nella membrana del ER sono aggiunti 2 *N*-acetilglucosammine *GlcNAc* e 5 mannosidi nella parte citosolica del ER.
2. Questo dolicoles precursore fa un flip: la catena del carboidrato si trova nel lume del ER. Sono aggiunti 4 mannosidi e 3 *GlcNAc* alla catena e il precursore si lega a un *Asn* della proteina da un oligosaccaril-trasferasi *OST*.

Il precursore legato alla proteina viene poi fatto maturare da glicosidasi e mannosidasi nel ER. Successivamente la proteina glicosilata si muove nel Golgi dove possono essere aggiunti o modificati altri zuccheri.

7.9.4.1.2 Maturazione dei precursori delle proteine *N*-glicosilate in Golgi La maturazione del precursore in ER e Golgi determina se la proteina deve rimanere ancorata alla membrana o essere secreta. All'oligosaccaride ad alto contenuto di mannosio vengono prima eliminati 3 mannosidi, dalla mannosidasi *I*, viene aggiunto un *GlcNAc* legata a *UDP*. Successivamente vengono rimossi altri tre mannosidi e aggiunte tre molecole (che non si legge dalle slides ?)

7.9.4.2 Glicosilazione *O*-linked

La glicosilazione *O*-linked avviene principalmente nelle proteine su serine o treonine post-traduzionalmente. Gli enzimi si trovano in diversi compartimenti del Golgi e il dolicoles-P non è coinvolto. Le proteine possono essere *O*-glicosilate con *GlcNAc*, fucosio, xilosio, galattosio, mannosio, acido sialico in base alla proteina e all'organismo. Questo processo è fondamentale per la produzione di proteine proteoglicani utilizzati per creare componenti della matrice extracellulare. Gli anticorpi sono spesso pesantemente *O*-glicosilati.

7.9.4.2.1 Processo di *O*-glicosilazione Il processo avviene in due fasi:

- Trasferimento di 1 *UDP-GalNAc* al β -OH di *S/T/K/P* attraverso una *N*-acetilgalattosammina trasferasi.
- I monosaccaridi come donatori di zucchero legati ai nucleotidi *UDP/GDP/CMP* sono aggiunti uno alla volta su un gruppo *OH* presente sugli zuccheri nella fase di estensione.

7.9.4.3 Gliptazione

La gliptazione è l'attacco covalente di un'ancora glicosil-fosfatidil-inositolo *GPI* alla terminazione *C* di una proteina. *GPI* contiene un carboidrato e lipidi. I carboidrati sono attaccati alla membrana da ude catene di acidi grassi. L'ancora viene assemblata nel ER e attaccata alla terminazione *C* della proteina allo stesso tempo in cui i segmenti transmembrana sono rotti. La proteina rimane così attaccata alla membrana solo dall'ancora *GPI*. Il carboidrato direziona la proteina alla superficie cellulare, dove può essere rilasciata da rottura dei lipidi. Questo processo è pertanto utile per i processi di segnalazione. La modifica avviene negli eucarioti e in certi archea e localizza proteine solubili all'esterno della superficie della membrana cellulare. La *GPI* viene utilizzata per l'adesione tra cellule e la trasduzione del segnale. Viene rimossa dalla fosfolipasi *C* che regola la dinamica di localizzazione della proteina.

7.9.4.3.1 Struttura di *GPI* L'ancora a *GPI* è composta da:

- Un linker fosfo-etanolamina che si lega alla terminazione *C* delle proteine obiettivo.
- Il nucleo del glicano.
- Una catena fosfolipidica che ancora la proteina alla membrana.

Il nucleo e la coda fosfolipidica sono variabili per dare specificità alla localizzazione e alla natura del legame.

7.9.4.3.2 Sintesi di *GPI* La sintesi di *GPI* coinvolge due processi: la creazione dell'ancora a *GPI* e l'attacco della proteina.

7.9.4.3.2.1 Sintesi dell'ancora La sintesi comprende diversi passaggi:

1. La sintesi inizia nella parte citoplasmatica del ER, poi avviene un flip interno verso il lume del ER.
2. Durante la sintesi vengono aggiunti residui di mannosio e altri zuccheri sulla molecola di fosfatidilinositolo inseriti nel bistrato della membrana del ER.
3. La faccia citoplasmatica contiene i substrati donatori, le derivate nucleotidiche degli zuccheri.
4. Nel lume il dolicholo-P-Mannosio e 2,3-fosfoetanolamina vengono aggiunti per l'attacco alla proteina.

7.9.4.3.2.2 Attacco della proteina L'ancora ora formata entra in contatto grazie al segnale di attaccamento all'ancora con l'amminoacido sito di legame della proteina sintetizzata e si attacca ad essa.

7.9.5 Ulteriori modifiche

Le proteine glicosilate possono essere ulteriormente modificate chimicamente alle code a polisaccaride attraverso:

- Sulfurilazione di mannosio e N-acetilglucosammina.
- Fosforilazione del mannosio.
- Acetilazione dell'acido sialico.

7.9.6 Glicosilazione specifica-specifica di proteine ricombinanti

7.10 Modifiche lipidiche

I lipidi idrofobici attaccati alle proteine le portano tipicamente nelle membrane oltre a regolare la loro funzione e stabilità. Si attaccano covalentemente ad amminoacidi specifici, tipicamente vicino una delle terminazioni. Le proteine possono avere più di una modifica lipidica e alcune sono reversibili. Avvengono nel citoplasma o nella parte citoplasmatica della membrana cellulare o nel lume dei pathway secretori (ER e Golgi).

7.10.1 Tipologie

Un tipo di modifica lipidica è l'aggiunta di un'ancora glicosilfosfatidilinositolo *GPI*.

7.10.1.1 Acilazione

L'acilazione o palmitoilazione o miristoilazione viene fatta da gruppi di acidi grassi.

7.10.1.1.1 Alla terminazione N L'acilazione alla terminazione *N* coinvolge l'attacco di miristoile, un acido grasso 14-*C* saturato attraverso un legame *N* ammine con una glicina. È generalmente stabile e permette per associazione di membrana di lunga durata e non è reversibile.

7.10.1.1.2 Alla terminazione C L'acilazione alla terminazione *C* coinvolge l'attacco di palmitoile, un acido grasso 16-*C* saturato al gruppo sulfidrilico *S* di una cisteina. Può essere rotta facilmente ed è pertanto reversibile. Viene utilizzata per modulare la localizzazione delle proteine e loro interazioni. Ha anche effetto sulla stabilità. Può avvenire in contemporanea all'acilazione sulla terminazione *N*.

7.10.1.2 Prelinazione

La prelinazione o farnesilazione e geranilgeranilazione viene fatta da gruppi di isoprenoidi. La prelinazione è l'aggiunta di un gruppo isoprenoide, tipicamente un 15-*C* farnesile o un 20-*C* geranilgeranile. Sono aggiunti attraverso un legame tioestere da una cisteina alla terminazione *C* di una proteina. Il gruppo carbossilico della cisteina viene poi metilato aumentando la sua affinità per la membrana. La maggior parte delle proteine prenilate si legano a *GTP*. Il motivo di legame è *CAAX*, dove *X* è glutammina, metionina o serina per il farnesile o leucina per il geranilgeranile.

7.10.2 Modifiche multiple

Le proteine possono avere più di un tipo di modifica. Per esempio *KRAS*, una *GTPasi* di segnalazione può essere farnesilata e palmitoilata. Se entrambe le modifiche sono presenti *KRAS* viene direzionata alla membrana plasmatica, mentre solo farnesilazione la direziona al Golgi.

7.11 ADP-ribosilazione

Si intende per *ADP-ribosilazione* l'aggiunta post-traduzionale di 1 o più molecole di *ADP-ribosio*.

7.11.1 Sintesi di ADP-ribosio

ADP-ribosio viene prodotta da NAD^+ (nicotinammide adenina dinucleotide). Avviene un legame 5'-5'' tra adenosina e ribosio. Il legame è molto resistente all'attività nucleasica (è anche presente nel cap 5' del mRNA).

7.11.2 Mono ADP-ribosilazione

La mono *ADP-ribosilazione* avviene nell'arginina nella sequenza consenso *Arg-Ser-Glu-X-Glu*, da parte di mono-*ADP-ribosiltrasferasi*. È reversibile da mono-*ADP-ribosil-glicidrolasi*. Ha un ruolo nella regolazione del trafficking vescicolare e nella trascrizione. Il legame avviene sul gruppo NH_2^+ dell'arginina, ovvero 1''-*NH*.

7.11.3 Poli *ADP*-ribosilazione

La poli *ADP*-ribosilazione avviene su vari amminoacidi da parte di poli-*ADP*-ribosilpolimerasi *PARP*. Gli *ADP*-ribosio sono attaccati in maniera lineare o ramificata. È reversibile da poli-*ADP*-ribosil-glicoidrolasi *PARG*. Ha ruolo nella replicazione del DNA, nella riparazione del danno al DNA, nella modifica istonica (rilassamento della cromatina) e regolazione della trascrizione. In caso di catena ramificata il legame avviene 1''-2'', in caso di catena lineare tra 1''-2'.

7.11.4 *ADP*-ribosilazioni possibili

- Mono-*ADP*-ribosilazione.
- Multi mono-*ADP*-ribosilazione.
- Oligo-*ADP*-ribosilazione.
- Poli-*ADP*-ribosilazione lineare.
- Poli-*ADP*-ribosilazione ramificata.
- Multi poli-*ADP*-ribosilazione.
- Mista mono-, oligo- e poli-*ADP*-ribosilazione.

2

7.11.5 *Corinebacterium diphterium*

La patogenicità di *Corynebacterium diphterium* è dovuta alla mono-*ADP*-ribosilazione di diftamide dalla tossina difterica. La tossina difterica codifica geni che si originano da un batteriofago che viene ricombinata stabilmente nel genoma di ceppi patogeni di *C. diphterium*. Questa inattiva *eEF2* mono-*ADP* ribosilandola bloccando la sintesi delle proteine.

7.12 Modifica chimica diretta

Modifiche dirette di amminoacidi, non catalizzate da enzimi possono essere importanti per la regolazione, mentre altre causano danno alle proteine.

7.12.1 Specie reattive all'ossigeno *ROS*

ROS sono tipicamente eliminate dalla cellula da attività enzimatica. Nonostante questo in particolari condizioni di stress ossidativo *ROS* possono reagire con le proteine e altre molecole biologiche. Le cisteine e metionine sono particolarmente suscettibili a questa modifica, diventando rispettivamente acido sulfonico e metionina sulfossido.

7.12.2 Specie reattive all'azoto

I residui di amminoacidi possono essere modificati da specie reattive all'azoto come l'ossido nitrico $N = O$. In particolare la cisteina diventa S-nitrocisteina attraverso S-nitrosilazione e la tirosina diventa tirosina nitrosilata attraverso O-nitrazione. Le modifiche possono avere effetto sulla funzione delle proteine: la S-nitrosilazione di metalloproteasi di zinco può impedire il legame dello zinco attivando la proteina.

Capitolo 8

DNA mobile

8.1 Panoramica degli elementi trasponibili

Il DNA può muoversi nel o fuori dal genoma attraverso trasposizione (non specifica alla sequenza) o ricombinazione conservativa sito specifica *CSSR*. Gli elementi trasponibili o trasposoni sono sequenze di DNA che possono muovere intorno al genoma e inserirsi a siti obiettivo. Si dividono in:

- Autonomi: i trasposoni codificano le proteine necessarie al loro movimento nel DNA.
- Non autonomi: i trasposoni si basano su proteine sintetizzate da altri trasposoni autonomi.

Il primo trasposone umano identificato fu *LINE L1* era un inserzione nell'esone 14 del fattore *VIII* della coagulazione e causava emofilia: la persona non è capace di bloccare il sanguinamento. Molte malattie umane sono dovute a inserimenti di trasposoni e circa $\frac{1}{200}$ delle nascite umane hanno una trasposizione diversa rispetto ai genitori. La quantità di trasposoni attiva nel genoma è molto bassa e non si spostano molto: tra lo 0.5-1%. Possono anche aiutare a livello di espressione di geni essenziali.

8.1.1 Trasposoni e malattie umane

8.1.1.1 *LINE L1*

Sui geni che causano emofilia. *RP2* per retinite pigmentosa X-linked, causa problemi con la vista e mancanza della visione periferica. *APC* che causa cancro del colon. *HBB* che causa beta-talassemia che causa stanchezza e un trasporto ridotto di ossigeno.

8.1.1.2 Delezione da ricombinazione omologa di due elementi trasponibili

La zona tra due elementi dopo la ricombinazione tra i due elementi può essere eliminata dal cromosoma. Un esempio è la sindrome di Alport's che agisce sui reni con abbondanza di sangue nell'urina. La delezione del Von Willebrand factor nella cascata del pathway della coagulazione questa si lega al fattore proteico *VIII* e se mancante la sua attività impedisce la coagulazione.

8.1.2 Eventi di trasposizione

Gli elementi trasponibili, *TE* o jumping genes sono elementi di DNA mobili che si muovono da una posizione all'altra sullo stesso o su un altro cromosoma. Influenzano l'espressione genica, causando

8.1. PANORAMICA DEGLI ELEMENTI TRASPONIBILI

mutazioni con proteine non-funzionali e possono riorganizzare la struttura cromosomica, con un ruolo attivo e passivo nell'evoluzione. I trasposoni possono:

- Inserirsi in un esone:
 - Troncamento.
 - Esonizzazione.
 - Splicing alternativo.
- Creare un enhancer:
 - Cambio della specificità del promotore.
 - Sotto-regolazione.
 - Repressione.
- Inserirsi in un introne:
 - Esonizzazione.
 - Esonizzazione e troncamento.
 - Splicing alternativo.
- Silenziamento di un promotore o un enhancer:
 - Sotto-regolazione.
 - Repressione.
- Distruggere un promotore o un enhancer:
 - Up-regolazione.

8.1.2.1 Risvolti positivi

Possono essere positivi nella risposta allo stress: nel caso dei batteri un *TE* si replica prima di fare un salto e lasciare o dare più capacità alla cellula di rispondere a condizioni di stress.

8.1.2.2 Piante

Nelle piante, per esempio i *TE* sono molto abbondanti (60%) e possono rispondere a condizioni di stress per eliminare o alzare l'espressione genica per rispondere allo stress. Le piante utilizzano questo sistema per regolare l'espressione genica e le proteine in quanto per le piante è l'unico modo di creare un ambiente meno sfavorevole.

8.1.2.2.1 Mais Il mais in natura originalmente è nero e non giallo. I colori diversi dei chicchi di mais sono conseguenze di *jumping genes*. Il chicco di mais contiene una base che si lega alla struttura di base. Al di sopra si trova l'embrione da cui si sviluppa la pianta, endosperma interno con la risorsa di carboidrati per l'embrione. La parte superiore o aleurone è la parte che contiene i pigmenti che determina il colore del mais che viene coperto dal pericarpo che tipicamente è trasparente. Studiando il mais McClintock nota mais di vari colori.

8.1.2.2.1.1 Colorazione del mais La colorazione del mais è una conseguenza del jumping gene che entra nella sequenza del gene che codifica per il colore. Si trova sul cromosoma 9, gene *C* che quando espresso crea un pigmento viola. Il colore viola è recessivo. Nel cromosoma 9 si notano gli elementi *Ac* e *Ds* due jumping genes. Questi si possono spostare all'interno del gene *C*. *Ac* codifica per una trasposasi che attiva la trasposizione di *Ds* e questo può trasporsi nel gene *C*. La trasposizione in un gene *C* causa un mutante formando il chicco giallo: *C^{Ds}* è dominante. Può ancora succedere un ulteriore evento di trasposizione, dove *Ac* attiva *Ds* che si sposta dal gene *C* che ritorna normale, portando a geni maculati. Pertanto se il trasposone *Ds* si inserisce nel gene *C* si nota un tessuto giallo. Trasposizione durante lo sviluppo precoce che ripristina il gene *C* causa una sezione colorata di viola grossa. Trasposizione durante lo sviluppo tardivo crea sezioni viola piccole.

8.1.3 Scoperta dei trasposoni

Gli elementi trasponibili vengono scoperti nel 1940 da Barbara McClintock e si nota come sono visibili nella pigmentazione dei chicchi di mais. Negli anni 60 invece si scoprono questi elementi nei batteri e negli anni 80 negli umani. I trasposoni non sono specifici alla sequenza e il 44% del genoma umano è occupato da trasposoni ed elementi ripetitivi simili ai trasposoni, ma solo una piccola proporzione di loro $< 0.05\%$ è attiva.

8.1.4 Caratteristiche della trasposizione

- La trasposizione di questi elementi non è specifica alla sequenza: tipicamente sono le condizioni di crescita che determinano dove questi elementi si spostano.
- La trasposizione dipende dall'enzima trasposasi che fa uscire l'elemento dal posto dove è e lo fa inserire nell'elemento obiettivo.
- L'inserzione di trasposoni risulta in una duplicazione del sito dove si inserisce il trasposone: la corta sequenza al sito di inserimento viene duplicata a causa della riparazione del danno al DNA.

8.1.4.1 Inserzione dei trasposoni

L'inserzione del trasposone può essere non-replicativa o replicativa:

- Non replicativa (classe *II*) taglia e incolla o trasposizione diretta: avviene solo l'escissione del DNA e si muove a un nuovo sito.
- Replicativa (classe *I*):
 - Tipo *I*: nick and paste la replicazione a DNA di un trasposone che non si muove su un sito diverso.
 - Tipo *II* o retro-trasposizione: il trasposone si replica attraverso un intermedio a cDNA.

8.1.5 Tipi di trasposoni

I trasposoni si trovano in tutti gli organismi e possono comprendere grandi parti dei loro genomi. La frequenza di nuove mutazioni attribuita ai trasposoni varia tra organismi. I tipi di trasposoni sono classificati in base a:

- Struttura e composizione.
- Proteine codificate.
- Metodo di trasposizione.

8.1.5.1 Solo a DNA cut and paste

I trasposoni più semplici a DNA non replicativi: codificano per una trasposasi e ripetizioni invertite a livello terminali o *TIR* che reclutano la trasposasi, si legano ad essa per liberare il trasposone. La trasposasi in caso di trasposone autonomo viene codificata all'interno. In assenza di trasposasi vengono detti non autonomi.

8.1.5.2 Elementi con lunghe terminazioni terminali *LTR*

Sono replicativi di classe *I* e sono presenti nei retrovirus. Sono autonomi e codificano diverse proteine come una trascrittasi inversa. Si muovono attraverso un intermedio a RNA. Alle terminazioni presentano dei long terminal repeat elements *LTR*.

8.1.5.3 Elementi non *LTR*

Sono replicativi di classe *I* e sono retrotrasposoni. Possono essere autonomi o non autonomi, codificano proteine con diverse proprietà. Si muovono attraverso un intermedio a RNA;

- Long interspersed elements *LINE*: codificano proteine che mediano la propria trasposizione.
- Short interspersed elements *SINE*: non codificano proteine per il proprio movimento e si basano su quelle codificate da *LINE*.

8.1.6 Elementi trasponibili nei procarioti

Sono a DNA e tre tipi:

- Elementi a insertion sequence *IS*.
- Trasposoni *Tn*.
- Fagi trasponibili *Mu* fago. Quando si integra nel genoma di *E. coli* può farlo in qualsiasi sequenza di coli e replicarsi ed amplificarsi nel suo genoma.

I trasposoni solo a DNA contengono sempre uno o più elementi di sequenza che sono affiancati a destra e sinistra da corte ripetizioni invertite riconosciute dalla trasposasi auto-codificata. Questi trasportano geni codificanti enzimi della trasposasi, fattori patogenici, resistenza agli antibiotici. Esistono virtualmente in tutti gli organismi e negli eucarioti contengono solo il gene della trasposasi.

8.1.6.1 Elementi a insertion sequence *IS*

Questi sono stati i primi trasposoni a essere identificati: *IS1* nell'operone del galattosio e in ceppi particolari fino a 19 copie. Solo in *E. coli* ne esistono fino a centinaia. Sono gli elementi trasponibili più semplici e lunghi tra le 700 e le 1500bp. Codificano per la trasposasi il cui gene è affiancato da sequenze di ripetizioni terminali dette *TIR*, *IR* o sequenze a mosaico *ME* a cui si lega la trasposasi. Le ripetizioni *IR* formano un complesso sinaptico con la trasposasi legata ad essi. Hanno principalmente un meccanismo replicativo ma possono averlo anche non-replicativo.

8.1.6.2 Trasposoni *Tn*

I trasposoni sono lunghi tra le 2100 e le 2300bp e hanno una composizione più complessa dagli elementi *IS*. Un elemento *Tn* è un trasposone composto: due elementi *IS* affiancano uno o più geni: quelli interni sono funzionali e quelli che codificano per le trasposasi si trovano all'interno degli elementi *IS*. Ogni elemento *IS* può trasporarsi indipendentemente o l'intero elemento *Tn* può agire come un trasposone singolo. Si trovano composti o non composti.

8.1.6.2.1 Composti Come *Tn5* e *Tn10* contengono segmenti centrali con geni per la resistenza agli antibiotici. Sono affiancati a entrambe le terminazioni da elementi *IS*, dello stesso tipo di un traspososne particolare. Gli elementi possono essere orientati in orientamenti inversi o lo stesso. La sequenza esterna *IS-50R* codifica per le trasposasi. La lunghezza totale può essere diverse migliaia di nucleotidi e si traspongono attraverso un meccanismo cut-and-paste non replicativo. Si traspongono attraverso il meccanismo non-replicativo cut-and-paste.

8.1.6.2.2 Non-composti Non contengono elementi *IS* ma semplici *IR* che affiancano il trasposone da entrambi le parti lunghe tra i 2 e i 20nt. Codificano per un gene di resistenza a un farmaco, una trasposasi e una resolvasi, rispettivamente per inserzione e ricombinazione. Si traspongono attraverso il meccanismo replicativo nick-and-paste.

8.1.6.2.3 Trasposone composto *Tn5* *Tn5* consiste di due abbastanza identici elementi *IS*:

- *IS-50L* guida il gene che codifica per la resistenza a canamicina, neomicina, bleomicina e streptomicina.
- *IS-50R* codifica la trasposasi.

Ogni *IS-50* è circondato da due sequenze invertite di 19bp dette outer and inner end *OE*, *IE* che definiscono la fine degli elementi *IS* a cui la trasposasi agisce. La metilazione di *IE* riduce l'abilità della trasposasi di legarsi e agire.

8.2 Panoramica dei trasposoni a DNA

Nei trasposoni solo a DNA non viene coinvolto nessun intermedio a DNA. La maggior parte di questi si muovono attraverso il meccanismo cut-and-paste non replicativo: il trasposone si escisse completamente e si inserisce nel target utilizzando una piccola quantità di replicazione per riparare i siti di inserimento. Alcuni si muovono attraverso il meccanismo replicativo nick-and-paste, ma solo nei batteri. Il trasposone rimane attaccato al DNA donatore ed è unito al target formando un coniugato o co-integrato attraverso ricombinazione omologa. Eventualmente il co-integrato si separa in due molecole, ognuna contenente un trasposone.

8.2.1 Trasposizione DNA-only non replicativa cut-and-paste

In questo caso si notano inverted repeats a sinistra e destra e la codifica della trasposasi al centro che viene tradotta. Nel target si trova un sito, tipicamente determinato da un numero di nucleotidi. Viene tagliato il target DNA, il trasposone si inserisce e si chiudono i gaps da DNA polimerasi e il sito di taglio viene replicato e a destra del trasposone.

8.2.1.1 Processo

Questo processo avviene per i *Tn* composti come *Tn5*, *Tn7* o *Tn10*. Avviene attraverso strutture di trasposoni:

1. Il DNA è riconosciuto da 2 trasposasi che si legano alle sequenze *IR*.
2. Le trasposasi dimerizzano e portano le terminazioni del trasposone vicine formando il traspososoma o complesso sinaptico. Una trasposasi che si lega a un *IR* va a tagliare quello opposto.

8.2. PANORAMICA DEI TRASPOSONI A DNA

Avviene pertanto un taglio incrociato, in modo da assicurare che le terminazioni siano vicine prima della rottura.

3. Questo attiva l'attività del trasposone e causa la sua uscita dal DNA donatore.
4. IL complesso transpososoma si lega al DNA target e le sue terminazioni 3'-OH fanno un attacco nucleofilo al sito target in presenza di Mg^{2+} causando una doppia rottura *ssDNA*.
5. I gaps a singolo filamento sono riempiti dall'attività di riparazione della DNA polimerasi dell'ospite.

8.2.1.2 Liberazione del trasposone

Ci sono 3 possibili meccanismi di auto-liberazione dei trasposoni.

8.2.1.2.1 *Tn7* Avviene una rottura a doppio filamento mediata da *TnsA*, che porta alla liberazione di due single strands.

8.2.1.2.2 Con hairpin *Tn10*, *Tn5* Avviene un taglio del filamento singolo e per liberarsi completamente per attaccare il target DNA si deve liberare l'altro filamento. I single strand il 3'-OH come donatore di elettroni si presenta come donatore di elettroni anche per l'altra rottura: fa nel DNA donatore un attacco nucleofilo per tagliare il filamento rimanente formando un hairpin. L'hairpin viene aperto da acqua che agisce come donatore di elettroni. In questo modo si può formare il legame fosfoesterico con il DNA accettore.

8.2.1.2.3 *Hermes* Avviene una reazione di transesterificazione con formazione di hairpin sul DNA donatore.

8.2.1.3 Taglio al sito donatore

Quando il DNA trasposone escinde dal sito donatore viene lasciato un gap che deve essere riparato. Il pathway potrebbe cambiare o non cambiare il sito donatore.

8.2.1.3.1 Non-homologous end joining *NHEJ* *NHEJ* riunisce le terminazioni. Il sito è raramente viene ripristinato allo stato originale, ma più spesso il sito è ripristinato in maniera approssimativa.

8.2.1.3.2 Homology-directed repair *HDR* *HDR* utilizza il cromatide fratello come un template di riparazione. Il trasposone è perfettamente ricopiato nel sito originale.

8.2.2 Trasposizione DNA-only replicativa nick-and-paste

Due esempi batterici sono *Mu phage* e *Tn3*.

8.2.2.1 *Tn3*

A livello della trascrizione la trasposasi codificata da *Tn3* catalizza la formazione di un cointegrato tra plasmidi donatori e recipienti. Durante il processo *Tn3* è replicato così che si trovi una copia dell'elemento ad ogni giunzione del cointegrato. La resolvasi prodotta dal gene *tnpR* risolve il cointegrato mediando la ricombinazione tra due elementi *Tn3*. Il plasmide donatore e recipiente si separano, ognuno con una copia di *Tn3*.

8.2.2.2 *Mu fage*

Il *Mu fage* è il trasposone più lungo e codifica per 55 proteine, con numerosi geni per la produzione delle proteine di testa e coda del fago. Il DNA si trova lineare nel fago. Dopo l'infezione si circolarizza nel DNA e si integra attraverso ricombinazione sito-specifica nel genome di *E. coli*. Dopo l'induzione dello stato litico si amplifica nel genoma di *E. coli* attraverso trasposizione per produrre 100 cromosomi virali e produce le proteine codificate. Vengono assemblate le particelle del fago che si escindono dal DNA e causano lisi. È l'unico elemento trasponibile che esiste in uno stato extracromosomico.

8.2.2.3 Processo

1. La trasposasi crea un taglio ssDNA ad entrambe le terminazioni 3' del DNA del trasposone.
2. Ogni terminazione 3'-OH si attacca al DNA obiettivo per creare una rottura a filamento singolo.
3. Il DNA donatore si trasferisce al DNA target formando un intermedio Shapiro con 2 interruzioni.
4. Una forcella di replicazione si forma all'interruzione di sinistra e il DNA viene sintetizzato per chiudere i gaps.
5. Si forma un co-integrato che si divide a metà a causa dell'attività ricombinante delle resolvasi auto-codificate.

8.3 Retrotrasposoni

I trasposoni coinvolgono intermedi a RNA e trascrittasi inverse. Si dividono in *LTR* (long terminal repeat) o elementi non *LTR*. I retrovirus *LTR* hanno fasi extracellulari e si trovano solo nei vertebrati. Elementi retroviral-like *LTR* si muovono solo intracellularmente e si trovano in funghi, piante ed animale. Elementi non-*LTR* si trovano in tutti i domini della vita. Possono inoltre costituire parti significative dei genomi eucarioti.

8.3.1 Trasposoni umani

Elemento	Numero di copie ×1000	Lunghezza totale <i>Mb</i>	Genoma %	Attività
Retrotrasposoni <i>LTR</i>	443	227	8.3	
Line (non- <i>LTR</i> autonomous)	868	462	20.4	
<i>LINE-1</i>	516	462	16.9	Attivo
<i>LINE-2</i>	315	88	3.2	
<i>LINE-3</i>	37	8	0.3	
Sine (non- <i>LTR</i> non-autonomous)	1558	360	13.3	
<i>Alu</i>	1090	290	10.6	Attivo con <i>LINE-1</i>
<i>MIR/MIR3</i>	468	69	2.5	
<i>SVA</i>	2.76	4	0.15	Attivo con <i>LINE-1</i>
Trasposoni a DNA	294	78	2.8	

8.3.2 Retrotrasposoni *LTR*

Gli *LTR* retrotrasposoni e i retrovirus sono simili nella struttura. Possiedono entrambe lunghe ripetizioni terminali. I retrovirus possiedono un elemento codificante il capside *ENV* necessario per la sopravvivenza al di fuori della cellula.

8.3.2.1 Attività

L'integrazione nel genoma ospite del retrotrasposone *LTR* è mediata da un integrasi codificata dall'elemento interno attraverso lo stesso meccanismo per gli elementi solo a DNA. La terminazione 3'-OH dell'elemento attacca il DNA target e la sequenza target è duplicata. Gli elementi *LTR* possono causare mutazioni di inserimento alterando l'espressione del gene avendo effetto sui promotori, enhancer e siti di splicing.

8.3.2.1.1 Processo

1. Gli elementi *LTR* sono trascritti in un mRNA provirus dalla cellula ospite.
2. Il provirus viene esportato nel nucleo e proteine sono prodotte nel mRNA>
3. Il mRNA e le proteine si assemblano in una particella simile a un virus.
4. Nella particella il RNA è convertito in cDNA da trascrittasi inverse.
5. Il DNA è integrato nel genoma dell'ospite da integrasi.
6. Nei retrovirus esiste uno stato extracellulare che richiede *ENV*>

8.3.2.1.2 Sintesi da trascrittasi inversa Le *LTR* che contengono corte sequenze ripetute *R* *U5* e *U3* sono i siti di inizio per la trascrittasi inversa. Sono sintetizzate da trascrizione inversa e passano di degradazione del RNA.

8.3.2.1.3 Escissione I trasposoni *LTR* si muovono attraverso trasposizione replicativa e pertanto non sono escissi dal sito donatore e non si deve riparare. Ricombinazione omologa può avvenire tra i due *LTR* di un trasposone causando la delezione del DNA tra gli *LTR*. Questo non è mobile.

8.3.3 Retrotrasposoni non-*LTR*

I trasposoni non *LTR* come *LINEs* e *SINEs* sono i più comuni nei genomi eucariot. Quando si muovono una copia di RNA si associa con il sito obiettivo e agisce come stampo per la trascrizione inversa. La maggior parte degli elementi *LINE* non hanno un promotore e non sono attivi con l'eccezione di *LINE-1*. *LINE ORF1* codifica per una proteina legante RNA, *ORF2* una proteina ibrida con attività endonucleasica e di trascrittasi inversa. Elementi *LINE* possiedono una sequenza poli-A nel loro DNA come i *SINE*. Gli elementi *SINE* non codificano le proteine richieste per la trasposizione e dipendono da *LINE*. Contengono *A* e *B* boxes per la trascrizione da RNA polimerasi III.

8.3.3.1 Attività

Gli elementi non-*LTR* si inseriscono attraverso trascrizione inversa al sito di inserimento. *ORF1* è un chaperon legante il DNA. *ORF2* è una singola proteina con attività endonucleasica e di trascrittasi inversa.

1. L'elemento *LINE* viene trascritto in RNA che viene tradotto.
2. Il sito target, regione ricca in *AT* è nicked dall'attività endonucleasica.
3. La coda poli-A del *LINE RNA* si inserisce nel dsDNA e si accoppia con la sequenza poli-T.
4. La trascrizione inversa si estende dal DNA utilizzando lo stampo a RNA.
5. Il RNA è degradato da *RNAasi H* /
6. IL secondo filamento di DNA del sito target viene rotto e il filamento prodotto agisce come stampo per la DNA polimerasi.
7. I processi di riparazione del DNA riempiono i gaps.

La trascrittasi inversa fallisce nel completare la terminazione 5' causando inserzioni troncate senza *ORF1* che non possono trasporre. La trasposizione ha meccanismo replicativo. Gli elementi *SINE* utilizzano le proteine codificate da *LINE-1* e la trasposizione avviene con lo stesso meccanismo.

8.3.3.2 *ALU* - elemento *SINE*

ALU (arthrobacter luteus restriction endonuclease element) è lungo 280bp e si traspone nel genoma umano causando malattie e cancro. Viene trascritto dalla RNA polimerasi III.

8.3.3.2.1 Integrazione L'integrazione di *ALU* può essere deleteria se distrugge l'espressione genica. Contribuisce inoltre a una diversità nell'espressione genica fornendo siti di legame per fattori di trascrizione. Possono introdurre eventi di ricombinazione come duplicazione o delezione di segmenti cromosomali. Accelera l'evoluzione e si trova spesso in sequenze geniche codificanti pre-mRNA come pare di introni o *UTR*.

8.3.3.3 Introni mobili di gruppo II

Un secondo gruppo di retrotrasposoni non-*LTR* sono gli introni mobili di gruppo *II*. Possono avere due tipi di mobilità: retrohoming e retrotrasposizione. Il retrohoming avviene con alta frequenza in un sito specifico con omologia considerevole alla sequenza intronica. La retrotrasposizione è più rara e avviene con frequenza più bassa a siti non specifici. Il primo passo per entrambi è lo splicing dell'elemento dal mRNA per formare un lariat intronico da maturasi. Nel retrohoming il RNA escisso fa splicing inverso in uno dei filamenti del DNA. L'endonucleasi codificata dall'introne *RME* fa un nick al DNA fuori dal sito di inserzione, una trascrittasi inversa fa una copia del DNA dell'introne attraverso la terminazione 3' del taglio del DNA come primer. Proteine batteriche rimuovono il RNA e una copia di DNA viene usata come filamento stampo dal lariat.

8.4 Controllo della trasposizione

La trasposizione può portare a diversità genica, ma troppa trasposizione è principalmente svantaggiosa. Ci sono pertanto meccanismi da parte della cellula capaci di prevenire eventi di trasposizione nel proprio genoma. La trasposizione può essere controllata positivamente o negativamente. Se le proteine dei trasposoni hanno un ruolo centrale nei passi catalitici quelle dell'ospite hanno un ruolo centrale nella trasposizione.

8.4.1 Meccanismi di controllo

8.4.1.1 Frequenza di trasposizione

La frequenza di trasposizione è collegata alla concentrazione di trasposasi. La trasposasi è sotto il controllo a livelli traduzionali e trascrizionali.

8.4.1.1.1 *Tn10* Caratteristiche specifiche del *IS10* mantengono la frequenza di trasposizione bassa:

- L'espressione della trasposasi è guidata da un promotore debole P_{in} .
- Un trascritto anti-senso viene prodotto da P_{out} .
- L'accoppiamento delle basi con il mRNA da P_{in} produce dsRNA che inibisce la traduzione.

8.4.1.2 Metilazione in P_{in}

La metilazione completa blocca la trasposizione impedendo l'espressione della trasposasi. Durante la replicazione del DNA, il DNA alla forcella di replicazione è emi-metilato causando un aumento dell'espressione dell'espressione della trasposasi da P_{in} . Siccome solo un filamento subisce la trasposizione ed è rotto, il filamento donatore può essere riparato da riparazione direzionata da omologia usando il braccio intatto come stampo.

8.4.1.2.1 *Drosophila* - trasposone P La trasposizione del trasposone *P* cut-and-paste della *Drosophila* è controllata da splicing alternativo. Nelle cellule germinali il mRNA è spliced correttamente in modo che la trasposizione possa avvenire permettendo la trasmissione dell'elemento *P* da una generazione alla prossima. Nelle cellule somatiche il terzo introne non viene rimosso producendo una trasposasi troncata non funzionale.

8.4.1.3 *Drosophila* - controllo nella linea germinale

Specifici loci in *Drosophila* producono grandi numeri di piccole molecole di RNA *piRNA* che bloccano la trasposizione utilizzando interferenza a RNA. Questi si accoppiano con le basi degli mRNA dei trasposoni.

8.4.1.3.1 RNA regolatori Regioni di DNA detti cluster di piRNA contengono trasposoni inattivi o frammentati che sono trascritti dalla RNA polimerasi II. Il lungo trascritto di piRNA viene processato dalla proteina Argonauta Piwi e vengono prodotti piRNA lunghi 23nt. Questi si legano al mRNA dei trasposoni attivi, causando la loro rottura da parte della proteina argonauta *Ago*. Il trasposone mRNA rotto si lega a un altro lungo piRNA trascritto parentale. La rottura e processamento causa una produzione di più piRNA, che sono importati nel nucleo con la proteina argonauta per silenziare la trascrizione dei trasposoni attivi a livello di DNA o trasposoni reclutando enzimi epigenetici.

8.5 Panoramica di *CSSR*

La ricombinazione sito-specifica conservativa avviene tra siti specifici alla sequenza con corte regioni di omologia. Questi siti comprendono:

- Gli spacer omologi con direzionalità.
- I siti di legame delle ricombinasi che affiancano ogni spacer.

Quattro ricombinasi si legano a due duplex. Il DNA viene tagliato in una maniera divisa, creando overhangs in modo che le molecole di DNA ricombinate contengono spacer eteroduplex. Il reclutamento per formare eteroduplex impone l'accuratezza dell'evento di ricombinazione. *CSSR* può causare diversi tipi di riarrangiamento genetico in base alla posizione relativa e orientamento dei siti di ricombinazione. Gli enzimi dei fagi che causano integrazione del genoma del fago nel DNA ospite sono detti integrasi ma non sono imparentati con le integrasi dei trasposoni. Due siti di ricombinazione nella stessa direzione su un cromosoma singolo circolare causa escissione. Due siti di ricombinazione che si trovano in direzioni opposte su un singolo cromosoma circolare causa inversione.

8.5.1 Ricombinasi sito-specifiche

Le ricombinasi specifiche alla sequenza sono topoisomerasi specifiche alla sequenza. Ce ne sono due famiglie di *CSSR* ricombinasi e in entrambe un amminoacido nucleofilo attacca il DNA. Non richiedono Mg^{2+} come co-fattore.

8.5.1.1 Ricombinasi tirosina

Le ricombinasi tirosina funzionano rompendo il DNA e formando un legame covalente *DNA-3'-P-tirosina* come intermedio di reazione.

8.5.1.1.1 Meccanismo La tirosina ricombinasi forma un dimero *R1-R2* e uno *R3-R4*. Avviene una rottura del filamento top da parte di *R1* e *R3*, in cui si forma il legame *3'-P-tirosina*. Le terminazioni libere 5' non legate alla ricombinasi possono invadere formando una giunzione di Holliday. Avviene una risoluzione verticale con una rottura del filamento bottom da *R2* e *R4*. Infine avviene uno scambio del filamento bottom per terminare la ricombinazione.

8.5.1.2 Ricombinasi serina

Le ricombinasi serina funzionano rompendo il DNA e formando un legame covalente *DNA-5'-P-serina* come intermedio di reazione.

8.5.1.2.1 Meccanismo La serina ricombinasi forma un dimero *R1-R2* e *R3-R4*. La rottura forma terminazioni *3'-OH*. Avviene uno scambio di partner e un legamento *3'-OH-5'-P*

8.5.1.3 Confronto

Le due famiglie non sono imparentate, sono strutturalmente diverse e con diversi meccanismi di reazione. La reazione da parte di entrambi i tipi di ricombinasi produce lo stesso risultato. Non usano *ATP*, un co-fattore bivalente e sintesi del DNA. In assenza di Mg^{2+} le catene laterali di istidine lisine o arginine clustered istigano l'attacco nucleofilo da parte di tirosina o serina.

8.5.2 Conversione *CSSR* di dimeri di DNA in monomeri

La conversione di un DNA circolare in due cerchi di DNA è una funzione importante di *CSSR*. La ricombinazione omologa può causare multimerizzazione. *CSSR* contribuisce alla segregazione dei cromosomi risolvendo cromosomi dimerici. Durante la replicazione del DNA di un plasmide su un genoma circolare la ricombinazione omologa tra il vecchio e il nuovo DNA sintetizzato può risultare in un DNA dimerico. Per riformare le due forme monomeriche deve avvenire una ricombinazione revertente sito-specifica come mediata dalla tirosina ricombinasi *XerCD*.

8.6 Integrazione ed escissione del batteriofago λ

Quando λ infetta *E. coli* e entra nello stato lisogenico avviene *CSSR* tra i siti *attP* e i siti batterici *attB*. Entrambi contengono una regione spacer *O* che è affiancata da diversi siti di legame per le tirosine ricombinasi codificate da λ dette integrasi. L'integrazione del DNA del fago ai siti *attP* e *attB* produce i siti *attR* e *attL*. L'escissione del profago, mediata dall'integrasi di λ ricrea i siti *attP* e *attB*. La struttura unica di ogni sito *att* causa in diverse richieste di reclutamento di proteine per la ricombinazione dei diversi substrati di DNA.

8.6.1 Integrasi λ *Int*

L'integrasi λ *Int* contiene due domini leganti il DNA: uno N terminale che si lega ai siti del braccio di *Int* e quello C terminale che si lega ai siti nucleari. Il piegamento del DNA è richiesto per il corretto legame di *Int*, la proteina *IHF* è richiesto. Il super-avvolgimento gioca un ruolo importante nella promozione del piegamento del DNA e l'avvolgimento del DNA *attP* intorno all'intrasoma *attP*.

8.6.2 Integrazione del fago λ

Il *attP+attB* con i suoi quattro tetrameri *R1*, *R2*, *R3* e *R4* forma l'intasoma integrativo. I domini catalitici dei due dimeri di *Int* causa uno scambio di filamento tra il filamento del fago e del batterio.

8.6.3 Escissione del fago λ

L'escissione richiede l'assemblaggio di *attL* e *attR* intasoma escissivo. Non sono sufficienti *IHF* e *Int* in quanto i bracci *P* e *P'* usati per piegare il DNA si trovano su diversi filamenti, pertanto si richiedono *Xis* e *FIS* per piegare il DNA. Dopo l'assemblaggio corretto i siti di *Int* nucleari sono piazzati affianco e può avvenire un cambio di filamento.

Capitolo 9

Strumenti e tecniche della biologia molecolare

9.1 Separazione di molecole biologiche

9.1.1 Separazione per elettroforesi su gel

La separazione di molecole biologiche avviene attraverso elettroforesi su Gel. Si possono usare due matrici sia per DNA, RNA che proteine, pur essendo queste interscambiabili.

9.1.1.1 Tipologie di gel

- Agarosio: formato da *D-galattosio* e *3,6-anidro-L-galattosio*, forma una struttura a maglie in gel. Un aumento di concentrazione causa una riduzione della dimensione dei pori.
- *PAGE*: polyacrylamide gel electrophoresis. Viene tipicamente utilizzato per le proteine.

9.1.2 Processo

9.1.2.1 Preparazione del gel

L'agarosio, polisaccaride di agarobiosio forma una trama a maglie. Durante la preparazione:

- Si pesa l'agarosio.
- Lo si sospende in buffer salini.
- Si porta a bollore per solubilizzare.
- Si pone il gel in una cassetta di plastica.
- Si aggiungono dei pettinini per formare i pozzetti dove verranno messi i campioni.

9.1.2.2 Separazione delle molecole

Dopo l'aggiunta dei campioni nei pozzetti si applica al gel una carica elettrica che porta alla formazione di due poli, uno positivo e uno negativo. RNA e DNA sono carichi negativamente a causa del fosfato e tenderanno a spostarsi verso l'elettrodo positivo. Il gel rallenta il movimento e permette una dimensione secondo grandezza:

- Molecole più grandi si troveranno in posizione più alta in quanto si muovono più lentamente.
- Molecole più piccole si troveranno in posizione più bassa in quanto si muovono più lentamente.

9.1.3 Colorazione delle molecole

Gli acidi nucleici vengono colorati con un fluorescente come l'etidio bromuro o *SYBR green*. Questa molecola è idrofobica e si intercala nel DNA inducendo mutazioni frameshift. Quando intercalata aumenta la fluorescenza di 20 volte quando illuminata da raggi *UV* grazie ai propri gruppi aromatici. Quando si intercala diminuisce il grido della molecola allungandola e inducendo un superavvolgimento negativo.

9.1.4 Stima della taglia del DNA

I frammenti di DNA sono comparati con degli standard di DNA detti markers, ovvero sequenze di DNA di una lunghezza prestabilita che determinano una scala.

9.1.5 *PFGE*

PFGE o elettroforesi a campo pulsato può risolvere grandi frammenti da *200kb* fino a *6000kb*. Viene utilizzata per grandi DNA come cromosomi. Il campo elettrico è pulsato e proviene da tre direzioni diverse. Queste si dispongono a forma esagonale con 6 elettrodi che lavorano a coppie opposte. La corsa del DNA è pertanto a zig-zag, ma il risultato è comunque una separazione verticale. Questo avviene in quanto il tempo di pulsazione di ogni coppia di elettrodi è equivalente.

9.1.6 Sodium dodecyl polyacrylamide gel electrophoresis

Sodium dodecyl polyacrylamide gel electrophoresis o *SDS-PAGE* utilizza *SDS*, un detergente che si lega alle zone idrofobiche delle proteine denaturandole. La proteina in questo modo torna ad avere una struttura primaria con cariche negative uniformi. Questo permette uno spostamento delle proteine dovuto unicamente alla loro massa. La denaturazione avviene grazie anche a un agente riducente, il *β -mercaptoetanololo*, che rompe i legami disolfuro quando riscaldato a 95°C per *5min*. Vengono utilizzati marker come scala e coloranti come nitrato d'argento e coomassie blue.

9.1.7 Concentrazione di gel

La concentrazione del gel viene determinata in base alla grandezza dei frammenti che si vogliono separare: maggiore la concentrazione più grandi i frammenti.

9.2 Amplificazione di sequenze di RNA e DNA

9.2.1 Panoramica

L'amplificazione delle sequenze di DNA o RNA specifici si rende necessaria per l'esaminazione o per una manipolazione ulteriore di date molecole.

9.2.2 Polymerase chain reaction

La polymerase chain reaction o *PCR* viene usata per amplificare DNA di interesse attraverso DNA polimerasi stabili e attive ad alte temperature ricavate da *Thermus aquaticus*. Coinvolge 3 passaggi principali ripetuti tra le 20 e le 40 volte. Il numero di molecole raddoppia ad ogni passaggio.

1. Denaturazione: del DNA a 95°C.
2. Annealing dei primers: a 50-60°C.
3. Sintesi: del nuovo DNA 72°C per 20 secondi per ogni *kb*.

I termociclatori automatizzano il processo.

9.2.3 Temperatura di annealing

Per calcolare la temperatura di annealing T_{ann} si deve considerare la temperatura di melting T_m :

$$T_{ann} = T_m - 2^\circ\text{C}$$

La temperatura di melting si calcola come:

$$T_m = [4^\circ\text{C} \cdot (n_G + n_C)] + [2^\circ\text{C} \cdot (n_T + n_A)]$$

9.2.4 *PCR* mutagenica

La *PCR* nonostante sia principalmente utilizzata per amplificare una sequenza di DNA fedelmente può anche essere usata per introdurre mutazioni o aggiungere sequenze alla fine delle molecole. Questo può per esempio essere fatto alle terminazioni dei primer. Queste presenteranno uno stampo per i cicli successivi al primo.

9.2.5 Amplificazione basata su *PCR* di RNA attraverso *cDNA*

Il RNA non può essere clonato direttamente, pertanto un DNA complementare deve essere sintetizzato attraverso una trascrittasi inversa. In quanto le terminazioni di una molecola di RNA potrebbero essere sconosciute a causa di splicing alternativo si utilizza *RACE* o rapid amplification of cDNA ends.

9.2.5.1 Rapid amplification of cDNA ends

Rapid amplification of cDNA ends o *RACE* è una procedura di *PCR* modificata che permette l'identificazione di terminazione di molecole di RNA. Nella *RACE 3'* un primer con una regione di poli-T e una sequenza a scelta viene annealed alla coda di mRNA poli-A. La trascrittasi inversa crea cDNA e il secondo filamento a DNA viene sintetizzato estendendo da un primer interno. Il primer interno e il primo possono essere usati successivamente in una *PCR* standard.

9.3 Clonaggio di DNA

9.3.1 Panoramica

È oggi possibile clonare geni. La maggior parte delle tecniche richiede che il DNA sia isolato e modificato.

9.3. CLONAGGIO DI DNA

9.3.1.1 Vettori di DNA

9.3.1.1.1 Plasmidi Introducendo un frammento di DNA in un plasmide molte copie del frammento possono essere generate. I plasmidi contengono tipicamente un marcatore selezionabile che permette la crescita delle sole cellule che lo hanno integrato. Possono integrare frammenti di DNA fino a 15kb.

9.3.1.1.2 Bacterial artificial chromosome I bacterial artificial chromosome *BAC* possono possedere inserzioni fino a centinaia di chilobasi.

9.3.1.1.3 Shuttle vector I shuttle vectors sono vettori capaci di crescere in diversi organismi. Necessitano di un origine di replicazione per ogni organismo in cui crescono e possiedono un marcatore diverso per ogni organismo

9.3.1.1.4 Marcatori per la selezione I marcatori per la selezione sono tipicamente geni di resistenza batterica.

9.3.1.2 Utilizzi

I cloni di DNA possono essere utilizzati per complementare una mutazione o per aggiungere tag a proteine in modo che possano essere purificate o visualizzate all'interno di cellule.

9.3.2 Isolamento dei plasmidi

1. Sospensione delle cellule.
2. Lisi delle cellule.
3. Neutralizzazione del lisato.
4. Legame del DNA dal surnatante a una matrice.
5. Lavaggio della matrice.
6. Diluizione del DNA.

9.3.3 Enzimi di restrizione

Fondamentali per un clonaggio ottimale sono gli enzimi di restrizione. Questi agiscono su sequenze palindromiche e sono derivati dai batteri che li usano come meccanismo di difesa contro i batteriofagi. Sono distinti in base al taglio che operano:

- Blunt *HpaI*.
- 3' overhangs *KpnI*.
- 5' overhangs *EcoRI*, *HINDIII*

Gli ultimi due formano delle estremità sticky coesive.

9.3.4 Strategie per il clonaggio genico

9.3.4.1 Ligazione

Nella ligazione il frammento e il vettore sono tagliati attraverso enzimi di restrizione in modo che le terminazioni delle molecole abbiano terminazioni complementari. Quando i due frammenti sono incubati insieme le terminazioni complementari formano accoppiamenti da basi e le due molecole si

9.4. MANIPOLAZIONE GENOMICA

uniscono. Queste possono essere unite covalentemente da DNA ligasi. Questo viene poi introdotto nei plasmidi per l'amplificazione attraverso trasformazione.

9.3.4.1.1 Trasformazione L'organismo più utilizzato per il clonaggio è *E. coli*.

9.3.4.2 Gibson cloning

Nel processo di Gibson cloning diversi frammenti possono essere clonati simultaneamente. I frammenti sono progettati in modo che possiedano terminazioni che corrispondono le terminazioni degli altri frammenti e del vettore. Sono mischiati con il vettore per permettere l'anneal e i gap sono riempiti da DNA polimerasi e i frammenti sono ligati insieme.

9.3.4.3 Ricombinazione omologa

I geni possono essere inseriti nei vettori attraverso ricombinazione omologa. I siti di ricombinazione sono inseriti al prodotti di PCR nello stesso modo aggiungendo siti di restrizione. La ricombinazione omologa avviene tra due regioni di DNA omologhe. Diverse sequenze alle due terminazioni dell'inserzione controllano la direzione dell'inserimento del frammento. Gli enzimi di ricombinasi sono necessari per la reazione in vitro.

9.3.5 Mutagenesi sito-diretta

La mutagenesi sito-diretta viene usata per alterare la sequenza di DNA in posizioni specifiche per creare proteine mutanti e studiare la loro funzione. Si effettua *PCR* dove vengono creati i primers e vengono usate polimerasi ad alta fedeltà. Le metilazioni del DNA vengono rimosse dall'enzima *DpnI*. La molecola viene amplificata e usata per trasformazione.

9.3.6 Libreria genica

Si intende per libreria genica una collezione di geni clonati in un plasmide. Vengono usate per identificare una mutazione. Si trasforma un mutante con la libreria in modo che ogni individuo possieda un vettore e gli organismi con la mutazione soppressa conterranno il vettore desiderato. Questo clone può essere isolato dalle cellule trasformate. Le librerie possono inoltre essere costruite da mRNA espressi.

9.4 Manipolazione genomica

9.4.1 Panoramica

9.4.1.1 Forward genetics

Nella forward genetics un mutante viene identificato grazie a un difetto in un certo processo. Il gene mutante identificato permette di comprendere meglio il processo e pathway.

9.4.1.2 Reverse genetics

Nella reverse genetics un gene viene mutato e si esamina la sua funzione. Mutazioni casuali creano una via molto comune molto usata per esaminare il genoma. Le mutazioni possono essere indotte con metodi mutageni. La ricombinazione omologa permette di creare mutazioni su un gene specifico.

9.4.2 Inserimento di un trasposone

L'inserimento di un trasposone avviene casualmente nei genomi e i geni possono essere rotti se un trasposone interrompe la loro sequenza. Un componente contiene il gene della trasposasi controllato da un promotore, mentre l'altro elemento un marcatore selezionabile. Dopo la trasposizione la sequenza che circonda il trasposone può essere isolata e usata per identificare la sua locazione nel genoma.

9.4.3 Genomi delle piante

I genomi delle piante sono stati mutati estensivamente utilizzando inserti a *T*-DNA. Le linee di *Arabidopsis* sono disponibili con inserti in quasi tutti i geni. I plasmidi *Ti* induttori del tumore da *Agrobacterium tumefaciens* contiene un'origine di replicazione con il T-DNA affiancato da sequenze di inserimento e i fattori di virulenza codificano un enzima di trasformazione. In natura la sezione *T* codifica ormoni delle piante che inducono una divisione cellulare incontrollata che causa tumori crown gall. Il T-DNA può essere sostituito da DNA di interesse.

9.4.3.1 Mutazioni

Le mutazioni avvengono quando *Agrobacterium* inserisce il plasmide *Ti* attraverso ricombinazione non omologa illegittima nel genoma della pianta.

9.4.4 Ricombinazione omologa

La ricombinazione omologa è utilizzata per creare cambi specifici. Un costrutto lineare viene inserito in un organismo e le regioni omologhe si allineano e ricombinano in modo che il DNA dell'ospite viene sostituito dal DNA costruito. La sequenza originale viene degradata. Addizioni o delezioni possono essere ottenute se le sequenze di DNA omologhe differiscono. Viene anche usata per aggiungere tag a geni.

9.4.4.1 Aumentare l'efficienza

La bassa efficienza di ricombinazione omologa può essere aumentata inducendo una rottura a doppio filamento nel locus di interesse utilizzando sistemi basati su *CRISPR* batterica.

9.4.4.1.1 Clustered regulatory interspaced short palindromic sequences Clustered regulatory interspaced short palindromic sequences o *Crispr* utilizza piccole lunghezze di DNA obiettivo e RNA non codificanti per dirigere l'attività della nucleasi *Cas9* verso il luogo desiderato.

9.5 Identificare la composizione di molecole biologiche

9.5.1 Sequenziamento del DNA

9.5.1.1 Sequenziamento chain-terminated dideoxy

Nel sequenziamento chain-terminated dideoxy un singolo primer viene annealato alla sequenza di interesse e il primer viene esteso con la DNA polimerasi I. La soluzione di reazione contiene tutti e 4 i deossi nucleotidi e una piccola quantità di dideoossi nucleotidi ognuno etichettato con una diversa fluorescenza. L'estensione del polinucleotide si ferma se un nucleotide dideoossi viene aggiunto alla

catena in quanto non possiede un gruppo 3'-OH disponibile per la reazione. La reazione genera un insieme di molecole a diverse grandezze ognuna terminante con un dideossinucleotide.

9.5.1.1.1 Lettura dei risultati Originariamente i frammenti etichettati radioattivamente venivano fatti correre su un gel *PAGE* molto lungo e letti ad occhio. Metodi odierni utilizzano una macchina sequenziatrice che separa i polinucleotidi attraverso elettroforesi su *PAGE* e misura il colore di ogni dimensione del frammento con un laser. I risultati accurati fino a 500bp a causa di esaurimento di nucleotidi sono mostrati sul cromatogramma.

9.5.1.2 Ottenere la sequenza genomica completa

La sequenza può essere ottenuta assemblando frammenti sequenziati.

1. Il DNA genomico viene rotto in piccoli frammenti di restrizione a cui sono attaccati linkers.
2. I frammenti sono amplificati secondo la sequenza linker.
3. I frammenti sono sequenziati.
4. Computazionalmente le sequenze linker sono rimosse dai dati di sequenza e le regioni sovrapposte sono usate per determinare l'ordine e l'assemblaggio delle sequenze di frammenti.

Viene richiesto sovra-sequenziamento a causa della sovrapposizione dei frammenti e per assicurare copertura completa ed accuratezza.

9.5.1.2.1 Problematiche

- Un sequenziamento 8× è richiesto per ottenere un'accuratezza del 99.9%.
- Le regioni ripetitive come centromeri e telomeri sono difficilmente allineabili ridottrando *YACs*.
- La velocità viene aumentata da macchine di sequenziamento moderne.

9.5.1.3 Next-generation sequencing

9.5.1.3.1 Illumina sequencing Durante l'illumina sequencing il DNA genomico viene frammentato, vengono legati adattatori e il DNA è denaturato. I frammenti sono aggiunti a una superficie con primer complementari agli adattatori legati. Il DNA viene amplificato da un bridge di dsDNA, viene denaturato per creare cluster di molecole amplificate identiche. Le molecole in ogni cluster sono sequenziate simultaneamente. Un primer specifico e nucleotidi etichettati fluorescentemente vengono utilizzati per la sintesi. Dopo ogni addizione il nucleotide aggiunto viene determinato dalla sua fluorescenza.

9.5.2 Sequenziamento delle proteine

9.5.2.1 Degradazione di Edman

La degradazione di Edman è una degradazione a passi dalla terminazione *N*. IL fenilisotiocianato è utilizzato per etichettare i residui *N* terminali in condizioni debolmente alcaline. Il risultante feniltioantoin *PHT*-derivato induce instabilità nel legame peptidico *N* terminale tra i primi due amminoacidi che possono essere idrolizzati causando la rottura del legame. Il complesso *PTH-aa* viene identificato attraverso cromatografia *HPLC*.

9.5.2.2 Spettrometria di massa

La spettrometria di massa viene usata per identificare proteine in miscele complesse. Le molecole ionizzate attraverso matrix assisted laser desorption/ionization *MALDI* o electron-spray ionization *ESI* permettono un calcolo del rapporto tra massa e carica. Il rapporto permette di determinare la massa di ogni frammento e quindi l'amminoacido in esso contenuto.

9.5.2.2.1 Altre ionizzazioni Le ionizzazioni *MALDI* e *ESI* sono soft e non rompono i peptidi. Ionizzazioni hard come *EI* sono usate per piccole molecole formando lo ione molecolare radical cationico.

9.5.3 BLAST

BLAST è un tool informatico utilizzato per comparare sequenze di acidi nucleici e di sequenze proteiche. Si confronta una query con tutte le sequenze depositate nel database. Quelle simili alla nostra query si dicono subjects e sono ordinate con scores di similarity.

9.6 Identificazione di specifiche molecole di DNA

9.6.1 Panoramica

Enzimi di restrizione tagliano il DNA in frammenti di certa lunghezza che possono essere comparati con frammenti di lunghezza conosciuta utilizzando elettroforesi su gel d'agarosio.

9.6.2 Ibridazione southern blot

Il southern blot viene utilizzato per identificare una specifica sequenza di DNA all'interno di un campione.

9.6.2.1 Processo

1. Il DNA viene tagliato con enzimi di restrizione.
2. I frammenti vengono separati con elettroforesi per dimensione.
3. I frammenti denaturati in filamenti singoli in ambiente alcalino presentano basi
- esposte e trasferite su una membrana.
4. Sonde di ssDNA marcato radioattivamente con sequenza complementare ibridano i frammenti con la sequenza target.
5. I frammenti ibridati vengono identificati con un autoradiogramma.

9.6.3 DNA fingerprint

Il genoma di persone non imparentate differisce per lo 0.1% a causa di mutazioni o polimorfismi. Le seconde sono una variazione di DNA senza fenotipo clinico che avviene in regioni del genoma che non codificano proteine.

9.6.3.1 Restriction fragment length polymorphism

Restriction fragment length polymorphism *RFLP* è identificato rompendo il DNA in frammenti con enzimi di restrizione. La lunghezza del frammento di restrizione è alterata se la variante genetica crea o distrugge un sito di restrizione.

9.6.3.1.1 Cause di varianza

9.6.3.1.1.1 Cambi a singola base Cambi a singola base nella sequenza nucleotidica *SNP* rappresentano il 90% del genoma umano e possono causare mutazioni.

9.6.3.1.1.2 Cambi nel numero di ripetizioni di certe sequenze Il numero variabile di ripetizioni tandem *VNTR* sono sequenze corte di DNA ripetute in tandem a specifiche locazioni. Il numero è variabile per individuo e causa una identificazione univoca per ogni individuo.

9.6.3.1.2 Utilizzo La rottura da parte di enzimi di restrizione del genoma seguita da ibridazione southern blot rivela frammenti di diversa lunghezza in base a quante ripetizioni di ogni frammento sono contenute.

9.6.3.2 Analisi *RFLP*

L'analisi *RFLP* viene usata in ambito forense per identificare un individuo grazie a materiali che contengono DNA come:

- Regioni mini satellite subtelomeriche ricche in *GC*.
- Regioni micro satellite regioni distribuite nel genoma.

Questi pattern sono ereditati attraverso regole mendeliane.

9.6.3.3 DNA fingerprinting

Il DNA fingerprinting avviene attraverso ibridazione southern blot in cui il DNA viene tagliato ai lati di *VNTR*. Si fa correre la digestione su gel e si fa un blot. La probe mostra pattern complessi con i microsatelliti marcati. Le analisi forensi sono effettuate con amplificazione *PCR*. I primer vengono generati in modo che siano ai lati dei loci.

9.6.3.3.1 Utilizzo dell'analisi *RFLP* per diagnosticare l'anemia falciforme Una mutazione del codone 6 dell'esone 1 è un sito di restrizione per l'enzima *MstII*. La *PCR* amplifica l'esone 1 che presenta due frammenti per sano e uno per malato. Le bande vengono marcate con sonde specifiche per l'esone 1.

9.6.4 Ibridazione di colonie

L'ibridazione di colonie permette di identificare un clone da una libreria batterica cresciuta su plates. È utile per lo screening di una libreria genetica per un singolo gene. Avviene una lisi con *SDS* e denaturazione del DNA che viene ibridato con sonde radioattive. Viene utilizzata la *PCR* con primer specifici.

9.6.5 Cariotipo

Il kariogramma viene ottenuto e reso visibile con colorazione *Giemsa* utilizzata nei test di diagnosi prenatale. Viene ottenuto attraverso tecniche invasive come amniocentesi o come amplificazione del materiale genetico nel sangue materno. Questo consente di visualizzare i cromosomi in metafase con pattern di bandeggio caratteristici. Le *G bands* sono bande costituite da eterocromatina, scure, mentre le chiare da eucromatina.

9.6.5.1 Processo

Il DNA viene colorato da un mix di blu di metilene che si lega al legame fosfodiesterico, eosina che si lega a siti cationici. Questo permette l'identificazione di anomalie cromosomiche.

9.6.5.2 Aberrazioni

Poliploidia, aneuploidia bilanciata e sbilanciata con perdita di materiale genetico. Inserzioni, delezioni. Mosaicismo. Disomia uniparentale in cui due cromosomi omologhi provengono dallo stesso genitore.

9.6.6 Fluorescent in situ hybridization

Fluorescent in situ hybridization o *FISH* fissa le cellule a un vetrino e il DNA viene sondato con un'analisi southern. Le sonde contengono tag fluorescenti.

9.6.7 Spectral karyotyping

Lo spectral karyotyping o *SKY* è una modifica di *FISH* in cui i cromosomi metafasici sono isolati e ognuno di essi è colorato diversamente in modo da semplificare l'analisi *Giemsa*.

9.6.8 Array comparative genomic hybridization

Array comparative genomic hybridization *aCGH* viene usata per identificare piccoli cambiamenti nella sequenza di DNA. Può misurare la variazione di numero di copie. Un microarray è un insieme di frammenti di DNA fissato a uno slide che rappresentano il genoma. Due campioni di DNA sono confrontati uno etichettato verde e uno rosso. Sono mischiati e ibridizzati a un microarray. Quando l'array è scansionato ogni punto avrà fluorescenza. Il colore è dipendente dal numero di copie relativo nel campione. Giallo in entrambe o rosso o verde se è maggiore in una delle due. Il rapporto tra i numeri di copie può essere confrontato lungo cromosomi o nel genoma.

9.7 Identificazione di specifiche molecole di RNA

9.7.1 Panoramica

L'espressione genica è spesso controllata dai livelli di trascrizione, pertanto è utile esaminare i livelli di RNA.

9.7.2 Ibridazione northern blot

Specifiche sequenze di RNA possono essere identificate attraverso l'ibridazione northern blot. Sono svolti nella stessa maniera dei southern blot ma il RNA non deve essere frammentato. La quantità di RNA nel campione può essere determinata in molti modi.

9.7.2.1 Reverse transcription quantitative polymerase chain reaction

Reverse transcription quantitative polymerase chain reaction *RT-qPCR* crea una copia di cDNA del RNA generato. Questo viene amplificato con una reazione a PCR con primer specifici. La quantità di prodotto generata è proporzionale alla quantità di RNA nel campione originale. La quantità di prodotto può essere determinata dopo ogni ciclo.

9.7.2.2 Metodo delta-delta Ct di quantificazione

La quantificazione del metodo delta-delta Ct, dove Ct è il primo ciclo della reazione in cui la fluorescenza è emessa la disopra della soglia.

1. Si calcola il ΔCT per normalizzare il campione escludendo variazioni di quantità di campione.

$$\Delta CT = CT_{medio\ GENE\ TARGET} - CT_{MEDIO\ GENE\ HOUSEKEEPING}$$

2. Calcolo del $\Delta\Delta CT$, confrontando i campioni incogniti rispetto a uno scelto come controllo.

$$\Delta\Delta CT = \Delta CT_{campione\ INCOGNITO} - \Delta CT_{campione\ CONTROLLO}$$

3. Calcolo del fold change, se il valore è maggiore di 2 il gene target è più espresso, meno espresso se minore di 0.5, se si trova tra questi due valori la differenza non è significativa.

$$2^{-\Delta\Delta CT}$$

9.7.3 Microarray per profilo trascrittomico

La tecnica dei microarray per profilo trascrittomico permette una quantificazione del mRNA. I microarray comprendono tutti i geni dell'organismo, pertanto è possibile fare una esamina dell'espressione genica del genoma. Se un trascritto viene espresso differenzialmente il microarray appare verde o rosso, altrimenti giallo. Viene sostituita da *RNA-seq*.

9.7.4 Trascrizione di un gene reporter

Per determinare l'espressione di un gene si possono fondere le sue regioni regolatorie a un gene reporter come β -galattosidasi o *GFP*.

9.8 Identificazione di specifiche proteine

9.8.1 Ibridazione western blot

L'ibridazione western blot utilizza anticorpi per individuare proteine specifiche.

1. Le proteine sono separate per dimensione attraverso elettroforesi su gel.
2. Le proteine sono trasferite dal gel su un filtro specializzato che lega le proteine.
3. Il filtro è esposto a anticorpi primari specifici della proteina di interesse.
4. Il filtro è esposto ad anticorpi secondari specifici all'anticorpo primario amplificando il segnale.
5. Gli antibiotici secondari sono coniugati con una molecola che può essere utilizzata per identificare la presenza dell'anticorpo.

9.8.1.1 Proteine senza anticorpi

Non è sempre possibile creare anticorpi per proteine. In questi casi un tag terminale può essere aggiunto alla proteina e usare anticorpi che lo riconoscono.

9.8.2 Stable isotope labelling with amino acid in culture

Stable isotope labelling with amino acid in culture *SILAC* viene usato per determinare la quantità relativa di proteine in due campioni attraverso spettrometria di massa. Le due colture sono fatte crescere in condizioni diverse, una contrassegnata con arginina pesante ^{13}C e l'altra con amminoacidi normali. Ogni frammento che contiene un arginina pesante sarà più pesante di $8Da$ nella spettrometria di massa. L'altezza relativa dei due picchi può essere usata per determinare la concentrazione relativa di proteine. In questo modo molte proteine possono essere analizzate insieme mostrando come cambi globali in espressione siano colpiti dal trattamento.

9.9 Identificazione di interazioni tra molecole

9.9.1 Panoramica

Dopo aver identificato una molecola è utile identificare quelle che interagiscono con essa.

9.9.2 Co-purificazione

Nella co-purificazione quando una molecola viene purificata in base al suo tag le molecole con cui interagisce rimangono attaccate e co-purificate.

9.9.3 Co-immunopurificazione

Nella co-immunopurificazione biglie ricoperte di anticorpi vengono usate per far precipitare la proteina e le molecole con cui interagisce. Questo metodo può essere usato per determinare le interazioni tra proteine e altre molecole. Può in particolare essere usata per identificare a che regioni di un cromosoma una proteina DNA-binding si attacca.

9.9.3.1 Chromatin immunoprecipitation

Si intende per chromatin immunoprecipitation *ChIP* il processo di co-immunoprecipitazione utilizzato per identificare a che regioni di un cromosoma una DNA-binding protein si attacca.

9.9.3.1.1 Processo

9.9. IDENTIFICAZIONE DI INTERAZIONI TRA MOLECOLE

1. Vengono trattate le cellule con formaldeide per creare un cross-link tra le proteine e il DNA.
2. Il DNA viene rotto in corti frammenti attraverso sonificazione.
3. Vengono usati anticorpi contro la proteina specifica per immunoprecipitarla.
4. Cross-linking inverso attraverso *SDS* a 65°C viene usato per separare le proteine dal DNA.
5. Vengono rimosse le proteine attraverso la proteasi *K*.
6. Viene identificata la sequenza attraverso sequenziamento *ChIP-seq* o ibridazione a un microarray *ChIP-chip*.

9.9.4 Analisi *CLIP*

L'analisi *CLIP* analizza i cross-link tra RNA e proteine. È simile a *ChIP*.

9.9.4.1 Processo

1. Avviene un crosslink tra proteine e RNA in cellule viventi.
2. Gli RNA crosslinked sono immunoprecipitati e digeriti.
3. Il crosslinking viene rimosso, il RNA purificato e il cDNA viene sintetizzato prima del sequenziamento o ibridazione.

9.9.5 Electrophoretic mobility shift assay

Electrophoretic mobility shift assay *EMSA* viene usato per determinare la forza di interazione tra le molecole. Una molecola di DNA si muove più lentamente quando legata a una proteina. Si nota come aumentando la concentrazione della proteina più molecole di DNA si muovono più lentamente. Analizzando diverse concentrazioni si può determinare la curva di legame della proteina.

9.9.6 DNA footprinting

DNA footprinting permette di identificare le regioni di DNA dove si lega la proteina. Una terminazione del DNA viene etichettata. La proteina si lega al DNA che viene digerito con un enzima che taglia casualmente. Questo in modo che ogni molecola di DNA venga tagliata in media una sola volta. Ogni regione di DNA legata a una proteina viene protetta dalla digestione. Questa forma il footprint della DNA binding protein. La separazione dei frammenti attraverso elettroforesi mostra delle bande mancanti dove la proteina si lega.

9.9.7 Colocalizzazione proteica attraverso microscopia a fluorescenza

Coloranti con colore diversi possono esser utilizzati alla volta, in modo da visualizzare diverse proteine. Le cellule vengono fissate e permeabilizzate prima del trattamento con anticorpi. Ne vengono utilizzati due, uno che riconosce l'epitopo e uno che è specifico al primario e con il fluoroforo attaccato.

9.10 Sequenziamento di molecole di DNA e RNA attraverso nanopori

9.10.1 Panoramica

Il sequenziamento di molecole di DNA e RNA attraverso nanopori non necessita di marcatura dei nucleotidi o di amplificazione di DNA ed è pertanto portatile, rapido e a basso costo.

9.10.2 Nanopori

Un nanoporo è un buco molto piccolo di circa $1.5nm$ in lunghezza in una membrana. Questo è maggiore di un singolo filamento ma minore di una doppia elica.

9.10.2.1 Creazione

Un nanoporo può essere creato attraverso sintesi da una molecola biologica o creando buchi su una superficie solida utilizzando un raggio di elettroni.

9.10.3 Sequenziamento a nanopori

Il sequenziamento a nanopori segue un principio base in cui un filamento di RNA o DNA sono guidati attraverso un nanoporo attraverso elettroforesi. Mentre ogni nucleobase passa attraverso il poro la corrente viene colpita e il cambio permette la lettura della sequenza. Ogni base infatti causa un cambio di corrente specifico e un tempo di attraversamento univoco. Un esempio di nanoporo biologico è la α -emolisina, mentre a strato solido sono silice SiO_2 o grafite.