

# Biologia molecolare della cellula

Giacomo Fantoni

Telegram: @GiacomoFantoni

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

**7 aprile 2020**

# Indice

## I Introduzione alla cellula 12

### 1 Introduzione 13

- 1.1 Microscopia . . . . . 13
  - 1.1.1 Risoluzione . . . . . 13
  - 1.1.2 Tecniche di microscopia ottica . . . . . 14
- 1.2 Storia . . . . . 14

### 2 Cellule e genomi 16

- 2.1 Caratteristiche universali della vita sulla Terra . . . . . 16
  - 2.1.1 Tutte le cellule conservano la loro informazione ereditaria nello stesso codice chimico lineare (DNA) . . . 16
  - 2.1.2 Tutte le cellule replicano la loro informazione ereditaria mediante polimerizzazione su uno stampo . . . . 17
  - 2.1.3 Tutte le cellule trascivono porzioni della loro informazione ereditaria nella stessa forma intermedia (RNA) . . . . . 17
  - 2.1.4 Tutte le cellule usano proteine come catalizzatori . . 18
  - 2.1.5 Tutte le cellule traducono RNA nello stesso modo . . 18
  - 2.1.6 Il frammento di informazione genica che corrisponde ad una proteina è un gene . . . . . 18
  - 2.1.7 La vita richiede energia libera . . . . . 19

- 2.1.8 Tutte le cellule funzionano come fabbriche biochimiche che impiegano le stesse strutture molecolari base . . 19

- 2.1.9 Tutte le cellule sono chiuse in una membrana plasmatica attraverso la quale nutrienti e materiali di scarto devono passare . . . . . 19

- 2.2 La diversità dei genomi e l'albero della vita . . . . . 19

- 2.2.1 Le cellule possono ottenere energia da una varietà di fonti di energia libera . . . 19

- 2.2.2 Alcune cellule fissano azoto e anidride carbonica per altre . . . . . 20

- 2.2.3 La più grande diversità biochimica esiste tra le cellule procariote . . . . . 20

- 2.2.4 L'albero della vita possiede tre rami principali: Batteri, Archei ed Eucarioti 20

- 2.2.5 Alcuni geni evolvono rapidamente, altri sono altamente conservati . . . . . 20

- 2.2.6 Nuovi geni sono generati da geni preesistenti . . . . . 21

- 2.2.7 La duplicazione genica permette la creazione di famiglie di geni imparentati in una stessa cellula . . . . 21

- 2.2.8 I geni possono essere trasferiti tra organismi . . . . 21

- 2.2.9 Molti geni sono comuni tra tutti i rami primari dell'albero della vita . . . . 22

- 2.2.10 Le mutazioni rivelano la funzione dei geni . . . . . 22

- 2.3 Informazione genetica negli eucarioti . . . . . 22

- 2.3.1 Le cellule eucariote potrebbero essersi originate come predatori . . . . . 22

- 2.3.2 Le moderne cellule eucariote si sono evolute da una simbiosi . . . . . 22

2.3.3	Gli eucarioti possiedono genomi ibridi . . . . .	23	3.2.5	Gli enzimi abbassano la barriera di energia di attivazione che blocca la reazione chimica . . . . .	28
2.3.4	I genomi eucarioti sono grandi . . . . .	23	3.2.6	Gli enzimi possono guidare i substrati lungo specifici cammini di reazioni . . . . .	29
2.3.5	Il genoma definisce il programma dello sviluppo multicellulare . . . . .	23	3.2.7	Come gli enzimi trovano i loro substrati, la rapidità dei movimenti molecolari . . . . .	29
2.3.6	Molti eucarioti vivono come cellule solitarie . . . . .	24	3.2.8	Il cambio di energia libera di una reazione determina se può avvenire spontaneamente . . . . .	29
<b>3</b>	<b>Chimica e bioenergetica della cellula</b>	<b>25</b>	3.2.9	Molecole vettore attivate sono essenziali per la biosintesi . . . . .	30
3.1	Le componenti chimiche della cellula . . . . .	25	3.2.10	L'ATP è il vettore attivo più utilizzato . . . . .	30
3.1.1	L'acqua è mantenuta da legami a idrogeno . . . . .	25	3.2.11	NADH e NADPH sono vettori di elettroni . . . . .	31
3.1.2	Quattro tipi di attrazione non covalente aiuta a unire le molecole nelle cellule . . . . .	25	3.2.12	Esistono molte altre molecole vettori nella cellula . . . . .	31
3.1.3	Alcune molecole polari formano acidi e basi in acqua . . . . .	26	3.2.13	La sintesi di polimeri biologici è guidata dall'idrolisi dell'ATP . . . . .	31
3.1.4	La cellula è formata da composti di carbonio . . . . .	26	3.3	Come le cellule ottengono energia dai nutrienti . . . . .	32
3.1.5	Le cellule contengono quattro principali famiglie di piccole molecole organiche . . . . .	26	3.3.1	La glicolisi è un cammino centrale per produrre ATP . . . . .	32
3.1.6	La chimica della cellula è dominata da macromolecole con proprietà notevoli . . . . .	27	3.3.2	La fermentazione produce ATP in assenza di ossigeno . . . . .	32
3.1.7	I legami non covalenti specificano la forma di una macromolecola e i suoi legami con le altre molecole . . . . .	27	3.3.3	La glicolisi mostra come gli enzimi accoppiano l'ossidazione alla conservazione di energia . . . . .	32
3.2	Catalisi e utilizzo dell'energia da parte delle cellule . . . . .	27	3.3.4	Gli organismi conservano le molecole nutrienti in serbatoi speciali . . . . .	33
3.2.1	Il metabolismo della cellula è organizzato dagli enzimi . . . . .	27	3.3.5	Zuccheri e acidi grassi sono degradati in acetil CoA nei mitocondri . . . . .	33
3.2.2	L'ordine biologico è reso possibile dal rilascio di calore dalla cellula . . . . .	28	3.3.6	Il ciclo di acido citrico genera NADH ossidando i gruppi acetili in CO <sub>2</sub> . . . . .	33
3.2.3	Le cellule ottengono energia ossidando molecole organiche . . . . .	28			
3.2.4	Ossidazione e riduzione coinvolgono il trasferimento di elettroni . . . . .	28			

3.3.7	Il trasporto di elettroni guida la sintesi della maggior parte dell'ATP nelle cellule . . . . .	34	4.1.13	Legami incrociati covalenti stabilizzano proteine extracellulari . . . . .	38
3.3.8	Amminoacidi e nucleotidi sono parte del ciclo azotato . . . . .	34	4.1.14	Le molecole proteiche spesso svolgono la funzione di subunità per l'assemblaggio di grandi strutture . . . . .	39
<b>4</b>	<b>Proteine</b>	<b>35</b>	4.1.15	I fattori di assemblaggio aiutano la formazione di complesse strutture biologiche . . . . .	39
4.1	Forma e struttura delle proteine . . . . .	35	4.1.16	Fibrille amiloide possono formarsi da molte proteine . . . . .	39
4.1.1	La forma di una proteina è specificata dalla sua sequenza degli amminoacidi . . . . .	35	4.1.17	Molte proteine contengono domini a bassa complessità che possono formare amiloidi reversibili . . . . .	40
4.1.2	Le proteine si ripiegano nella conformazione a minore energia . . . . .	36	4.2	Funzione delle proteine . . . . .	40
4.1.3	L' $\alpha$ -elica e il $\beta$ -foglietto sono pattern di piegamento comuni . . . . .	36	4.2.1	Tutte le proteine si legano ad altre molecole . . . . .	40
4.1.4	I domini proteici sono unità modulari da cui le proteine più grandi sono costruite . . . . .	36	4.2.2	La conformazione superficiale di una proteina determina la sua chimica . . . . .	40
4.1.5	Poche delle catene polipeptidiche possibili saranno utili alla cellula . . . . .	36	4.2.3	La comparazione di sottosequenze tra membri di una famiglia proteica evidenzia i siti di legame cruciali . . . . .	41
4.1.6	Le proteine possono essere classificate in molte famiglie . . . . .	37	4.2.4	Le proteine si legano con altre proteine attraverso molti tipi di interfacce . . . . .	41
4.1.7	Alcuni domini proteici si trovano in molte proteine diverse . . . . .	37	4.2.5	I siti di legame degli anticorpi sono estremamente versatili . . . . .	41
4.1.8	Il genoma umano codifica un complesso insieme di proteine, rivelando che molto rimane sconosciuto . . . . .	37	4.2.6	La costante di equilibrio misura la forza di legame . . . . .	41
4.1.9	Proteine più grandi contengono più di una catena polipeptidica . . . . .	38	4.2.7	Gli enzimi sono catalizzatori potenti e molto specifici . . . . .	42
4.1.10	Alcune proteine globulari formano lunghi filamenti elicoidali . . . . .	38	4.2.8	Il legame con il substrato è il primo passo nella catalisi enzimatica . . . . .	42
4.1.11	Molte proteine hanno forme allungate e fibrose . . . . .	38	4.2.9	Gli enzimi velocizzano la reazione stabilizzando selettivamente stati transitivi . . . . .	42
4.1.12	Le proteine contengono una grande quantità di catene polipeptidiche intrinsecamente disordinate . . . . .	38	4.2.10	Gli enzimi possono usare simultaneamente catalizzatori acidi e basici . . . . .	42

4.2.11	Il lisozima illustra come un enzima funziona . . . . .	43	4.2.24	Un elaborato sistema di coniugazione di ubiquitina è utilizzato per marcare le proteine . . . . .	46
4.2.12	Molecole strettamente legate aggiungono altre funzioni alle proteine . . . . .	43	4.2.25	I complessi proteici con parti interscambiabili fanno un efficiente uso dell'informazione genica . . . . .	47
4.2.13	Complessi multienzima aiutano ad aumentare il tasso del metabolismo cellulare . . . . .	43	4.2.26	Una proteina GTP-legante mostra come grandi movimenti proteici siano possibili . . . . .	47
4.2.14	La cellula regola le attività catalitiche dei propri enzimi	44	4.2.27	Le proteine motrici producono grandi movimenti nelle cellule . . . . .	47
4.2.15	Gli enzimi allosterici possiedono due o più siti di legame che interagiscono . .	44	4.2.28	I trasportatori di membrana imbrigliano energia per pompare le molecole attraverso le membrane . .	48
4.2.16	Due ligandi i cui siti di legame sono accoppiati devono avere effetto reciprocamente sul legame dell'altro . . . . .	44	4.2.29	Le proteine formano grandi complessi che funzionano come macchinari proteici . . . . .	48
4.2.17	Insiemi di proteine simmetrici producono transizioni allosteriche cooperative . .	45	4.2.30	Impalcature concentrano insiemi di proteine che interagiscono . . . . .	48
4.2.18	Molti cambi nelle proteine sono guidati dalla fosforilazione proteica . . .	45	4.2.31	Molte proteine sono controllate da modifiche covalenti che le dirigono verso siti specifici all'interno della cellula . . . . .	48
4.2.19	Una cellula eucariotica contiene un gran numero di proteina chinasi e proteina fosfatasi . . . . .	45	4.2.32	Una complessa rete di interazioni proteiche sottosta alle funzioni della cellula . .	48
4.2.20	La regolazione della proteina chinasi Src rivela come una proteina può funzionare come un microprocessore	45			
4.2.21	Le proteine che si legano e regolano il GTP sono onnipresenti regolatori . . .	46	<b>II</b>	<b>Meccanismi genetici basici</b>	<b>50</b>
4.2.22	Le proteine regolatorie GAP e GEF controllano l'attività delle proteine GTP-leganti determinando se è legato GTP o GDP . . . . .	46	<b>5</b>	<b>DNA, cromosomi e genomi</b>	<b>51</b>
4.2.23	Le proteine possono essere regolate dall'aggiunta covalente di altre proteine .	46	5.1	La struttura e la funzione del DNA	51
			5.1.1	Una molecola di DNA consiste di due catene di nucleotidi complementari .	51
			5.1.2	La struttura del DNA mette a disposizione un meccanismo per l'ereditarietà . .	52

5.1.3	Negli eucarioti il DNA è racchiuso nel nucleo cellulare . . . . .	52	5.3.4	La cromatina acquisisce varietà addizionale attraverso inserzioni specifiche al sito di un insieme di varianti di istoni . . . . .	56
5.2	DNA cromosomico e il suo confezionamento nella fibra cromatina . . . . .	52	5.3.5	Le modifiche covalenti e le varianti istoniche agiscono insieme per controllare le funzioni dei cromosomi . . . . .	56
5.2.1	Il DNA eucariotico è confezionato in un insieme di cromosomi . . . . .	52	5.3.6	Un complesso di proteine lettrici e scrittrici possono espandere modifiche alla cromatina lungo un cromosoma . . . . .	57
5.2.2	I cromosomi contengono lunghe stringhe di geni . . . . .	53	5.3.7	Barriere nella sequenza di DNA bloccano l'espansione dei complessi di lettura-scrittura separando domini di cromatina vicini . . . . .	57
5.2.3	La sequenza nucleotidica del genoma umano mostra come i geni sono ordinati . . . . .	53	5.3.8	La cromatina nei centromeri rivela come le varianti degli istoni possono creare strutture speciali . . . . .	57
5.2.4	Ogni molecola di DNA che forma un cromosoma lineare deve contenere un centromero, due telomeri e un'origine di replicazione . . . . .	53	5.3.9	Alcune strutture cromatiniche possono essere ereditate direttamente . . . . .	58
5.2.5	Le molecole di DNA sono altamente conservate nei cromosomi . . . . .	54	5.3.10	L'attivazione e la repressione di strutture di cromatina può essere ereditato epigeneticamente . . . . .	58
5.2.6	I nucleosomi sono unità base della struttura del cromosoma eucariote . . . . .	54	5.3.11	Le strutture di cromatina sono importanti per le funzioni del cromosoma eucariotico . . . . .	58
5.2.7	La struttura del nucleo del nucleosoma rivela come il DNA è condensato . . . . .	54	5.4	La struttura globale dei cromosomi . . . . .	58
5.2.8	I nucleosomi hanno strutture dinamiche e sono soggetti a cambi catalizzati da complessi rimodellanti a cromatina ATP-dipendenti . . . . .	55	5.4.1	I cromosomi sono piegati in grandi anelli di cromatina . . . . .	58
5.2.9	I nucleosomi sono tipicamente condensati in una fibra cromatinica . . . . .	55	5.4.2	Cromosomi politene sono utili per visualizzare le strutture di cromatina . . . . .	58
5.3	La struttura e la funzione della cromatina . . . . .	55	5.4.3	Esistono multiple forme di cromatina . . . . .	59
5.3.1	L'eterocromatina è altamente organizzata e limita l'espressione genica . . . . .	55	5.4.4	Gli anelli di cromatina decondensano quando i geni al loro interno sono espressi . . . . .	59
5.3.2	Lo stato eterocromatico è autopropagante . . . . .	56			
5.3.3	I nuclei istonici sono modificati covalentemente a molti siti diversi . . . . .	56			

5.4.5	La cromatina può muoversi a siti specifici nel nucleo per alterare l'espressione dei geni . . . . .	59	5.5.9	Cambi in sequenze precedentemente conservate possono aiutare a decifrare passi critici dell'evoluzione . . .	63
5.4.6	Reti di macromolecole formano un insieme di ambienti biochimici distinti nel nucleo . . . . .	60	5.5.10	Mutazioni nella sequenza di DNA che controlla l'espressione genica ha guidato molti dei cambi evolutivi nei vertebrati . . .	63
5.4.7	I cromosomi mitotici sono altamente condensati . . . .	60	5.5.11	La duplicazione genica fornisce una fonte di novità genica durante l'evoluzione	63
5.5	Come si evolvono i genomi . . . . .	60	5.5.12	Geni duplicati divergono . .	64
5.5.1	Il confronto tra i genomi rivela sequenze di DNA attraverso la loro evoluzione del throughput di evoluzione	61	5.5.13	L'evoluzione del gene globina mostra come la duplicazione del DNA contribuisce all'evoluzione degli organismi . . . . .	64
5.5.2	L'alterazione dei genomi è causata da fallimento del normale meccanismo per la copia e il mantenimento del DNA e da elementi del DNA trasponibili . . . . .	61	5.5.14	I geni che codificano nuove proteine possono essere creati ricombinando gli esoni . . . . .	64
5.5.3	La sequenza genomica di due specie differisce in proporzione al tempo in cui si sono evolute separatamente	61	5.5.15	Mutazioni neutrali si diffondono diventando fissate in una popolazione, con probabilità dipendente dalla dimensione della popolazione . . . . .	65
5.5.4	L'albero filogenetico costruito dal confronto delle sequenze di DNA traccia le relazioni di tutti gli organismi . . . . .	61	5.5.16	Molto può essere imparato dall'analisi della variazione tra gli umani . . . . .	65
5.5.5	Un confronto tra cromosomi umani e dei topi mostra come la struttura dei genomi diverge . . . . .	62	<b>6</b>	<b>Replicazione, riparazione e ricombinazione del DNA</b>	<b>67</b>
5.5.6	La dimensione del genoma dei vertebrati riflette i tassi relativi di addizione e perdita del DNA in un lignaggio . . . . .	62	6.1	Il mantenimento della sequenza di DNA . . . . .	67
5.5.7	Si può inferire la sequenza di alcuni genomi antichi . .	62	6.1.1	I tassi di mutazione sono estremamente bassi . . . . .	67
5.5.8	Il confronto tra sequenze multispecie identifica sequenze di DNA conservato di funzione sconosciuta . . .	63	6.1.2	I tassi di mutazione bassi sono necessari per la vita come la conosciamo . . . .	68
			6.2	Meccanismi di replicazione del DNA	68
			6.2.1	L'accoppiamento delle basi sottostà la replicazione e riparazione del DNA . . . . .	68
			6.2.2	La forcella di replicazione del DNA è asimmetrica . .	68

6.2.3	L'alta fedeltà del meccanismo di replicazione del DNA richiede molti meccanismi di correzione . . . . .	69	6.3.4	Negli eucarioti la replicazione del DNA avviene durante un'unica parte del ciclo vitale della cellula . . .	73
6.2.4	Solo la replicazione nella direzione 5'-3' permette una correzione efficiente . .	69	6.3.5	Diverse regioni dello stesso cromosoma si replicano a tempi distinti durante la fase S . . . . .	73
6.2.5	Un enzima che polimerizza i nucleotidi sintetizza corti primer di RNA sul filamento in ritardo . . . . .	70	6.3.6	Un grande complesso multisubunità si lega alle origini di replicazione eucariotiche . . . . .	73
6.2.6	Proteine speciali aiutano l'apertura della doppia elica sopra la forcella di replicazione . . . . .	70	6.3.7	Nuovi nucleosomi sono assemblati in coda alla forcella di replicazione . . . .	74
6.2.7	Un anello che scivola mantiene una DNA polimerasi in movimento sul DNA . .	70	6.3.8	I telomeri sono condensati in strutture specializzate che proteggono la fine dei cromosomi . . . . .	74
6.2.8	Le proteine a una forcella di replicazione cooperano per formare una macchina di replicazione . . . . .	71	6.3.9	La lunghezza dei telomeri è regolata dalla cellula e dagli organismi . . . . .	75
6.2.9	Un sistema di riparazione direzionato al filamento rimuove gli errori di replicazione che sfuggono alla macchina di replicazione . .	71	6.4	Riparazione del DNA . . . . .	75
6.2.10	La DNA topoisomerasi impedisce l'ingarbugliamento del DNA durante la replicazione . . . . .	71	6.4.1	Senza la riparazione del DNA, danni spontanei cambierebbero rapidamente la sequenza del DNA . . . . .	75
6.2.11	La replicazione del DNA è fondamentalmente simile in eucarioti e batteri . . . . .	72	6.4.2	La doppia elica del DNA è prontamente riparata . . .	75
6.3	Inizializzazione e completamento della replicazione del DNA nei cromosomi . . . . .	72	6.4.3	Danni al DNA possono essere rimossi attraverso molti cammini . . . . .	75
6.3.1	La sintesi del DNA inizia alle origini di replicazione .	72	6.4.4	L'accoppiamento di riparazione a asportazione di nucleotidi con la trascrizione garantisce che il DNA più importante è riparato efficientemente . . . . .	76
6.3.2	I cromosomi batterici hanno un'origine singola per la replicazione . . . . .	72	6.4.5	La chimica delle basi del DNA facilita l'individuazione dei danni . . . . .	76
6.3.3	I cromosomi eucariotici contengono multiple origini di replicazione . . . . .	73	6.4.6	Speciali translesioni della DNA polimerasi sono usate durante emergenze . .	76
			6.4.7	Rotture nel doppio filamento sono efficientemente riparate . . . . .	77



6.4.8	Danno al DNA ritarda la progressione del ciclo cellulare . . . . .	77	6.5.3	Alcuni virus usano un meccanismo di trasposizione per muoversi nel cromosoma della cellula ospite . . .	81
6.4.9	Riparazione omologa . . . .	77	6.5.4	Retrotrasposoni simili ai retrovirus assomigliano ai retrovirus ma non possiedono un capsido . . . . .	81
6.4.10	L'accoppiamento delle basi del DNA guida la ricombinazione omologa . . . . .	77	6.5.5	Una grande frazione del genoma umano è composta di retrotrasposoni nonvirali	81
6.4.11	La ricombinazione omologa può riparare perfettamente rotture a doppio filamento nel DNA . . . . .	78	6.5.6	La ricombinazione conservativa specifica al sito può riordinare il DNA reversibilmente . . . . .	82
6.4.12	Il cambio di filamenti avviene grazie la proteina RecA/Rad51 . . . . .	78	<b>7</b>	<b>Come una cellula legge il genoma, dal DNA alle proteine</b>	<b>83</b>
6.4.13	La ricombinazione omologa può salvare forcelle di replicazione danneggiate . .	78	7.1	Dal DNA all'RNA . . . . .	83
6.4.14	Le cellule regolano l'uso della ricombinazione omologa durante la riparazione del DNA . . . . .	78	7.1.1	Le molecole di RNA hanno un unico filamento . . . . .	83
6.4.15	La ricombinazione omologa è cruciale per la meiosi . . . . .	79	7.1.2	La trascrizione produce RNA complementare a un filamento di DNA . . . . .	84
6.4.16	La ricombinazione meiotica comincia con una rottura a doppio filamento programmata . . . . .	79	7.1.3	L'RNA polimerasi causa la trascrizione . . . . .	84
6.4.17	Le giunzioni di Holliday sono formate durante la meiosi . . . . .	79	7.1.4	Le cellule producono diverse categorie di molecole di RNA . . . . .	84
6.4.18	La ricombinazione omologa produce crossover e non-crossover durante la meiosi . . . . .	79	7.1.5	Segnali codificati nel DNA indicano l'RNA polimerasi dove iniziare e dove finire .	84
6.4.19	La ricombinazione omologa solitamente causa conversione genica . . . . .	80	7.1.6	I segnali di inizio e terminazione della trascrizione sono eterogenei nella sequenza nucleotidica . . .	85
6.5	Trasposizione e ricombinazione specifica al sito conservativa . . . .	80	7.1.7	L'iniziazione della trascrizione negli eucarioti richiede molte proteine . . . . .	85
6.5.1	Attraverso la transizione gli elementi genetici mobili possono inserirsi in ogni sequenza di DNA . . . . .	80	7.1.8	La RNA polimerasi II richiede un insieme di fattori di trascrizione generali . . .	86
6.5.2	Trasposoni solo a DNA si possono muovere attraverso un meccanismo a copia-incolla . . . . .	81	7.1.9	La polimerasi II richiede attivatori, mediatori e proteine per la modifica della cromatina . . . . .	86

7.1.10	L'allungamento della trascrizione negli eucarioti richiede proteine accessorie .	86	7.1.26	Il nucleo contiene una varietà di aggregati subnucleari . . . . .	91
7.1.11	La trascrizione crea tensione superelocale . . . . .	87	7.2	Da RNA a proteine . . . . .	91
7.1.12	L'allungamento della trascrizione è strettamente accoppiato con il processamento dell'RNA . . . . .	87	7.2.1	Una sequenza di mRNA è decodificata in insiemi di tre nucleotidi . . . . .	91
7.1.13	L'incappucciamento dell'RNA è la prima modifica dei pre-mRNA eucariotici .	87	7.2.2	Molecole di tRNA combinano amminoacidi ai codoni	92
7.1.14	L'RNA splicing rimuove le sequenze di introni dal pre-mRNA . . . . .	87	7.2.3	tRNA sono modificati covalentemente prima che escano dal nucleo . . . . .	92
7.1.15	Sequenze di nucleotidi segnalano dove avviene lo splicing . . . . .	87	7.2.4	Specifici enzimi accoppiano amminoacidi con la molecola di tRNA appropriata	92
7.1.16	Lo splicing dell'RNA è svolto dallo spliceosoma .	88	7.2.5	La modifica da tRNA sintetasi assicura accuratezza .	92
7.1.17	Lo spliceosoma usa l'idrolisi dell'ATP per produrre una serie di riordinamenti RNA-RNA . . . . .	88	7.2.6	Gli amminoacidi sono aggiunti alla terminazione C <sup>-</sup> di una catena polipeptidica crescente . . . . .	93
7.1.18	Altre proprietà del pre-mRNA e la sua sintesi spiega le scelte dei siti propri di splice . . . . .	88	7.2.7	Il messaggio di RNA è decodificato nei ribosomi .	93
7.1.19	La struttura cromatinica influenza l'RNA splicing .	89	7.2.8	Fattori di allungamento portano avanti la traduzione e ne aumentano l'accuratezza . . . . .	93
7.1.20	Lo splicing di RNA mostra plasticità notevole . . . . .	89	7.2.9	Molti processi biologici superano le limitazioni intrinseche all'accoppiamento di basi complementari .	94
7.1.21	RNA splicing catalizzato dallo spliceosoma si è probabilmente evoluto da meccanismi di autosplicing	89	7.2.10	L'accuratezza nella traduzione richiede una spesa di energia libera . . . . .	94
7.1.22	Enzimi di processamento dell'RNA generano la terminazione 3' degli mRNA eucariotici . . . . .	89	7.2.11	Il ribosoma è un ribozima .	94
7.1.23	mRNA maturi eucariotici sono esportati selettivamente dal nucleo . . . . .	90	7.2.12	La sequenza nucleotidica negli mRNA segnala dove iniziare la sintesi delle proteine . . . . .	95
7.1.24	Anche gli RNA non codificanti sono sintetizzati e processati nel nucleo . . . .	90	7.2.13	I codoni di fine marcano la fine della traduzione . . . .	95
7.1.25	Il nucleolo è una fabbrica di produzione di ribosomi .	91	7.2.14	Le proteine sono create su poliribosomi . . . . .	96
			7.2.15	Ci sono piccole variazioni nel codice genetico standard	96

7.2.16	Meccanismi di controllo della qualità agiscono per prevenire la traduzione di mRNA danneggiati . . . . .	96	8.2.3	La dimerizzazione dei regolatori di transizione aumenta la loro affinità e specificità per il DNA . . . . .	100
7.2.17	Alcune proteine cominciano a piegarsi mentre vengono sintetizzate . . . . .	96	8.2.4	I regolatori di trascrizione si legano cooperativamente al DNA . . . . .	101
7.2.18	Accompagnatori molecolari aiutano a guidare il piegamento della maggior parte delle proteine . . . . .	97	8.2.5	La struttura del nucleosoma promuove legami cooperativi di regolatori di trascrizione . . . . .	101
7.2.19	La cellula utilizza diversi tipi di accompagnatori . . . . .	97	8.3	I regolatori di trascrizione attivano e disattivano i geni . . . . .	101
7.2.20	Regioni idrofobiche esposte forniscono segnali critici per il controllo di qualità delle proteine . . . . .	97	8.3.1	Il repressore triptofano disattiva i geni . . . . .	101
7.2.21	Il proteosoma è una proteasi compartimentalizzata con siti attivi reclusi . . . . .	98	8.3.2	I repressori disattivano i geni e gli attivatori li attivano . . . . .	102
7.2.22	Molte proteine sono controllate da distruzione regolata . . . . .	98	8.3.3	Un attivatore e un repressore controllano l'operone Lac . . . . .	102
<b>8</b>	<b>Controllo dell'espressione genica</b>	<b>99</b>	8.3.4	L'innestamento del DNA può avvenire durante la regolazione genica dei batteri . . . . .	102
8.1	Una panoramica del controllo dei geni . . . . .	99	8.3.5	Interruttori complessi controllano la trascrizione genica negli eucarioti . . . . .	102
8.1.1	Diversi tipi di cellule sintetizzano diversi insiemi di RNA e proteine . . . . .	99	8.3.6	Una regione di controllo di un gene eucariotico consiste di un promotore e molte sequenze cis-regolatorie . . . . .	103
8.1.2	Segnali esterni possono causare il cambio dell'espressione dei geni di una cellula . . . . .	99	8.3.7	I regolatori di trascrizione eucariotici lavorano in gruppi . . . . .	103
8.1.3	L'espressione dei geni può essere regolata a molti dei passaggi nel cammino da DNA a RNA a proteine . . . . .	99	8.3.8	Le proteine attivatrici promuovono l'assemblaggio di RNA polimerasi al punto di inizio della trascrizione . . . . .	103
8.2	Controllo della trascrizione da parte di proteine leganti a specifiche sequenze . . . . .	100	8.3.9	Gli attivatori di trascrizione eucariotici direazionalano la modifica di strutture cromatiniche locali . . . . .	104
8.2.1	La sequenza dei nucleotidi nella doppia elica di DNA può essere letta da proteine . . . . .	100	8.3.10	Gli attivatori di trascrizione possono promuoverla rilasciando la RNA polimerasi dai promotori . . . . .	104
8.2.2	I regolatori di trascrizione contengono motivi strutturali che possono leggere sequenze di DNA . . . . .	100			

8.3.11	Gli attivatori di trascrizione lavorano sinergicamente . . . . .	104	8.4.3	L'imprinting genomico si basa sulla metilazione del DNA . . . . .	108
8.3.12	Repressori di trascrizione eucariotici possono inibire la trascrizione in molti modi	105	8.4.4	Alterazioni globali al cromosoma nella struttura cromatinica possono essere ereditate . . . . .	109
8.3.13	Sequenze di DNA isolanti impediscono a regolatori di influenzare geni distanti . .	105	8.4.5	Meccanismi epigenetici assicurano che pattern stabili di espressione genica possono essere trasmessi a cellule figlie . . . . .	109
8.3.14	Meccanismi genetici molecolari che creano e mantengono tipi di cellula specializzati . . . . .	105	8.5	Controlli post trascrizionali . . . . .	109
8.3.15	Complessi interruttori genici che regolano lo sviluppo della Drosophila sono creati da molecole più piccole . . . . .	105	8.5.1	L'attenuazione della trascrizione causa la terminazione prematura di alcune molecole di RNA . . . . .	109
8.3.16	Il gene Eve della Drosophila è regolato da controlli combinatori . . . . .	106	8.5.2	Riboswitches rappresentano probabilmente antiche forme di controllo genico . .	110
8.3.17	I regolatori di trascrizione sono messi ingioco da segnali extracellulari . . . .	106	8.5.3	RNA splicing alternativo può produrre diverse forme di proteine dallo stesso gene	110
8.3.18	Il controllo dei geni combinatorio diversi tipi di cellule . . . . .	106	8.5.4	La definizione di un gene è stata modificata . . . . .	110
8.3.19	Combinazioni di regolatori di trascrizione master specificano il tempo della cellula controllando l'espressione di molti geni .	106	8.5.5	Un cambio nel sito della rottura del trascritto a RNA e dell'aggiunta di Poli-A può cambiare la terminazione C di una proteina . . . . .	110
8.3.20	Cellule specializzate devono rapidamente attivare e disattivare insiemi di geni .	106	8.5.6	Modifiche dell'RNA possono cambiare il significato del messaggio RNA . . . . .	111
8.3.21	Cellule differenziate mantengono la loro identità . .	107	8.5.7	Il trasporto di RNA dal nucleo può essere regolato .	111
8.3.22	Circuiti di trascrizione permettono alla cellula di compiere operazioni logiche	107	8.5.8	Alcuni mRNA sono localizzati in regioni specifiche del citosol . . . . .	111
8.4	Meccanismi che rinforzano la memoria cellulare in piante e animali	107	8.5.9	Le regioni non tradotte 5' e 3' controllano la traduzione dei propri mRNA	112
8.4.1	Pattern di metilazione del DNA possono essere ereditati quando le cellule dei vertebrati si dividono . . .	107	8.5.10	La fosforilazione di un fattore di iniziazione regola globalmente la sintesi proteica . . . . .	112
8.4.2	Isole ricche di CG sono associate con molti geni nei mammiferi . . . . .	108			

## Parte I

# Introduzione alla cellula

# Capitolo 1

## Introduzione

### 1.1 Microscopia

I microscopi possono essere divisi in due categorie principali: i microscopi ottici (composti, semplici o a fluorescenza) e i microscopi elettronici. Mentre i primi utilizzano la luce e lenti ingrandenti i secondi utilizzano fasci di elettroni direzionati attraverso campi magnetici. Diverse tecniche di microscopia elettronica includono trascrizione, sezione, freeze fracture e scansione. I microscopi elettronici generano immagini unicamente in bianco e nero e si rende pertanto necessario utilizzare dei falsi colori in un secondo momento, ma hanno come vantaggio una maggiore risoluzione delle immagini. Per un microscopio ottico l'ingrandimento dell'immagine è dato dall'ingrandimento dell'obiettivo moltiplicato per l'ingrandimento degli eye-pieces:

$$M_{microscope} = M_{objective} \cdot M_{eyepieces}$$

#### 1.1.1 Risoluzione

La risoluzione di un immagine è una misura del dettaglio che essa contiene e se attraverso tecniche digitali l'ingrandimento è illimitato la risoluzione no. Si indica con risoluzione la minor distanza tra due punti che possono essere distinti come separati. La risoluzione dipende da parametri dell'utente finale e da parametri fisici.

##### Parametri fisici

I parametri fisici che determinano la risoluzione sono:

- Il corretto allineamento del sistema ottico del microscopio.
- La lunghezza d'onda della luce ( $\lambda$ ): maggiore la lunghezza d'onda minore la risoluzione.
- L'apertura numerica (NA) dell'obiettivo e del condensatore. Questo parametro indica la capacità di un obiettivo di raccogliere luce e risolvere dettagli ad una distanza fissata dall'oggetto. Dipende dall'ingrandimento e dall'indice di rifrazione del medium tra microscopio e oggetto (aria, acqua, olio). Maggiore l'indice di rifrazione maggiore il numero di apertura e maggiore la risoluzione.

### 1.1.2 Tecniche di microscopia ottica

#### Microscopia a cambio di fase

Nella microscopia a cambio di fase viene sfruttato lo shift di fase della luce quando attraversa il corpo che si vuole osservare. L'interpretazione dello shift dà origine a diverse possibili considerazioni sull'oggetto.

#### Microscopia a fluorescenza

La microscopia a fluorescenza sfrutta il fenomeno della fluorescenza: certe molecole, quando colpite da certe lunghezze d'onda, si eccitano. Successivamente quando la molecola ritorna dallo stato eccitato a quello basale emette a sua volta un'onda luminosa di lunghezza d'onda maggiore di quella con cui era stata colpita (legge di Stokes). Per questa tecnica vengono tipicamente utilizzate delle proteine come GFP (green fluorescent protein), YFP, CFP tipicamente estratte da organismi (nel caso della GFP da una medusa) che permettono pertanto la loro codifica nel DNA della cellula. Queste proteine vengono aggiunte ad una proteina in modo da riuscire ad osservarne il comportamento. Si deve prestare attenzione al fatto che per massa o struttura questa aggiunta potrebbe creare una variazione in comportamento della proteina oggetto di studio. I fattori essenziali per la microscopia a fluorescenza sono:

- Eccitazione di alta intensità.
- Filtri di eccitazione e emissione appropriati.
- Autofluorescenza minima nell'oggetto di studio.
- Utilizzo di un olio di immersione non fluorescente.
- Antifade reagents.

Questa tecnica permette pertanto di osservare proteine in cellule vive a differenza dell'antibody staining. Un'importante applicazione è l'immunoistochimica.

**FRET** La tecnica FRET si utilizza per determinare la prossimità di due diverse proteine: si attaccano ad esse due molecole fluorescenti tali che l'emissione della prima eccita la seconda, che a sua volta emette luce. Se questo accade si è dimostrato che le due proteine sono vicine.

**Photobleach** Questa tecnica viene utilizzata per mostrare la velocità di movimento di una proteina della cellula: si trova una cellula di controllo e quella oggetto di studio entrambe con la proteina fluorescente. La seconda viene sottoposta ad un raggio ad alta potenza (photoactivation) che distrugge la parte fluorescente della proteina. Si osserva ora la velocità con cui la luminescenza torna nella zona colpita, confrontandola con la cellula di controllo e si crea il grafo di velocità di movimento della proteina.

## 1.2 Storia

Mendel è considerato il padre della genetica moderna. Fu un monaco benedettino che visse nell'800 che introdusse la quantificazione in biologia. I suoi esperimenti riguardanti piante di piselli hanno osservato che prendendo due linee pure con un fenotipo diverso e incrociandole nella prima generazione si ottenevano piante con un solo fenotipo. Successivamente incrociando tra di loro piante della

prima generazione osservò una ricomparsa dell'altro fenotipo in rapporto di 3 a 1. Mendel associò questa scoperta alla presenza di due alleli, uno dominante e uno recessivo che determinavano il fenotipo. Successivamente determinò anche che due caratteri vengono trasmessi in modo indipendente. Nel 1871 viene scoperto il DNA grazie al lavoro di Miescher che chiama nucleina in quanto si trova nel nucleo della cellula. Il suo lavoro riguardava l'estrazione delle cellule da bende impregnate di pus, ricche pertanto di materiale cellulare. Dall'estrazione della nucleina, ricca di fosforo e acida, notò che la sua concentrazione era lineare con il numero di cellule e le cellule che si dividevano molto ne possedevano di più.

Dal 1909 gli esperimenti di Morgan gli valsero il premio Nobel per le sue scoperte riguardanti il ruolo dei cromosomi nell'ereditarietà. Riconobbe infatti che la sede dell'informazione genetica sono i cromosomi, isolati dalla *Drosophila*, utilizzata a causa dei suoi grandi cromosomi. Riuscì pertanto a confermare che i geni sono depositati nei cromosomi nel nucleo della cellula, che sono organizzati in una lunga riga nei cromosomi, che tratti dipendenti l'uno dall'altro corrispondono a geni che sono in siti vicini sui cromosomi e scoprì il fenomeno del crossover.

Nel 1928 Griffith studiò lo *Streptococcus Pneumoniae*, un patogeno presente in due varianti con diversa patogenicità: Rough (R) e Smooth (S). Se la variante S viene iniettata in un topo questo muore, mentre con R non muore. Dopo aver ucciso il patogeno di tipo S e averlo iniettato in un topo sano, questo non si ammalava, ma unendo a S morto un agente R vivo il topo, iniettato con la combinazione, moriva. Viene pertanto ipotizzato il passaggio orizzontale di materiale genetico.

Nel 1943 Avery conferma che il DNA è la molecola responsabile del trasferimento genetico orizzontale utilizzando lo stesso agente di Griffith riuscì a separare in coltura l'estratto cellulare nelle varie componenti molecolari: proteine, DNA, lipidi e polisaccaridi. Trattando le cavi con queste molecole separatamente scoprì che il responsabile della trasmissione genica orizzontale era il DNA.

Nel 1952 Hershey e Chase marcarono DNA con  $P^{32}$  e le proteine con  $S^{35}$  dei virus batteriofagi (che replicano il proprio DNA attraverso batteri fino a farli esplodere). Durante l'infezione di questi virus sull'*Escherichia coli*, separati tramite frullazione e con il loro materiale estratto attraverso lisi si nota come la maggior parte conteneva  $P^{32}$ , determinando il DNA come la molecola responsabile dell'infezione.

Nel 1953 Watson e Crick riuscirono a determinare la struttura del DNA e a formulare il dogma centrale della biologia molecolare: Il passaggio da DNA a mRNA può avvenire anche in senso contrario



**Figura 1.1:** Il dogma centrale della biologia molecolare

nel caso dei retrovirus.

Nel 1996 Mello e Fire evidenziarono l'importanza dell'RNA e del microRNA, vincendo il premio Nobel per la scoperta dell'interferenza dell'RNA e la proprietà dell'RNA a doppio filamento di interferire e spegnere l'espressione genica. Il microRNA è un RNA con 20-22 nucleotidi a singolo filamento, scoperto prima nelle piante che lo usano per difendersi dalle infezioni virali. L'RNA a doppio filamento fa in modo che il gene codificante la proteina corrispondente venga silenziato in modo che non esprima la proteina.



## Capitolo 2

# Cellule e genomi

Tutti gli esseri viventi sono fatti di cellule e queste unità di materia vivente hanno in comune lo stesso macchinario che svolge le stesse funzioni di base. Si nota pertanto il contrappunto tra enorme differenza tra gli individui se osservati all'esterno ma una straordinaria somiglianza nei meccanismi fondamentali.

### 2.1 Caratteristiche universali della vita sulla Terra

Ciascuna specie presente sulla Terra è diversa e si riproduce fedelmente producendo una progenie che appartiene alla stessa specie. L'organismo genitore passa l'informazione che specifica in modo dettagliato le caratteristiche che la progenie avrà. Questo fenomeno, detto ereditarietà è ciò che distingue la vita da altri processi. L'organismo vivente inoltre consuma energia libera per creare e mantenere la sua organizzazione che spinge un sistema complesso di processi chimici nel modo specificato dall'informazione ereditaria. Sia che l'organismo sia costituito da una singola cellula o da raggruppamenti di cellule specializzate collegati da sistemi complessi di comunicazione è stato generato da divisioni cellulari di una singola cellula. La singola cellula è perciò il veicolo dell'informazione ereditaria che definisce la specie e include il meccanismo necessario alla sua copia.

#### 2.1.1 Tutte le cellule conservano la loro informazione ereditaria nello stesso codice chimico lineare (DNA)

Tutte le cellule viventi sulla Terra conservano le loro informazioni sotto forma di molecole a doppio filamento di DNA, lunghe catene polimeriche accoppiate senza ramificazioni, formate sempre dagli stessi quattro tipi di monomeri, con nomi derivati da un alfabeto a quattro lettere (A, C, G, T). Questi monomeri sono attaccati in una lunga sequenza lineare che codifica l'informazione genica. Essendo il DNA una struttura utilizzata da tutte le cellule viventi il DNA di un essere umano è leggibile, copiabile e interpretabile da una cellula batterica (e viceversa). Utilizzando metodi chimici si può leggere la sequenza completa di monomeri in qualunque molecola di DNA e decifrare così l'informazione ereditaria contenuta in qualsiasi organismo.

### 2.1.2 Tutte le cellule replicano la loro informazione ereditaria mediante polimerizzazione su uno stampo

I meccanismi che rendono possibile la vita dipendono dalla struttura della molecola di DNA a doppio filamento. Ciascun monomero nel DNA, detto nucleotide, è composto da uno zucchero (deossiribosio) con un gruppo fosfato attaccato e una base che può essere adenina (A), guanina (G), citosina (C) o timina (T). I nucleotidi creano una catena polimerica composta da un'ossatura ripetitiva zucchero-fosfato con una serie di basi che sporgono da un lato. Essendo l'unità zucchero-fosfato asimmetrica la catena ha una polarità che determina l'ordine di lettura. Il polimero di DNA viene esteso aggiungendo monomeri ad una estremità. Essendo la base uguale per tutti in teoria qualsiasi base può essere aggiunta in qualsiasi momento. Essendo che nella cellula vivente il DNA viene sintetizzato su uno stampo formato da un filamento preesistente di DNA le basi che sporgono dal filamento esistente si legano a basi del filamento che viene sintetizzato, secondo una regola rigida definita da strutture complementari delle basi: A si lega a T e C a G. Questi accoppiamenti di basi mantengono i nuovi monomeri in posizione e controllano la scelta del nuovo monomero da aggiungere. In questo modo si crea una struttura a doppio filamento composta da due catene complementari di nucleotidi e un'ossatura con polarità inversa. I nucleotidi di ciascun filamento si uniscono tra di loro attraverso legami covalenti, mentre con i corrispettivi nell'altro con legami ad idrogeno. I due filamenti si avvolgono l'uno sull'altro formando la struttura a doppia elica. Essendo i legami tra le basi deboli i due filamenti possono separarsi in modo da fornire lo stampo per una nuova replicazione. Questo processo di replicazione del DNA avviene con ritmi, controlli e molecole ausiliarie diverse, ma le basi sono universali: il DNA è il depositario dell'informazione e la polimerizzazione a stampo è il modo con cui l'informazione viene copiata e propagata.

### 2.1.3 Tutte le cellule trascrivono porzioni della loro informazione ereditaria nella stessa forma intermedia (RNA)

Si rende necessario esprimere le informazioni del DNA in modo da guidare la sintesi di altre molecole nella cellula. Anche questo processo è universale e produce principalmente RNA e proteine. Il processo inizia con una polimerizzazione su stampo detta trascrizione, in cui segmenti del DNA sono usati come stampo per la sintesi di molecole più corte di acido-ribonucleico o RNA. In seguito durante la traduzione queste molecole dirigono la sintesi di proteine. Nell'RNA l'ossatura è formata da ribosio e la timina viene sostituita dall'uracile (U). Durante la trascrizione i monomeri dell'RNA sono allineati e scelti per la polimerizzazione su un filamento stampo di DNA. Il risultato è un polimero che rappresenta una parte dell'informazione genetica della cellula. Lo stesso segmento di DNA può essere utilizzato per la sintesi di molti trascritti identici di RNA. Si noti pertanto come se il DNA rimane unico e stabile per la cellula l'RNA è monouso e prodotto in massa. Questi trascritti sono intermedi nel trasferimento dell'informazione genetica: servono soprattutto da RNA messaggero (mRNA) che guida la sintesi di proteine secondo le istruzioni conservate nel DNA. Le molecole di RNA possiedono anche strutture caratteristiche che possono conferire capacità chimiche specializzate. Essendo a filamento singolo possono ripiegarsi all'indietro su se stesse per formare legami deboli tra le basi, situazione che avviene quando segmenti della sequenza sono localmente complementari. La forma viene pertanto dettata dalla sequenza e può permettere alla molecola di essere scelta selettivamente e di catalizzare modificazioni chimiche cruciali per alcuni dei processi più antichi e fondamentali della cellula.

### 2.1.4 Tutte le cellule usano proteine come catalizzatori

Anche le proteine sono lunghe catene polimeriche non ramificate formate dall'unione in serie di monomeri comuni a tutti gli esseri viventi. Portano un'informazione sotto forma di una sequenza lineare di simboli. Dopo l'acqua sono l'elemento più presente nella cellula. I monomeri sono detti amminoacidi e ne esistono 20, sono tutti formati dalla stessa struttura centrale standard che permette la formazione di catene a cui è attaccato un gruppo laterale che determina il carattere chimico specifico. Ciascuna molecola proteica o polipeptide si ripiega in una forma tridimensionale precisa con siti reattivi sulla sua superficie. Questi polimeri di amminoacidi si legano con alta specificità ad altre molecole e agiscono da enzimi che catalizzano reazioni in cui vengono rotti o creati legami covalenti e dirigono pertanto la maggioranza dei processi chimici. Hanno anche funzione strutturale, motile, di rilevazione segnali, che viene determinata in base alla sequenza di amminoacidi creata dalla sequenza genica. Il circuito a feedback di catalizzazione del processo di duplicazione del DNA da parte delle proteine, che viene poi utilizzato per produrre proteine e RNA è alla base del comportamento autocatalitico e capace di autoriprodursi degli organismi viventi.

### 2.1.5 Tutte le cellule traducono RNA nello stesso modo

La traduzione dell'informazione genica è un processo complesso. L'informazione contenuta in un mRNA è letta in gruppo di tre nucleotidi alla volta: ciascuna tripletta di nucleotidi (codone) codifica un singolo amminoacido. Questa codifica porta a ridondanza in quanto ci sono 64 codoni ma 20 amminoacidi. Il codice è letto dal tRNA (RNA transfer). Ogni tipo di tRNA possiede ad un'estremità un amminoacido specifico e all'altra estremità una sequenza di tre nucleotidi (anticodone) che gli permette tramite l'accoppiamento di basi di riconoscere un gruppo di codoni nell'mRNA. Per la sintesi proteica una successione di molecole di tRNA cariche degli amminoacidi deve legarsi all'mRNA e gli amminoacidi devono essere uniti per espandere la catena proteica e i tRNA liberati dal loro carico devono essere rilasciati. Questo insieme di processi viene eseguito dal ribosoma formato da due catene principali di RNA detto rRNA (RNA ribosomiale) e da un gran numero di proteine diverse. Questa struttura si attacca ad un'estremità dell'mRNA e si sposta lungo di essa catturando molecole di tRNA cariche e mettendo insieme gli amminoacidi in modo da formare una nuova catena proteica.

### 2.1.6 Il frammento di informazione genica che corrisponde ad una proteina è un gene

Le molecole di DNA contengono le specifiche per migliaia di proteine. Sequenze speciali nel DNA servono come punteggiatura, indicando dove l'informazione di ciascuna proteina inizia e finisce. Ogni segmento del DNA è trascritto in una molecola di mRNA, codifica di diverse proteine. Tali segmenti sono i geni. Le molecole di RNA trascritte possono essere processate in modi diversi in modo da generare un insieme di versioni alternative di una proteina. Oltre a questo alcune parti sono trascritte in RNA con funzioni catalitiche, strutturali o regolatorie. Un gene viene pertanto definito come il segmento della sequenza di DNA corrispondente a una singola proteina o un insieme di proteine varianti o a una singola molecola di RNA catalitica, strutturale o regolatoria. L'espressione dei geni è regolata in base alla necessità della proteina o RNA che producono. Questo avviene grazie a lunghezze di DNA regolatorio che intramezzano la sequenza che legano delle proteine che controllano il tasso della trascrizione. In questo modo il genoma della cellula determina la natura delle proteine e quando devono essere prodotte.

### 2.1.7 La vita richiede energia libera

Una cellula vivente è un sistema chimico dinamico che opera lontano dall'equilibrio chimico. Affinchè una cellula cresca deve prendere dall'ambiente energia e materiali per far avvenire le reazioni. Questo consumo di energia è ciò che tiene la cellula in vita. L'energia è anche fondamentale per la propagazione dell'informazione genetica. Il processo che guida la formazione dei legami che determinano le molecole all'interno della cellula richiede energia per formare legami abbastanza forti da resistere a pressioni esterne.

### 2.1.8 Tutte le cellule funzionano come fabbriche biochimiche che impiegano le stesse strutture molecolari base

Siccome tutte le cellule creano DNA, RNA e proteine, devono contenere e manipolare una collezione simile di zuccheri semplici, nucleotidi e amminoacidi e altre sostanze come l'ATP (adenina trifosfato) per la sintesi di DNA e RNA e come trasportatore di energia libera.

### 2.1.9 Tutte le cellule sono chiuse in una membrana plasmatica attraverso la quale nutrienti e materiali di scarto devono passare

Un'altra caratteristica universale è che ogni cellula è rinchiusa da una membrana citoplasmatica che agisce da barriera selettiva e permette alla cellula di concentrare i nutrienti e di conservare i prodotti delle sintesi, escrendo i materiali di scarto. Le molecole che formano la membrana sono anfipatiche, ovvero formate da una parte idrofobica e una idrofila. Queste molecole poste in acqua si aggregano spontaneamente in modo da avvicinare tra di loro (e allontanare dall'acqua) la parte idrofobica. Tali tipi di molecole, come i fosfolipidi si aggregano in acqua per formare vescicoli a bistrato. Tipicamente la coda idrofobica è costituita da polimeri di idrocarburi e dimostra perfettamente la tendenza della cellula di formare molecole le cui proprietà chimiche causano un auto-assemblamento nella struttura necessaria alla cellula. Naturalmente il confine della cellula deve poter permettere il passaggio di alcuni elementi e sono pertanto presenti proteine di trasporto di membrana il cui compito di trasportare le molecole che entrano ed escono dalla cellula.

## 2.2 La diversità dei genomi e l'albero della vita

Il successo degli organismi viventi li ha portati ad occupare qualsiasi luogo sulla Terra. La maggior parte di questi organismi rimane però microscopica e invisibile all'occhio nudo.

### 2.2.1 Le cellule possono ottenere energia da una varietà di fonti di energia libera

Alcuni organismi come animali, funghi e molti batteri ottengono energia cibandosi di altri organismi viventi o dei prodotti chimici organici che producono, chiamati organotrofi. Altri la ottengono dal mondo non vivente e si dividono in due categorie: quelli che raccolgono l'energia dalla luce solare (fototrofi) e quelli che la ottengono da sistemi inorganici ricchi di essa (litotrofi). Gli organismi organotrofi non potrebbero esistere senza questi due. Gli organismi fototrofi includono batteri, alghe e piante, i litotrofi sono microscopici e vivono in condizioni impossibili per l'uomo. Alcuni litotrofi ottengono energia da reazioni aerobiche, altri da reazioni anaerobiche.

### 2.2.2 Alcune cellule fissano azoto e anidride carbonica per altre

La maggior parte della materia che compone le proteine è composta di idrogeno, carbonio, azoto, ossigeno, zolfo e fosforo, molto presenti in ambienti non viventi, ma non in una forma chimica facilmente incorporata in molecole biologiche. Una gran quantità di energia è necessaria per guidare le reazioni che usano le molecole inorganiche di  $N_2$  e  $CO_2$  per fissare  $N$  e  $C$  e renderli disponibili agli organismi. Pertanto classi di cellule sono specializzate per fare questo lavoro. Si nota pertanto come le cellule viventi possano variare infinitamente in aspetti della loro biochimica.

### 2.2.3 La più grande diversità biochimica esiste tra le cellule procariote

Gli organismi viventi, in base alla loro struttura cellulare possono essere divisi in eucarioti e procarioti. Gli eucarioti mantengono il proprio DNA all'interno di una membrana intracellulare chiamata il nucleo, mentre i procarioti non possiedono questo compartimento. La maggior parte delle cellule procariote sono piccole e vivono principalmente come individui indipendenti o in comunità vagamente organizzate. Possiedono tipicamente un rivestimento protettivo spesso detto parete cellulare sotto il quale si trova una membrana plasmatica che contiene un singolo compartimento citoplasmatico che contiene tutte le molecole necessarie per la vita. Vivono in una varietà di nicchie ecologiche e sono molto vari nelle loro capacità biochimiche.

### 2.2.4 L'albero della vita possiede tre rami principali: Batteri, Archei ed Eucarioti

La classificazione delle cose viventi dipende da somiglianze comuni, che, come mostrato da Darwin, suggerisce un antenato comune relativamente recente. I procarioti sono classificati in termini della loro biochimica e necessità nutrizionali. L'analisi genomica fornisce un modo più semplice e diretto di determinazione di relazioni evolutive. La sequenza di DNA di un organismo definisce la sua natura in maniera esaustiva e permette una facile comparazione con le informazioni corrispondenti in altre forme viventi e pertanto una determinazione immediata della loro distanza evolutiva. Questo ha permesso la distinzione dei procarioti in batteri e archei e come la prima cellula eucariota si sia generata da una cellula archea penetrata in un batterio antico.

### 2.2.5 Alcuni geni evolvono rapidamente, altri sono altamente conservati

Sia durante la conservazione che la copiatura dell'informazione genetica ci possono essere degli errori che alterano la sequenza nucleotidica creando delle mutazioni. Pertanto quando una cellula si divide le sue figlie sono spesso non del tutto uguali tra di loro. Questa mutazione può essere del tutto ininfluenza, migliorare la cellula o causare seri problemi. Questi cambiamenti possono essere mantenuti grazie alla selezione naturale ed è immediato notare come il terzo tipo di mutazione raramente verrà propagato. Alcune parti del genoma possono cambiare più facilmente: un segmento di DNA che non codifica proteine e non ha significativi ruoli regolatori può cambiare ad un tasso limitato unicamente dalla frequenza degli errori casuali, mentre un gene che codifica una proteina essenziale o una molecola di RNA genera quasi sempre una cellula che viene eliminata. Geni di quest ultimo tipo sono detti altamente conservati. Questi sono i geni da osservare se si vogliono determinare le relazioni tra gli organismi più lontani. La classificazione nei tre domini si basa sull'analisi delle componenti di rRNA dei ribosomi.

### 2.2.6 Nuovi geni sono generati da geni preesistenti

Il materiale dell'evoluzione sono sequenze di DNA preesistenti e l'innovazione può accadere in molti modi:

- Mutazione intragenica: un gene casuale può essere modificato da cambiamenti nella sua sequenza di DNA attraverso errori che accadono principalmente nel processo della replicazione del DNA.
- Duplicazione genica: un gene esistente può essere accidentalmente duplicato in modo da creare un paio di geni identici all'interno della cellula che possono successivamente divergere.
- Mescolamento dei segmenti di DNA: due o più geni esistenti possono rompersi e raggrupparsi creando un gene ibrido consistente di un segmento che prima apparteneva a geni diversi.
- Trasferimento intracellulare orizzontale: un segmento di DNA può essere trasferito dal genoma di una cellula ad un'altra.

Ognuno di questi cambi lascia delle tracce caratteristiche ed è chiaro come siano avvenuti tutti.

### 2.2.7 La duplicazione genica permette la creazione di famiglie di geni imparentati in una stessa cellula

Una cellula duplica il suo intero genoma ogni volta che si divide, ma può accadere che la duplicazione abbia degli errori con una conservazione di segmenti originali e duplicati in una singola cellula. Una volta che un gene viene così duplicato, una delle coppie è libera di mutare e specializzarsi in una funzione diversa. Diverse iterazioni danno origine a famiglie di geni che possono essere trovati nello stesso genoma. Attraverso questo processo gli individui di una specie possiedono diverse varianti di un gene primordiale. Si chiamano ortologi quei geni che si trovano in due specie diverse e derivano dallo stesso gene ancestrale nell'ultimo antenato comune, mentre si chiamano paraloghi quei geni che sono risultati da una duplicazione genica in un singolo genoma e probabilmente hanno ora funzioni differenti. Entrambi si classificano come geni omologhi.

### 2.2.8 I geni possono essere trasferiti tra organismi

I procarioti forniscono un buon esempio di trasferimento orizzontale dei geni da una specie di cellule all'altra. I segni più ovvi derivano da sequenze virali (dei batteriofagi). I virus sono piccoli pacchetti di materiale genetico evoluti come parassiti sui processi biochimici e riproduttivi della cellula. Non sono organismi viventi ma servono come vettori per il trasferimento di geni. Un virus si replica in una cellula, emerge da essa con un involucro protettivo e penetra, infettandola, un'altra cellula, che può essere anche di un'altra specie. Spesso la cellula infettata può morire, ma alcune volte il DNA virale potrebbe persistere nell'host per molte generazioni come un passeggero innoquo come plasmide o come sequenza inserita nel genoma regolare. Nei viaggi i virus possono recuperare frammenti di DNA dal genoma della cellula host e trasportarli in un'altra. Questi scambi sono comuni in procarioti ma rari in eucarioti di specie diverse. Molti procarioti possono recuperare anche DNA non virale dall'ambiente. Attraverso questi metodi batteri e archei possono acquisire geni da cellule vicine facilmente. Gli scambi orizzontali hanno un analogo tra gli eucarioti nell'attività sessuale.

#### 2.2.9 Molti geni sono comuni tra tutti i rami primari dell'albero della vita

La sequenza di un gene permette di capire la sua funzione, confrontandola con un database preesistente. Data la sequenza genomica di organismi rappresentativi per archei, batteri e eucarioti e considerando gli scambi orizzontali si nota come i geni in comune sono principalmente quelli del sistema di traduzione, trascrizione e trasporto di amminoacidi.

#### 2.2.10 Le mutazioni rivelano la funzione dei geni

L'analisi dei geni dipende da due approcci complementari: genetica e biochimica. La prima comincia con uno studio dei mutanti: si trova un organismo in cui il gene è alterato e si esaminano gli effetti sulla struttura e prestazioni. La biochimica analizza invece la funzione delle molecole. Combinando le due è possibile trovare quelle molecole la cui produzione dipende da un dato gene determinando allo stesso tempo il ruolo delle molecole nelle operazioni dell'organismo. La biologia molecolare ha permesso un rapido progresso in quanto si possono testare le contribuzioni dei geni all'attività del loro prodotto costruendo geni artificiali che combinano parte di un gene e parte di un altro. Gli organismi possono essere ingegnerizzati per produrre l'RNA o la proteina specificata dal gene in grandi quantità.

### 2.3 Informazione genetica negli eucarioti

Le cellule eucariote sono più grandi ed elaborate rispetto alle cellule procariote, come i loro genomi. La dimensione maggiore è accoppiata con radicali differenze strutturali e funzionali. Inoltre le cellule eucariote formano organismi multicellulari che arrivano a livelli di complessità maggiori.

#### 2.3.1 Le cellule eucariote potrebbero essersi originate come predatori

Per definizione le cellule eucariote mantengono il proprio DNA nel nucleo. L'involucro nucleare, un doppio strato di membrana circonda il nucleo e separa il DNA dal citoplasma. Sono circa 1000 volte più voluminose rispetto ai procarioti e possiedono un citoscheletro elaborato, un sistema di filamenti proteici nel citoplasma che forma, insieme ad altre proteine un sistema strutturale e motile. Esiste una serie di membrane simile alla membrana citoplasmatica che racchiudono diversi spazi nella cellula, molti dei quali riguardano digestione e secrezione. Non hanno una parete cellulare gli organismi procarioti unicellulari sono chiamati protozoi e possono cambiare la loro forma e intrappolare altre cellule e piccoli oggetti attraverso fagocitosi. Una delle ipotesi sull'origine delle cellule eucariote riguarda una cellula primordiale predatrice. I movimenti rapidi erano necessari alla caccia e il nucleo necessario alla protezione del genoma rispetto ai movimenti del citoscheletro.

#### 2.3.2 Le moderne cellule eucariote si sono evolute da una simbiosi

Tutte le cellule eucariote contengono i mitocondri, piccoli corpi nel citoplasma incapsulati in un doppio strato di membrana che consumano ossigeno e intrappolano energia dall'ossidazione di molecole nutritive come gli zuccheri per produrre la maggior parte dell'ATP. Sono simili in dimensioni ai piccoli batteri e possiedono il proprio genoma nella forma di una molecola di DNA circolare, i propri ribosomi e il proprio rRNA. È generalmente accettato che si sono originati da una forma di batterio aerobico fagocitato da una cellula ancestrale anaerobica. Sfuggendo alla digestione questi batteri si sono evoluti in simbiosi con la cellula fagocitante e la sua progenie, ricevendo protezione

e nutrimento in cambio di energia. Recenti analisi genomiche suggeriscono che la prima cellula eucariota si sia formata rispetto a una cellula archea fagocitata da un batterio aerobico. La maggior parte delle cellule eucariote di piante e alge contengono anche un'altra classe di membrane chiamate i cloroplasti che svolgono la fotosintesi che come i mitocondri possiedono il proprio genoma. Si sono quasi certamente generati come batteri simbiotici fotosintetici, acquisiti da cellule eucariote che avevano già i mitocondri. Una cellula eucariota equipaggiata con i cloroplasti non necessita di predare in quanto riceve nutrimento dalla fotosintesi e pertanto hanno perso la capacità di muoversi, scambiandola con la formazione di una parete cellulare protettiva. I funghi, come le cellule animali possiedono mitocondri ma non cloroplasti e una spessa parete esterna che limita la loro abilità di muoversi rapidamente o di fagocitare altre cellule. Si sono evoluti in spazzini, nutrendosi degli scarti di altre cellule o delle cellule stesse morte.

#### 2.3.3 Gli eucarioti possiedono genomi ibridi

L'informazione genetica degli eucarioti ha un'origine ibrida: dall'ancestrale cellula archea anaerobica e dal batterio che vi si è adattato come simbiote. La maggior parte dell'informazione è conservata nel nucleo, ma una piccola quantità rimane all'interno del mitocondrio e nei cloroplasti. Quando il DNA di mitocondri e cloroplasti è separato dal DNA nucleare e analizzato e sequenziato si nota come siano versioni degeneri e ridotte dei corrispondenti genomi batterici. La ragione di questo è che molti geni si sono spostati da essi al DNA del nucleo che mostra pertanto chiare prove dell'origine batterica.

#### 2.3.4 I genomi eucarioti sono grandi

La selezione naturale ha favorito mitocondri con piccoli genomi, mentre il genoma nucleare sembra essersi ingrandito, probabilmente la dimensione maggiore era un vantaggio nella vita predatoria. Grazie all'accumulazione di segmenti di DNA derivati dagli elementi trasportabili parassitici il genoma della maggior parte degli eucarioti è molti ordini di grandezza maggiore rispetto ai batteri e agli archei. Questa maggiore disponibilità porta ad avere più geni e maggior parti di DNA non codificante (98% per gli esseri umani). Molto del DNA non codificante è quasi certamente inutile, ma una parte svolge l'attività di regolare l'espressione dei geni adiacenti, cruciale per la formazione di sistemi multicellulari complessi.

#### 2.3.5 Il genoma definisce il programma dello sviluppo multicellulare

Le cellule in piante e animali sono estremamente varie, ma comunque derivano tutte dalla stessa cellula e contengono per la maggior parte copie identiche dello stesso genoma. Le differenze risultano dal modo in cui le cellule fanno uso selettivo delle proprie istruzioni geniche secondo gli indizi che ricevono dall'ambiente sviluppante. La cellula si comporta come una macchina multifunzione con sensori e abilità di esprimere diversi insiemi di geni secondo la sequenza di segnali che riceve. Il genoma in ogni cellula è grande abbastanza da accomodare le informazioni per l'intero sistema multicellulare, di cui viene usata un'unica parte. Un numero di geni codifica proteine che regolano l'attività di altri geni detti regolatori di trascrizione e si legano, direttamente o indirettamente al DNA regolatorio adiacenti ai geni da controllare. Le cellule sono inoltre in grado di inviare segnali con i propri vicini, pertanto lo stesso sistema di controllo governa ogni cellula, con diverse conseguenze in base ai messaggi scambiati. Il risultato è un preciso insieme di cellule in stati diversi.



### **2.3.6 Molti eucarioti vivono come cellule solitarie**

Molte specie di cellule eucariote formano organismi unicellulari come predatori (protozoi) o come fotosintetizzatori (algae unicellulari) o come spazzini (funghi unicellulari o lieviti). L'anatomia dei protozoi è elaborata e include strutture complesse come sensori, organi motili e di attacco. Nei termini di lignaggio e sequenze di DNA questi organismi presentano molte più differenze rispetto alle controparti multicellulari.

## Capitolo 3

# Chimica e bioenergetica della cellula

### 3.1 Le componenti chimiche della cellula

Gli organismi viventi sono composti da un piccolo sottoinsieme di elementi: carbonio (C), idrogeno (H), azoto (N) e ossigeno (O) formano il 96.5% del peso della cellula. Gli atomi di questi elementi sono legati da legami covalenti in modo da formare molecole in quanto sono più resistenti delle energie termiche all'interno della cellula e sono rotti solo durante specifiche reazioni con altri atomi e molecole. Due molecole diverse possono essere tenute insieme da legami non covalenti molto più deboli.

#### 3.1.1 L'acqua è mantenuta da legami a idrogeno

Le reazioni all'interno della cellula avvengono in ambiente acquoso, pertanto la vita si basa sulle proprietà chimiche dell'acqua. In ogni molecola d'acqua ( $H_2O$ ) i due atomi di H sono legati all'atomo O da due legami covalenti altamente polari, pertanto si trova una distribuzione ineguale di elettroni che causa una regione carica positivamente verso gli atomi H e negativamente verso l'O. Quando una parte carica positivamente si avvicina a una negativa si formano legami a idrogeno, molto meno forti di quelli covalenti e facilmente rotti dall'energia termica delle molecole. Questi legami durano pertanto un periodo breve. Questi legami sono responsabili dello stato liquido dell'acqua, dell'alta tensione superficiale e punto di ebollizione. Alcune molecole come gli alcoli che possiedono legami polari possono formare legami a idrogeno con l'acqua si dissolvono facilmente in acqua e sono chiamate idrofile (zuccheri, DNA, RNA e la maggior parte delle proteine). Le molecole idrofobiche invece sono apolari e non formano legami a idrogeno e pertanto non si dissolvono nell'acqua, un importante esempio sono gli idrocarburi, in cui gli H sono legati con gli atomi di C attraverso legami non polari, questa proprietà è sfruttata dalle cellule le cui membrane sono costruite da molecole con lunghe catene idrocarburiche.

#### 3.1.2 Quattro tipi di attrazione non covalente aiuta a unire le molecole nelle cellule

Molto della biologia dipende dagli specifici legami causati da legami non covalenti: attrazione elettrostatica (legami ionici), legami a idrogeno e attrazioni di van der Waals e un quarto fattore che

è la forza idrofobica. Nonostante ognuna di queste forze da sola sarebbe troppo debole per essere efficace si sommano tra loro in modo da creare una forte attrazione tra due molecole separate. Si noti anche come formando un'interazione competitiva con queste molecole l'acqua riduca fortemente la forza delle attrazioni elettrostatiche e dei legami a idrogeno.

#### 3.1.3 Alcune molecole polari formano acidi e basi in acqua

Una delle reazioni chimiche più significative nella cellula occorre quando una molecola con un legame covalente altamente polare tra un idrogeno e un altro atomo si dissolve in acqua. Tale idrogeno ha quasi completamente perso il proprio elettrone e pertanto esiste quasi come nucleo di idrogeno caricato positivamente ( $H^+$ ). Quando la molecola polare viene circondata da molecole d'acqua il protone viene attratto dalla loro carica parzialmente negativa e si può dissociare dalla molecola originale formando uno ione idronio ( $H_3O^+$ ). La reazione inversa accade molto velocemente, pertanto in soluzione acquosa i protoni continuano a spostarsi tra una molecola e l'altra. Le sostanze che compiono questa reazione sono dette acidi e maggiore la concentrazione di  $H_3O^+$ , più acida la soluzione. Questo ione risulta presente anche in acqua pura a causa del continuo movimento di protoni, in una concentrazione  $10^{-7}M$ . Per convenzione la concentrazione di  $H_3O^+$  è riferita come la concentrazione di  $H^+$  e espressa utilizzando la scala del pH, logaritmica. L'acqua pura ha un valore di 7 ed è detta neutra. Per valori di pH maggiori di 7 è detta basica, per valori minori detta acida. Gli acidi si caratterizzano in forti o deboli in base a quanto facilmente perdono i protoni in acqua. Molti degli acidi importanti per la cellula sono deboli. A causa dell'effetto sulla natura delle molecole dei protoni liberi l'acidità all'interno della cellula deve essere regolata. L'opposto di un acido è una base, molecole che accettano un protone in soluzione acquosa, ancora una volta nelle cellule sono presenti per la maggior parte basi deboli. Acidi e basi hanno azioni contrastanti e tendono ad annullare reciprocamente il loro effetto, pertanto l'interno della cellula è mantenuto vicino alla neutralità da buffer, acidi e basi deboli che tendono a compiere scambi di protoni ad un pH vicino a 7, mantenendo l'ambiente della cellula costante.

#### 3.1.4 La cellula è formata da composti di carbonio

Senza considerare l'acqua e gli ioni inorganici come il potassio, la maggior parte delle molecole nella cellula sono basate sul carbonio, un atomo con la capacità di formare grosse molecole. Siccome il carbonio è piccolo e ha possiede quattro elettroni liberi nel livello esterno può formare legami covalenti con altri atomi e con sè stesso, in modo da formare catene ed anelli in modo da generare molecole grandi e complesse. I composti del carbonio sono detti molecole organiche. Alcune combinazioni di atomi compaiono ricorrentemente nella cellula e ognuno di questi gruppi possiede proprietà chimiche e fisiche proprie che influiscono sul loro comportamento.

#### 3.1.5 Le cellule contengono quattro principali famiglie di piccole molecole organiche

Le molecole della cellula sono basate sul carbonio e hanno pesi molecolari tra 100 e 1000 e contengono circa 30 atomi di carbonio. Sono solitamente trovate libere in soluzione. Alcune sono usate come monomeri per costruire macromolecole polimeriche, altre agiscono come fonti di energia e sono divise e trasformate in altre piccole molecole con più ruoli nella cellula. Sono molto meno presenti rispetto alle macromolecole. Tutte le molecole organiche sono sintetizzate e divise dallo stesso insieme di componenti, pertanto i composti nella cellula sono chimicamente simili e possono essere classificati per la maggior parte in zuccheri, acidi grassi, nucleotidi e amminoacidi.

#### 3.1.6 La chimica della cellula è dominata da macromolecole con proprietà notevoli

Le macromolecole sono le molecole organiche più abbondanti per peso nella cellula. Sono le strutture principali per la costruzione e la cellula e definiscono le proprietà degli organismi viventi. Le macromolecole nella cellula sono polimeri che sono costruiti legando covalentemente piccole molecole organiche (monomeri) in lunghe catene. Le proteine sono abbondanti e versatili, alcune servono da enzimi, che catalizzano tutte le reazioni interne alla cellula, altre hanno funzione strutturale, o per compattare il DNA nei cromosomi, altre ancora agiscono come produttrici di forza motile. Nonostante le reazioni chimiche per la formazione di polimeri varino tra proteine, acidi nucleici e polisaccaridi in tutte la molecola cresce grazie all'addizione di un monomero in una fine di una catena in una reazione di condensazione, in cui una molecola di acqua è persa con ogni subunità aggiunta. Questa operazione richiede gli stessi enzimi per tutta la molecola ed è pertanto facilmente serializzabile. Tranne i polisaccaridi i monomeri che formano le macromolecole esistono in diverse varianti e richiedono pertanto una sequenza precisa di addizione per formare la macromolecola corretta.

#### 3.1.7 I legami non covalenti specificano la forma di una macromolecola e i suoi legami con le altre molecole

La maggior parte dei legami covalenti nella macromolecola permettono una rotazione degli atomi che legano, permettendo grande flessibilità garantendo ad essa un gran numero di conformazioni possibili quando l'energia termica causa rotazioni. Nonostante questo la maggior parte delle macromolecole biologiche sono altamente costrette a causa di un gran numero di legami non covalenti che si formano tra diverse parti della stessa molecola e causano la macromolecola in una conformazione particolare, determinata dalla sequenza lineare dei monomeri. Questo accade nella maggior parte delle proteine e in molte delle piccole molecole di RNA. Questi legami non covalenti possono anche creare forte attrazione tra molecole diversi in modo da creare un'interazione molecolare con alta specificità e con vari gradi di affinità, permettendo rapida dissociazione dove necessario. Questo processo è fondamentale per tutte le catalisi biologiche, permettendo il loro funzionamento come enzimi e per la creazione di strutture cellulari complesse.

### 3.2 Catalisi e utilizzo dell'energia da parte delle cellule

Una delle principali differenze tra organismi viventi e non viventi è che i primi creano e mantengono ordine. Per farlo necessitano di operare un insieme di reazioni chimiche in cui alcune molecole sono separate per mettere a disposizione altre molecole per costruirne altre.

#### 3.2.1 Il metabolismo della cellula è organizzato dagli enzimi

Le reazioni chimiche che una cellula performa accadrebbero normalmente unicamente ad alte temperature, pertanto ogni reazione richiede un'accelerazione specifica nella sua reattività. Questo fatto permette alla cellula di controllare la sua chimica. Il controllo è operato da catalizzatori biologici specializzati, proteine detti enzimi o RNA detto ribosomi. Ogni enzima catalizza una delle possibili reazioni che sono connesse in serie in modo che il prodotto di una sia il substrato di un'altra. Questi cammini lineari sono legati uno con l'altro, formando un insieme di reazioni interconnesse che permettono alla cellula di sopravvivere, crescere e riprodursi. Esistono due principali flussi di reazioni chimiche: quelle cataboliche che separano i nutrienti in piccole molecole, generando energia e alcune piccole molecole fondamentali e anaboliche o biosintetiche in cui le piccole molecole e l'energia

vengono utilizzate per guidare la sintesi delle molecole che formano la cellula. Insieme costituiscono il metabolismo della cellula.

#### 3.2.2 L'ordine biologico è reso possibile dal rilascio di calore dalla cellula

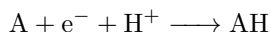
Le cellule devono ridurre il proprio livello di entropia e pertanto deve recuperare energia dall'ambiente sotto forma di cibo o fotoni, che viene poi utilizzata per generare l'ordine necessario. Nel corso di queste reazioni una parte dell'energia utilizzata viene trasformata in calore. L'energia, nel caso delle cellule animali viene ottenuta rompendo i legami dei nutrienti e viene trasformata in energia termica. La cellula non può beneficiare del calore rilasciato a meno che queste reazioni che generano energia siano accoppiati direttamente con i processi che generano l'ordine molecolare.

#### 3.2.3 Le cellule ottengono energia ossidando molecole organiche

Tutte le cellule animali e vegetali utilizzano l'energia conservata in legami chimici di molecole organiche, sia che siano zuccheri sintetizzati dalla fotosintesi sia che siano ottenuti mangiando. L'energia è stratta da un processo di ossidazione graduale. L'atmosfera contiene molto ossigeno e in presenza di ossigeno la forma di carbonio più stabile è la  $CO_2$  e quella dell'idrogeno  $H_2O$ . Una cella è pertanto capace di ottenere energia permettendo a carbonio e idrogeno delle molecole di combinarsi con l'ossigeno per produrre  $CO_2$  e  $H_2O$ . Questo processo è chiamato respirazione aerobica. La fotosintesi e la respirazione sono processi complementari. Si nota pertanto come l'utilizzo di carbonio formi un grande ciclo che coinvolge l'intera biosfera.

#### 3.2.4 Ossidazione e riduzione coinvolgono il trasferimento di elettroni

L'ossidazione nella cellula avviene attraverso l'uso di enzimi in cui il metabolismo prende le molecole attraverso un numero di reazioni che raramente coinvolgono la diretta addizione di ossigeno. L'ossidazione si riferisce a quel processo in cui elettroni sono trasferiti da un atomo all'altro (il processo inverso è la riduzione). Essendo che il numero di elettroni deve essere conservato durante una reazione ossidazione e riduzione accadono contemporaneamente: una molecola guadagna un elettrone e un'altra lo perde. Questi termini si riferiscono anche a un parziale spostamento di elettroni in un legame covalente: quando se ne crea uno polare l'atomo dalla parte del delta positivo acquisisce una parziale carica positiva ed è detto ossidato. Quando una molecola recupera un elettrone recupera anche un protone allo stesso momento e l'effetto netto è l'addizione di un atomo di idrogeno alla molecola:



Queste reazioni, dette di idrogenazione sono riduzione, mentre quelle inverse, di deidrogenazione sono dette ossidazioni. In una molecola organica avviene un'ossidazione quando il numero di legami C-H diminuisce, una riduzione quando aumenta. Le cellule utilizzano gli enzimi per catalizzare le ossidazioni attraverso una sequenza di reazioni che permettono il raccolto dell'energia prodotta.

#### 3.2.5 Gli enzimi abbassano la barriera di energia di attivazione che blocca la reazione chimica

Si noti come le reazioni chimiche procedono spontaneamente unicamente nella direzione che porta alla perdita di energia libera (energeticamente favorevoli). Essendo che le molecole negli esseri viventi si trovano in uno stato energetico relativamente stabile è necessario, affinché una reazione inizi di un'energia di attivazione, creato da collisioni randomiche insolitamente energetiche, che diventano

violente maggiore è l'energia. La chimica di una cellula è altamente controllata e il superamento del livello di energia è svolto da enzimi che si legano con un'altra molecola (substrato) in modo da ridurre l'energia di attivazione necessaria per la reazione. Questi enzimi sono detti catalizzatori e aumentano il tasso delle reazioni chimiche in quanto permettono maggiori collisioni randomiche con le molecole circostanti.

#### 3.2.6 Gli enzimi possono guidare i substrati lungo specifici cammini di reazioni

Un enzima non può cambiare il punto di equilibrio per una reazione in quanto aumenta anche il tasso della reazione inversa. Nonostante questo sono capaci di guidare le reazioni verso un cammino specifico in quanto sono altamente selettivi e molto precisi, catalizzando un'unica reazione, pertanto ogni enzima selettivamente abbassa l'energia di attivazione di una delle reazioni chimiche possibili che il substrato può svolgere. In questo modo insiemi di enzimi possono direzionare ognuna delle molecole lungo cammini specifici. Ogni enzima possiede un sito attivo, uno spazio in cui solo particolari substrati possono legarsi e dopo la reazione rimangono invariati e possono pertanto funzionare più volte.

#### 3.2.7 Come gli enzimi trovano i loro substrati, la rapidità dei movimenti molecolari

Un enzima può catalizzare la reazione di migliaia di substrati al secondo. L'attacco veloce è possibile perché il movimento causato dal calore è estremamente veloce. Questi movimenti molecolari sono classificati in:

- Movimento traslatorio: il movimento della molecola da un posto all'altro.
- Vibrazioni: il movimento di atomi legati da legami covalenti tra di loro.
- Rotazioni.

Questi movimenti aiutano ad unire le superfici di molecole che interagiscono. Il movimento delle molecole causa il processo di diffusione e in questo modo ogni molecola collide con un gran numero di altre molecole al secondo. La distanza netta alla fine di una passeggiata casuale è proporzionale alla radice del tempo impiegato. Essendo che gli enzimi si muovono più lentamente dei substrati si possono considerare fermi. Il tasso di incontro dell'enzima con il suo substrato dipende dalla concentrazione dell'ultimo. Una collisione del substrato con il sito attivo causa immediatamente la creazione del sistema enzima-substrato. L'alta specificità è data dal fatto che una forma errata causa legami covalenti più deboli dell'agitazione termica.

#### 3.2.8 Il cambio di energia libera di una reazione determina se può avvenire spontaneamente

Nonostante gli enzimi velocizzino le reazioni non possono forzare reazioni energeticamente sfavorevoli. Questo tipo è però necessario per alcune operazioni della cellula, pertanto gli enzimi accoppiano reazioni energeticamente favorevoli in modo da produrre energia che viene utilizzata per le reazioni sfavorevoli e produrre ordine biologico. Il livello di energia libera ( $G$ ) esprime l'energia disponibile per fare lavoro e viene presa in considerazione quando il sistema subisce un cambiamento ( $\Delta G$ ), critico in quanto è una diretta misura della quantità di disordine creata nell'universo da una reazione. Reazioni energeticamente favorevoli hanno un  $\Delta G$  negativo, quelle sfavorevoli positivo e possono

avvenire unicamente se accoppiate con reazioni favorevoli tale che il  $\Delta G$  totale rimanga negativo. La concentrazione dei reagenti influenza il cambio di energia libera e la direzione di una reazione. A causa di questo per comparare le reazioni si deve utilizzare il cambio di energia libera standard o  $\Delta G^\circ$ , definito alla concentrazione per reagenti di  $1 \frac{M}{L}$ . Per una reazione  $Y \longrightarrow X$  a  $37^\circ C$ ,  $\Delta G^\circ$  è in relazione a  $\Delta G$  come:

$$\Delta G = \Delta G^\circ + RT \ln \frac{[X]}{[Y]}$$

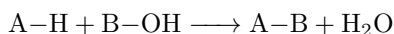
Dove  $R$  è la costante dei gas reali e  $T$  la temperatura assoluta. Si noti come al procedere della reazione il rapporto tra i reagenti cambia e avvicina  $\Delta G$  a zero, dove si raggiunge l'equilibrio chimico e non esiste un cambio di energia libera per guidare la reazione in nessuna direzione e pertanto il rapporto di prodotto e substrato raggiunge un valore costante  $K$  detta costante di equilibrio. I  $\Delta G$  delle reazioni accoppiate sono additivi, pertanto una reazione sfavorevole può essere guidata da una favorevole nel caso la seconda segua la prima.

#### 3.2.9 Molecole vettore attivate sono essenziali per la biosintesi

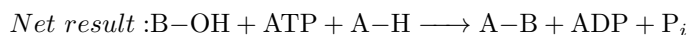
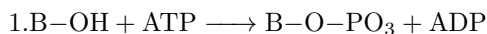
L'energia rilasciata dall'ossidazione delle molecole deve essere temporaneamente conservata prima che possa essere canalizzata nella sintesi. Nella maggior parte dei casi questo avviene come energia chimica di legame in un sottoinsieme di molecole dette molecole vettore con dei legami covalenti ricchi di energia. Queste molecole si diffondono rapidamente all'interno della cellula trasportando l'energia. Questi vettori attivati conservano energia in una forma facilmente scambiabile sottoforma di un gruppo chimico o elettrone mantenuto ad un livello energetico alto e può servire sia come fonte energetica che di materiali. Sono detti anche coenzimi. Esempi di queste molecole sono ATP, NADH e NADPH. La formazione di un vettore attivo è accoppiata con una reazione energeticamente favorevole come ossidazione: la quantità di calore rilasciata viene ridotta di una quantità pari a quella conservata nei legami, sufficiente a iniziare un'altra reazione.

#### 3.2.10 L'ATP è il vettore attivo più utilizzato

L'ATP è la molecola utilizzata per conservare e utilizzare l'energia. È sintetizzato in una reazione di fosforilazione in cui un gruppo fosfato è aggiunto all'ADP (adenina difosfato). Quando richiesto l'ATP dona il suo pacchetto di energia attraverso l'idrolisi in ADP e fosfato inorganico. L'ADP generato ritorna poi disponibile per la fosforilazione. Questa reazione è accoppiata con molte altre di sintetizzazione. Una tipica reazione biosintetica è una in cui due molecole  $A$  e  $B$  sono unite per produrre  $A-B$  durante la condensazione:



Esiste un cammino indiretto in cui un'accoppiamento con l'idrolisi dell'ATP causa la reazione di avvenire: l'idrolisi viene utilizzata per convertire  $B-OH$  in un composto intermedio più energetico che poi interagisce direttamente con  $A-H$  per formare  $A-B$ . Il meccanismo più semplice coinvolge il trasferimento di un fosfato dall'ATP a  $B-OH$  per creare  $B-O-PO_3$ :



La reazione di condensazione è pertanto forzata dall'idrolisi in un cammino catalizzato da un enzima.

### 3.2.11 NADH e NADPH sono vettori di elettroni

Altri vettori che partecipano nelle reazioni di ossidazione-riduzione sono parte delle reazioni accoppiate. Questi vettori sono specializzati nel trasportare elettroni mantenuti ad alta energia e atomi di idrogeno. I principali è  $\text{NAD}^+$  (dinucleotide adenina nicotinammide) e  $\text{NADP}^+$  (dinucleotide adenina nicotinammide fosfato) che raccolgono un pacchetto di energia corrispondente a due elettroni e un protone  $\text{H}^+$  e sono convertiti in NADH (dinucleotide adenina nicotinammide ridotto) e NADPH (dinucleotide adenina nicotinammide fosfato ridotto) che possono pertanto essere considerate vettori di ioni idruri. Il NADPH è prodotto durante un insieme di reazioni cataboliche che producono energia: due atomi di idrogeno sono rimossi da una molecola di substrato. Entrambi gli elettroni ma non un  $\text{H}^-$  sono aggiunti all'anello di  $\text{NADP}^+$  per formare NADPH, mentre il protone  $\text{H}^+$  è rilasciato in soluzione. Il NADPH rilascia facilmente lo ione idruro in una reazione di ossidazione-riduzione in modo da raggiungere uno stato più stabile ottenendo un rilascio di energia libera negativa. Il gruppo fosfato del NADPH dà alla molecola una forma completamente diversa rispetto al NADH, rendendole riconoscibili a enzimi completamente diversi in quanto si rende necessario regolare due reazioni di trasferimento di elettroni indipendentemente. Il NADPH opera con gli enzimi che catalizzano le reazioni anaboliche di sintesi per le molecole ricche di energia, mentre il NADH opera come intermedio nel sistema di reazioni cataboliche che generano ATP. La genesi delle due molecole avviene essa stessa indipendentemente in modo da avere un controllo fine sul rapporto tra  $\text{NAD}^+$  e NADH (mantenuto alto) e tra  $\text{NADP}^+$  e NADPH, mantenuto basso.

### 3.2.12 Esistono molte altre molecole vettori nella cellula

Altri vettori attivati possono trasportare un gruppo in un legame ad alta energia e instabile. Il resto della molecola, solitamente la parte più grande, permette il riconoscimento da parte di enzimi. Molte di queste parti contengono un nucleotide (solitamente adenosina). Si noti come ATP trasferisce fosfato, NADPH elettroni e idrogeno. Altri vettori attivati trasferiscono gruppi utilizzati per la biosintesi e sono generati in reazioni accoppiate con l'idrolisi dell'ATP.

### 3.2.13 La sintesi di polimeri biologici è guidata dall'idrolisi dell'ATP

Le molecole della cellula sono composte da subunità unite in una reazione di condensazione, in cui i costituenti di una molecola d'acqua ( $\text{OH}$  e  $\text{O}$ ) sono rimossi dai reagenti. Conseguentemente l'azione inversa, la rottura di tutti e tre i tipi di polimeri avviene attraverso idrolisi catalizzata, energeticamente favorevole. Acidi nucleidici, proteine e polisaccaridi sono polimeri prodotti dall'addizione ripetuta di monomeri a un capo di una catena che cresce. La condensazione in ogni passaggio richiede energia data dall'idrolisi di un nucleoside trifosfato. Per ogni tipo di macromolecola esiste un cammino catalizzato: il gruppo  $-\text{OH}$  che viene rimosso nella reazione di condensazione è prima attivato legandosi con una molecola che lo porta ad uno stato energetico elevato, attraverso una serie di intermedi ad alta energia. Ogni vettore attivato ha dei limiti nella sua capacità di guidare una reazione sintetica: il  $\Delta G$  dell'idrolisi dell'ATP dipende dalla concentrazione dei reagenti, ma si trova tra  $-46$  e  $-54 \frac{\text{kJ}}{\text{M}}$  e potrebbe pertanto guidare una reazione con  $\Delta G = +40 \frac{\text{kJ}}{\text{M}}$  che non è abbastanza per alcune reazioni. In questi casi l'idrolisi viene alterata in modo che produca AMP e pirofosfato ( $\text{PP}_i$ ) che si idrolizza in un passo successivo. L'intero processo crea un  $\Delta G = -100 \frac{\text{kJ}}{\text{M}}$ . Una reazione che avviene in questo modo è la sintesi degli acidi nucleici dal nucleoside trifosfato. La condensazione ripetitiva può essere orientata verso la testa o verso la coda. Nella head polymerization, il legame reattivo è trasportato alla fine del polimero crescente e deve pertanto essere rigenerato ogni volta che un monomero viene aggiunto e il monomero trasporta con sé il legame che viene utilizzato per il prossimo monomero. Nella tail polymerization il legame è trasportato da ogni monomero.



## 3.3 Come le cellule ottengono energia dai nutrienti

La riserva costante di energia necessaria alla generazione e mantenimento dell'ordine delle cellule è ottenuta dai legami chimici energetici nelle molecole dei nutrienti. Proteine, lipidi e polisaccaridi che li compongono sono rotti in piccole molecole prima che la cellula possa usarle attraverso digestione enzimatica e successivamente entrano il cytosol della cellula, dove avviene una graduale ossidazione. Gli zuccheri sono un'importante fonte di energia e vengono ossidati in passi onorati in anidride carbonica ( $\text{CO}_2$ ) e acqua.

### 3.3.1 La glicolisi è un cammino centrale per produrre ATP

Il processo principale per l'ossidazione degli zuccheri è la glicolisi, una serie di reazioni che produce ATP senza ossigeno molecolare. Avviene nel cytosol e include molti microorganismi anaerobici. Durante la glicolisi una molecola di glucosio è convertita in due molecole di piruvato a tre atomi di carbonio. Per ogni molecola di glucosio sono idrolizzate due molecole di ATP per fornire energia nei primi passaggi, ma alla fine sono prodotte quattro molecole di ATP, con un guadagno netto di due ATP e di NADH. La glicolisi coinvolge 10 reazioni separate in ognuna delle quali viene prodotto uno zucchero intermedio e catalizzata da un enzima diverso. L'ossidazione avviene rimuovendo elettroni grazie a  $\text{NAD}^+$  producendo NADH dal carbonio derivato dalla molecola di glucosio. La natura a passaggi rilascia l'energia in piccoli pacchetti che possono essere salvati in un vettore attivo. Alcuni dell'energia rilasciata guida la sintesi di ATP da ADP e  $\text{P}_i$  e altra rimane negli elettroni nel vettore NADH. Negli organismi aerobici le molecole di NADH donano gli elettroni a una catena di trasporto e il  $\text{NAD}^+$  viene utilizzato ancora per la glicolisi.

### 3.3.2 La fermentazione produce ATP in assenza di ossigeno

Per la maggior parte delle cellule la glicolisi è solo un preludio al passaggio finale della rottura dei nutrienti: il piruvato è trasportato ai mitocondri, dove è convertito in  $\text{CO}_2$  e acetile CoA, il cui gruppo acetile è successivamente ossidato in  $\text{CO}_2$  e  $\text{H}_2\text{O}$ . In contrasto, per molti organismi anaerobici la glicolisi rimane la fonte principale di ATP. In queste condizioni anaerobiche il piruvato e gli elettroni del NADH rimangono nel cytosol. Il piruvato è convertito in prodotti secreti dalla cellula come etanolo e  $\text{CO}_2$  nei lieviti o in acido lattico nei muscoli. In questo processo il NADH libera i suoi elettroni ed è convertito in  $\text{NAD}^+$  per mantenere la reazione di glicolisi. Questi processi sono chiamati fermentazioni.

### 3.3.3 La glicolisi mostra come gli enzimi accoppiano l'ossidazione alla conservazione di energia

Due reazioni centrali della glicolisi (passi 6 e 7) convertono lo zucchero intermedio gliceraldeide 3-fosfato in 3-fosfogliceraldeide, ossidando un gruppo aldeide in un acido carbossilico. La reazione completa rilascia energia libera per convertire una molecola di ADP in una di ATP e per trasferire due elettroni e protoni dall'aldeide al  $\text{NAD}^+$ , contemporaneamente liberando abbastanza calore da rendere la reazione energeticamente favorevole con  $\Delta G = -12.5 \frac{\text{kJ}}{\text{M}}$ . La reazione è guidata da due enzimi che si legano agli zuccheri intermedi. Il primo enzima (gliceraldeide 3-fosfato deidrogenasi) forma un legame covalente temporaneo con l'aldeide attraverso un gruppo  $-\text{SH}$  e catalizza la sua ossidazione attraverso  $\text{NAD}^+$  nel suo sito attivo. Il legame substrato-legame è successivamente rotto da uno ione fosfato per produrre un intermedio ad alta energia che viene rilasciato dall'enzima. Questo secondo intermedio si lega al secondo enzima fosfoglicerato chinasi che catalizza il trasferimento

del fosfato all'ADP, formando l'ATP e completando il processo di ossidazione dell'aldeide in acido carbossilico. La rottura del legame fosfato crea l'energia necessaria per la sintesi dell'ATP.

#### 3.3.4 Gli organismi conservano le molecole nutrienti in serbatoi speciali

Tutti gli organismi devono mantenere un rapporto ATP-ADP alto per rimanere in vita. Per compensare a lunghi periodi senza accesso a nutrienti gli animali conservano gli acidi grassi come goccioline di grasso composte di trigliceridi insolubili principalmente nel citoplasma di cellule di grasso specializzate (adipociti). Per conservazione a breve termine lo zucchero è conservato come glicogeno, la cui sintesi e degradazione sono rapidamente regolate al bisogno. Quando le cellule necessitano di più ATP di quanto riescano a produrre dai nutrienti nel sangue degradano il glicogeno in glucosio 1-fosfato che viene convertito in glucosio 6-fosfato per la glicolisi. Quantitativamente il grasso è più presente del glicogeno in quanto è più efficiente in quanto libera più energia ed è insolubile in acqua. Lo zucchero e l'ATP necessario alle piante è prodotto in organelli separati come i cloroplasti per la fotosintesi e i mitocondri per l'ATP. Essendo questi organelli isolati da una membrana e mancanti in alcune cellule, gli zuccheri sono esportati dai cloroplasti ai mitocondri di tutte le cellule. Durante i periodi di eccesso di capacità fotosintetica i cloroplasti convertono dello zucchero in grassi e amido, un polimero del glucosio analogo al glicogeno. I grassi nelle piante sono trigliceridi e differiscono unicamente nel tipo di acidi grassi che predominano. Entrambi sono conservati nei cloroplasti. Quando il glucosio raggiunge un livello soglia negli animali i trigliceridi sono idrolizzati per produrre acidi grassi e glicerolo e sono trasferiti nel flusso sanguigno. Gli acidi grassi sono ossidati direttamente.

#### 3.3.5 Zuccheri e acidi grassi sono degradati in acetile CoA nei mitocondri

Nei metabolismi aerobici il piruvato prodotto dalla glicolisi nel citosol è trasportato nei mitocondri, dove è decarbossilato da un complesso di tre enzimi detto il complesso piruvato deidrogenasi. Il prodotto sono una molecola di  $\text{CO}_2$ , una di NADH e acetil CoA. Gli acidi grassi dal flusso sanguigno sono trasportati nei mitocondri dove sono ossidati. Ogni molecola di acido grasso (attivata, acile grasso CoA) è rotta completamente da un ciclo di reazione che rompe due atomi di carbonio dalla coda carbossile generando una molecola di acetile CoA per ogni ciclo. Una molecola di NADH e una di  $\text{FADH}_2$  sono prodotte. La maggior parte di questa energia rimane conservata nelle molecole di acetile CoA e pertanto entra in gioco il ciclo di reazione di acido citrico, in cui il gruppo acetile ( $-\text{COCH}_3$ ) viene ossidato in  $\text{CO}_2$  e  $\text{H}_2\text{O}$ .

#### 3.3.6 Il ciclo di acido citrico genera NADH ossidando i gruppi acetili in $\text{CO}_2$

Il ciclo di acido citrico o ciclo di acido tricarbossilico o ciclo di Krebs crea i due terzi dell'ossidazione totale dei composti di carbonio nelle cellule e i suoi prodotti principali sono  $\text{CO}_2$  e elettroni ad alta energia nella forma di NADH. La prima è secreta come scarto, mentre i secondi sono passati a catene di trasporto di elettroni legate alla membrana, eventualmente combinandosi con l'ossigeno per formare acqua. Il ciclo in sé non richiede ossigeno, ma viene utilizzato in reazioni seguenti e avviene nei mitocondri. Il gruppo acetile non è ossidato direttamente: il gruppo è trasferito da CoA a una molecola di ossalacetato per formare l'acido tricarbossilico o acido citrico, che viene gradualmente ossidato permettendo la produzione di molecole vettore ad alta energia. Alla fine delle otto reazioni l'ossalacetato viene rigenerato e rientra in un altro ciclo. La molecola di  $\text{FADH}_2$  (flavina adenina dinucleotide ridotto) viene prodotta da FAD e il ribonucleoside trifosfato GTP dal

GDP, molto simile dell'ATP, il cui trasferimento del gruppo fosfato all'ADP produce ATP. L'energia conservata in NADH e FADH<sub>2</sub> viene utilizzata per la produzione di ATP attraverso fosforilazione ossidativa, che richiede ossigeno gassoso. L'acqua mette a disposizione gli atomi di ossigeno necessari alla produzione di CO<sub>2</sub>. Alcuni amminoacidi passano dal citosol ai mitocondri, dove sono convertiti in acetile CoA o altri intermedi del ciclo di Krebs che insieme alla glicolisi forma un punto per le reazioni biosintetiche producendo intermedi come ossalacetato e  $\alpha$ -chetoglutarato. Alcune di queste sostanze prodotte dal catabolismo sono trasferite dal mitocondrio al citosol, dove vengono utilizzate in reazioni anaboliche come precursori per la sintesi di molte molecole essenziali.

#### 3.3.7 Il trasporto di elettroni guida la sintesi della maggior parte dell'ATP nelle cellule

La maggior parte dell'energia chimica rilasciata nell'ultima parte della degradazione di un nutriente permette il trasferimento da parte di NADH e FADH<sub>2</sub> degli elettroni che hanno preso durante l'ossidazione del nutriente a catene di trasporto di elettroni, incorporate nella membrana interna del mitocondrio. Come gli elettroni passano lungo la catena di molecole elettrone-accettori e molecole elettrone-donatori passano sequenzialmente a stati di energia minore. Questa energia pompa protoni H<sup>+</sup> lungo la membrana dalla matrice mitocondriale allo spazio intermembrana e al citosol, generando un gradiente di ioni H<sup>+</sup> che serve come fonte di energia per la cellula per reazioni come la generazione di ATP attraverso fosforilazione di ADP. Alla fine della serie di trasferimenti gli elettroni sono passati a molecole di ossigeno gassoso diffuso nel mitocondrio che producono acqua. Questo processo detto fosforilazione ossidativa succede anche nella membrana plasmatica dei batteri. L'ossidazione di una molecola di glucosio in H<sub>2</sub>O e CO<sub>2</sub> è utilizzata per produrre 30 molecole di ATP.

#### 3.3.8 Amminoacidi e nucleotidi sono parte del ciclo azotato

Gli atomi di azoto e zolfo passano da composto a composto e all'ambiente in una serie di cicli reversibili. Essendo l'azoto poco reattivo solo alcuni organismi sono in grado di incorporarlo attraverso fissazione. La maggior parte dell'azoto organico deriva pertanto dal passaggio tra organismi. L'azoto viene ricevuto da proteine e acidi nucleici che vengono rotti in amminoacidi e nucleotidi e viene utilizzato per produrre proteine e acidi nucleici. Circa metà degli amminoacidi sono essenziali per l'uomo: non possono essere sintetizzati. I nucleotidi possono essere sintetizzati attraverso cammini biosintetici specializzati. Tutto l'azoto nelle basi purine e pirimidine sono derivati dall'amminoacido glutamina, dall'acido aspartico e dalla glicina, mentre gli zuccheri del ribosio e del desossiribosio dal glucosio. Gli amminoacidi non utilizzati in biosintesi possono essere ossidati per creare energia metabolica. Carbonio e idrogeno formano CO<sub>2</sub> e H<sub>2</sub>O, mentre gli atomi di azoto sono spostati in varie forme e appaiono come urea, che viene secreta. Ogni amminoacido viene processato diversamente. Lo zolfo, per essere usato per la vita, deve essere ridotto a solfuro (S<sub>2</sub><sup>-</sup>). Lo stato di ossidazione dello zolfo richiesto per la biosintesi di metionina, cisteina, del coenzima A e i centri a ferro-zolfo per il trasporto di elettroni. La riduzione dello zolfo comincia in batteri, funghi e piante, dove enzimi e ATP sono utilizzati per creare un cammino per l'assimilazione. Gli esseri umani non possono compiere questo processo.

## Capitolo 4

# Proteine

Le proteins costituiscono la maggior parte della massa secca della cellula. Svolgono le principali funzioni strutturali e la maggior parte delle funzioni della cellula: gli enzimi mettono a disposizione le superfici che catalizzano le reazioni chimiche. Quelle incorporate nella membrana plasmatica formano canali e pompe per il controllo del passaggio di piccole molecole nella cellula, hanno funzione di messaggeri intra e inter cellulari, si occupano del movimento della cellula, sono anticorpi, tossine, ormoni e miuulle altre funzioni.

### 4.1 Forma e struttura delle proteine

#### 4.1.1 La forma di una proteina è specificata dalla sua sequenza degli amminoacidi

Esistono 20 amminoacidi diversi che vengono codificati dal DNA di un organismo, ognuno con proprietà chimiche diverse. Una molecola protena è costituita da una lunga catena senza ramificazioni di questi amminoacidi, ognuno legato ai suoi vicini attraverso un legame peptidico covalente. Sono pertanto dette polipeptidi. Ogni tipo di proteina possiede una sequenza di amminoacidi unica e ne esistono migliaia di tipi. La sequenza ripetuti di atomi lungo il nucleo della catena polipeptidica è detto il polypeptide backbone, al quale si attaccano quelle porzioni di amminoacidi non coinvolte nel legame peptidico e che conferiscono all'amminoacido le sue proprietà uniche. Le 20 catene laterali differenti degli amminoacidi differiscono nelle proprietà: alcune sono nonpolari e idrofobiche, altre negativamente o positivamente cariche, altre formano legami covalenti. Le proteine formano una catena flessibile che può ripiegarsi in infiniti modi. Tale struttura può essere determinata da molti legami non covalenti che si formano tra una parte della catena e l'altra e sono: legami idrogeno, attrazioni elettrostatiche e forze di Van der Waals che in parallelo possono mantenere insieme due regioni della catena. La forza di questo gran numero di legami non covalenti determina la stabilità di ogni forma ripiegata. Può entrare anche in gioco una forza apparente di attrazione idrofobica. La forma viene pertanto influenzata fortemente dalla distribuzione degli amminoacidi polari e non polari. I non polaritendono a trovarsi all'interno della molecola in modo da evitare il contatto con l'acqua, mentre i gruppo polari verso l'esterno. Amminoacidi polari all'interno della proteina sono tipicamente legati con altri amminoacidi polari o al backbone polipeptidico attraverso legami a idrogeno.

##### 4.1.2 Le proteine si ripiegano nella conformazione a minore energia

La maggior parte delle proteine possiedono una particolare struttura tridimensionale. Tale conformazione è quella che minimizza la sua energia libera. La sequenza di amminoacidi contiene tutte le informazioni necessarie alla struttura della proteina. La maggior parte si piega in una singola conformazione stabile, che può variare leggermente quando interagiscono con altre molecole. Nelle cellule una proteina detta molecular chaperone assiste il processo di piegatura legandosi a catene polipeptidiche parzialmente piegate e aiutandole verso il cammino più favorevole. Sono richiesti per prevenire la formazione di aggregati proteici attraverso regioni idrofobiche temporaneamente esposte. Rendono pertanto il processo più affidabile. La maggior parte delle proteine si trovano ad una lunghezza da 50 a 2000 amminoacidi. Proteine grandi consistono di domini proteici, unità strutturali che si piegano indipendentemente.

##### 4.1.3 L' $\alpha$ -elica e il $\beta$ -foglietto sono pattern di piegamento comuni

Analizzando la struttura tridimensionale delle proteine diventa chiaro come esistano due pattern di piegamento regolari. L' $\alpha$ -elica e il  $\beta$ -foglietto e sono causati dal legame a idrogeno tra i gruppi N-H e C=O nel backbone polipeptidico, senza coinvolgere elementi della catene secondarie degli amminoacidi. Pertanto, nonostante siano incompatibili rispetto ad alcuni amminoacidi molte sequenze le possono formare. In entrambi i casi la proteina adotta una conformazione regolare e ripetuta. La parte centrale di molte proteine contiene regioni estese di  $\beta$ -foglietti che possono fermare segmenti vicini di backbone polipeptidico con la stessa orientazione (catene parallele) o che si piega su sè stessa (catene antiparallele). Un  $\alpha$ -elica viene generata quanto una singola catena polipeptidica si torce su sè stessa per formare un cilindro rigido. Un legame a idrogeno si forma ogni quarto legame peptidico, legando il C=O di un peptide con il N-H di un altro. L'elica pertanto compie un giro completo ogni 3.6 amminoacidi. Questa conformazione è abbondante nelle proteine di membrana come quelle di trasporto e i recettori, specialmente nella parte che attraversa la stessa e composta da amminoacidi non polari. In altre proteine le  $\alpha$ -eliche si avvolgono su sè stesse formando una bobina arrotolata, che si forma quando due o quattro  $\alpha$ -eliche hanno la maggior parte delle catene laterali non polari da una parte, in modo che possano avvolgersi tra di loro.

##### 4.1.4 I domini proteici sono unità modulari da cui le proteine più grandi sono costruite

Vengono distinti quattro livelli di organizzazione nella struttura di una proteina: la struttura primaria è la sequenza di amminoacidi, lunghezze di catena polipeptidica che formano  $\alpha$ -eliche e  $\beta$ -foglietti sono la struttura secondaria, la completa organizzazione tridimensionale viene detta struttura terziaria e se la proteina è costituita da un complesso di più catene polipeptidiche la struttura è detta quaternaria. Si intende per dominio proteico una sottostruttura prodotta da una qualsiasi parte contigua di catena polipeptidica che può piegarsi indipendentemente rispetto alle altre. Contengono solitamente tra i 40 e i 350 amminoacidi e sono le unità modulari che costituiscono proteine più grandi. Diversi domini di una proteina sono solitamente associati con diverse funzioni. Le proteine più piccole contengono un singolo dominio, mentre le più grandi anche a dozzine, spesso connessi da corte e non strutturate lunghezze di catena polipeptidica che formano cardini flessibili tra i domini.

##### 4.1.5 Poche delle catene polipeptidiche possibili saranno utili alla cellula

Le possibili combinazioni di una catena polipeptidica di lunghezza  $n$  sono  $20^n$  e pertanto solo una piccola frazione di questo insieme crea una conformazione tridimensionale stabile, circa una in un

miliardo. La selezione naturale ha portato la cellula a scegliere quelle proteine che, oltre a possedere tale conformazione, possiedono proprietà chimiche finemente regolate in modo da permettere alla proteina di catalizzare una particolare reazione o per svolgere la funzione strutturale richiesta.

##### 4.1.6 Le proteine possono essere classificate in molte famiglie

Una volta che la proteina si è voluta per formare una conformazione stabile con un'utilità può essere modificata attraverso meccanismi genetici in modo da creare nuove proteine con diverse funzioni. Questo processo porta alla nascita di famiglie di proteine con sequenza e conformazione simili ma con funzione distinte. La struttura di diversi membri di una famiglia di proteine è conservata maggiormente rispetto alla sequenza di amminoacidi. Molti cambi di amminoacidi sono neutri, senza effetto sulla struttura e funzione della proteina. Le proteine che subiscono cambi maligni vengono scartate durante il processo evolutivo.

##### 4.1.7 Alcuni domini proteici si trovano in molte proteine diverse

Le proteine formate da multipli domini si sono formate dall'unione accidentale di sequenze di DNA che codificano ogni dominio. Nel processo evolutivo del mescolamento del dominio, molte larghe proteine si sono evolute attraverso l'unione di domini preesistenti in nuove combinazioni. Come risultato si sono create nuove superfici leganti alla giustapposizione dei domini, dove si trovano molti dei siti funzionali della proteina. Un sottoinsieme di domini è stato molto mobile durante l'evoluzione, con strutture versatili dette moduli proteici. Alcuni domini possiedono un nucleo stabile formato da  $\beta$ -foglietti con anelli sporgenti di catena polipeptidica. Gli anelli sono situati per formare siti di legame per altre molecole. Il loro successo evolutivo è dovuto al fatto che mettono a disposizione una base per la generazione di siti di legame, richiedendo unicamente piccoli cambi agli anelli esterni. Si possono inoltre integrare facilmente all'interno di altre proteine in quanto possiedono alle terminazioni  $N^-$  e  $C^-$ . Quando il DNA codifica tale dominio svolge una duplicazione a tandem. I domini duplicati con questo ordinamento in linea possono essere collegati per formare strutture estese con sé stessi o altri domini. Queste strutture estese rigide sono comuni in matrici di molecole extracellulari e nella porzione extracellulare di proteine recettrici sulla superficie della cellula. Altri tipi di domini sono detti plug-in con i legami  $N^-$  e  $C^-$  vicini. Dopo il riordinamento genetico sono messi come inserimenti in una regione ad anello di una sequenza proteina. La frequenza di utilizzo dei domini differisce tra tipi di organismi. Il complesso maggiore di istocompatibilità (MHC) possiede un dominio di riconoscimento degli antigeni è presente unicamente negli umani, con funzioni specializzate e sono stati selezionati fortemente durante evoluzioni recenti. Molte coppie di domini si trovano insieme in molte proteine: la maggior parte delle proteine che contengono coppie di due domini si sono sviluppate relativamente tardi durante l'evoluzione.

##### 4.1.8 Il genoma umano codifica un complesso insieme di proteine, rivelando che molto rimane sconosciuto

La sequenziazione del genoma umano ha rivelato che contiene circa 21000 geni che codificano proteine e che i vertebrati hanno ereditato la maggior parte delle proteine dagli invertebrati, nonostante in media ogni proteina sia più complessa. Il mescolamento dei domini durante l'evoluzione ha portato alla creazione di molte nuove combinazioni di domini. La maggior varietà delle proteine permette più possibili interazioni proteina-proteina.

##### 4.1.9 Proteine più grandi contengono più di una catena polipeptidica

Gli stessi legami non covalenti che permettono il piegamento della proteina le permettono di legarsi con altre proteine per formare strutture più grandi nella cellula. Ogni regione di una superficie di una proteina che può interagire con altre molecole si dice sito di legame. Una proteina ne può contenere diversi. Se tale sito riconosce la superficie di una seconda proteina il legame tra le due catene polipeptidiche crea una proteina più larga con una geometria definita. Ogni catena polipeptidica in tale proteina è detta subunità proteica. Nel caso più semplice due catene polipeptidiche con la stessa conformazione possono legarsi testa-a-testa, formando un complesso simmetrico di due subunità mantenuto dall'interazione tra due siti di legame identici. Molte proteine contengono due o più tipi di catene polipeptidiche.

##### 4.1.10 Alcune proteine globulari formano lunghi filamenti elicoidali

Una proteina si dice globulare se la catena polipeptidica si ripiega su se stessa formando una forma compatta simile ad una palla con una superficie irregolare. Alcune di queste proteine si possono combinare formando lunghi filamenti se ogni molecola possiede un sito di legame complementare ad un'altra regione sulla superficie della stessa molecola. Essendo che ogni subunità si lega alle altre allo stesso modo e il legame non è mai una retta la struttura complessiva assumerà una forma ad elica.

##### 4.1.11 Molte proteine hanno forme allungate e fibrose

Gli enzimi tendono ad essere proteine globulari: nonostante molte sono larghe e complicate con multiple subunità, la maggior parte hanno una forma arrotondata. Ci sono funzioni che richiedono che ogni molecola proteica occupi una lunga distanza. Queste proteine possiedono generalmente una struttura semplice allungata e sono dette proteine fibrose. Il citoscheletro è formato da forme chiamate filamenti intermedi simili a corde. Sono abbondanti all'esterno della cellula, dove formano la maggior parte della struttura del gel della matrice extracellulare che aiuta a legare collezioni di cellule insieme per formare tessuti. Le proteine di matrice sono secrete dalle cellule e si assemblano in fibrilli lunghi.

##### 4.1.12 Le proteine contengono una grande quantità di catene polipeptidiche intrinsecamente disordinate

Un'altra molecola abbondante nella matrice proteica è l'elastina, un polipeptide altamente disordinato, il cui disordine è fondamentale per la sua funzione di produrre una mesh che può essere spostata da una conformazione all'altra. Hanno funzioni importanti nei siti di legame, prendendo una forma solo quando incontrano la molecola che legano. Una funzione predominante di queste parti è per l'appunto formare siti di legami con altre proteine ad alta specificità ma alterati da fosforilazione o defosforilazione o modifiche iniziate da eventi di segnale. Vengono utilizzate come legame per mantenere due domini proteici in prossimità in modo da permettere al substrato di muoversi tra i siti attivi in un complesso multienzima. Creano microregioni con una consistenza simile a gel che limita la diffusione.

##### 4.1.13 Legami incrociati covalenti stabilizzano proteine extracellulari

Molte proteine si trovano all'esterno della membrana plasmatica della cellula o sono secrete come matrice extracellulare. Tutte queste sono esposte alle condizioni dell'ambiente. Per mantenere la loro struttura la catena polipeptidica è stabilizzata da legami covalenti incrociati che possono

legare due amminoacidi nella stessa proteina o legare differenti catene polipeptidiche. Il legame più comune è quello zolfo-zolfo o legami disolfuro che si formano quando la cella prepara le proteine per l'esportazione. La loro formazione è catalizzata nel reticolo endoplasmatico da un enzima che lega due paia di gruppi  $-SH$  di cisteina adiacenti nella proteina piegata. Non cambiano la conformazione della proteina ma si limitano a rafforzarne la struttura. Tali legami non riescono a formarsi nel citosol a causa del gran numero di agenti riduttori.

##### 4.1.14 Le molecole proteiche spesso svolgono la funzione di subunità per l'assemblaggio di grandi strutture

Lo stesso principio che permette alle proteine di associarsi tra di loro formando anelli o filamenti operano per generare larghe strutture formate da un insieme di diverse macromolecole come complessi enzimatici, ribosomi, virus e membrane che non sono composti da una singola molecola formata da legami covalenti, ma dall'assemblaggio di molte molecole attraverso legami non covalenti che formano le subunità della struttura. Questo processo presenta dei vantaggi:

- Una grande struttura costruita da subunità ripetute richiede meno informazione genetica.
- Sia l'assemblaggio che il disassemblaggio possono essere controllati e resi reversibili in quanto i legami di associazione hanno poca energia.
- Gli errori nella sintesi della struttura sono evitati più facilmente.

Alcune subunità proteiche si assemblano in larghi fogli piatti in cui altre subunità sono ordinate in pattern esagonali. In alcuni casi proteine di membrana sono originate come un bistrato lipidico. Il foglio esagonale può essere facilmente trasformato in un tubo o una sfera cava che si legano a molecole di DNA e RNA nell'involucro dei virus. La formazione di strutture chiuse come anelli, tubi o sfere crea maggiore stabilità in quanto aumenta il numero di legami tra le subunità ed essendo creata da legami mutualmente dipendenti e cooperativi può essere assemblata o disassemblata da piccoli cambi di una subunità. Molte strutture nella cellula sono capaci di auto-assemblaggio. L'informazione per la creazione di molte delle macromolecole è contenuta nelle subunità stesse.

##### 4.1.15 I fattori di assemblaggio aiutano la formazione di complesse strutture biologiche

Nei casi in cui le molecole non si possano formare spontaneamente da una soluzione dei componenti sono necessarie informazioni supplementari fornite da enzimi e altre proteine che svolgono la funzione di stampi, ovvero fattori di assemblaggio che guidano la costruzione ma non fanno parte della struttura finale.

##### 4.1.16 Fibrille amiloide possono formarsi da molte proteine

Una classe di strutture proteiche, utilizzata per delle funzioni della cellula può contribuire a certe malattie quando non controllata. Sono  $\beta$ -foglietti autoriproduttori chiamati fibrille amiloide. Si costruiscono da una serie di catene polipeptidiche identiche che si stratificano una sull'altra creando uno stack di  $\beta$ -foglietti orientati perpendicolarmente all'asse formando un filamento cross-beta. Una porzione della proteina possiede la capacità di formare tali strutture in quanto possiede differenti sequenze e seguono diversi cammini. In condizioni normali il meccanismo che controlla le proteine si degrada con l'età, permettendo occasionalmente alle proteine di formare aggregati patologici che possono essere rilasciati dalla cellula morta e aggregarsi all'esterno come fibrille amiloide nella



matrice extracellulare che in casi estremi possono uccidere altre cellule e danneggiare i tessuti. Queste molecole possono anche essere utilizzate dalla cellula per concentrare in granuli secretori molecole da espellere.

### 4.1.17 Molte proteine contengono domini a bassa complessità che possono formare amiloidi reversibili

Un grande insieme di domini a bassa complessità possono formare fibre amiloidi che hanno ruoli funzionali nel nucleo e nel citoplasma, sono normalmente senza struttura e consistono di lunghezze di sequenze di amminoacidi lunghe centinaia di monomeri con un sottoinsieme dei 20 amminoacidi. Queste strutture formate sono unite da legami non covalenti deboli e si dissociano facilmente in presenza dei segnali corretti. Molte proteine con tale dominio contengono un altro insieme di domini che si lega specificamente ad altre proteine o al RNA, pertanto la loro aggregazione controllata può formare un idrogele che unisce queste e altre molecole in strutture puntate chiamate corpi intracellulari o granuli. mRNA specifico può essere inviato in questi granuli dove è salvato fino a che reso disponibile da un disassemblaggio controllato del nucleo formato dalla struttura amiloide.

## 4.2 Funzione delle proteine

Le proteine spesso possiedono parti mobili la cui azione meccanica è accoppiata ad un evento chimico che le dona le capacità che sottostanno i processi dinamici della cellula.

### 4.2.1 Tutte le proteine si legano ad altre molecole

Le interazioni fisiche di una proteina con altre molecole determinano le sue proprietà biologiche. Tutte le proteine si legano con altre molecole con legami che possono essere stretti o deboli e temporanei. Il legame ha un alto livello di specificità con una proteina detta ligando. Tale abilità dipende dalla capacità della proteina di formare un insieme di legami non covalenti deboli e interazioni idrofobiche che essendo deboli sono effettivi unicamente quando si formano simultaneamente e pertanto possibili solo se la superficie del ligando si adatta alla proteina. Il sito che si associa con il ligando, detto sito di legame consiste di una cavità nella superficie costituito da un particolare ordinamento di amminoacidi che appartengono a diverse parti della catena che vengono avvicinate dal piegamento. Regioni separate forniscono siti di legame per diversi ligandi permettendo la regolazione dell'attività della proteina. Gli atomi che si trovano all'interno della proteina contribuiscono alla forma della superficie e alle proprietà fisico-chimiche.

### 4.2.2 La conformazione superficiale di una proteina determina la sua chimica

Le capacità chimiche delle proteine richiedono che i gruppi chimici nella superficie interagiscono in modi che aumentino la reattività chimica di una o più catene laterali di amminoacidi. Queste interazioni sono divisibili in due categorie. L'interazione con diverse parti vicine della catena possono impedire l'accesso di molecole d'acqua al sito di legame in quanto possono formare legami a idrogeno che possono competere con i ligandi per i siti. Questo avviene in quanto l'acqua tende a formare legami con altre molecole d'acqua pertanto risulta sfavorevole che una singola molecola si separi dalla rete creata. Il raggrupparsi di catene laterali polari può alterare la loro reattività: se un gran numero di catene laterali cariche negativamente vengono forzate insieme la loro affinità per le cariche

positive aumenta di molto e quando amminoacidi interagiscono tra di loro gruppi solitamente non reattivi come  $-\text{CH}_2\text{OH}$  possono diventare reattivi permettendo la creazione o rottura dei legami covalenti selezionati. La reattività della superficie proteica non dipende pertanto unicamente dalla sequenza di amminoacidi ma anche dal loro orientamento relativo.

#### 4.2.3 La comparazione di sottosequenze tra membri di una famiglia proteica evidenzia i siti di legame cruciali

Il sequenziamento genomico permette di raggruppare molti domini proteici in famiglie che mostrano la loro evoluzione da un antenato comune. I membri di una famiglia sono simili. Si può utilizzare il tracciamento evolutivo per identificare i siti in un dominio proteico che sono fondamentali alla sua funzione. I siti che legano altre molecole sono i più probabili ad essere mantenuti invariati e pertanto gli amminoacidi che non sono cambiati in tutti i membri della famiglia sono mappati su un modello della struttura tridimensionale di un membro della famiglia e le posizioni invarianti formano dei cluster sulla superficie della proteina.

#### 4.2.4 Le proteine si legano con altre proteine attraverso molti tipi di interfacce

Le proteine possono legarsi tra di loro in molti modi. In molti casi una porzione della superficie di una proteina entra in contatto con un anello esteso di una catena polipeptidica con una seconda proteina (interazione superficie-stringa). Un secondo tipo di interfaccia si forma quando due  $\alpha$ -eliche da due proteine si accoppiano formando una coiled-coil, che si trova spesso in famiglie regolatrici dei geni. Il modo più comune di interazione è attraverso la corrispondenza precisa tra due superfici con un'interazione molto stretta data dal gran numero di legami che si formano e molto specifica.

#### 4.2.5 I siti di legame degli anticorpi sono estremamente versatili

Tutte le proteine si devono legare a particolari ligandi per compiere la propria funzione. Gli anticorpi o immunoglobuline sono proteine prodotte dal sistema immunitario in risposta a molecole esterne. Ogni anticorpo si lega strettamente con una molecola obiettivo particolare disattivandola o marcandola per la distruzione. L'anticorpo deve riconoscere l'obiettivo o antigene con grande specificità. Sono molecole a forma di Y con due siti di legame identici complementari a una piccola parte della superficie di una molecola antigenica. Sono formati da molti anelli di catena polipeptidica che protrudono dalla fine di un paio di domini proteici giustapposti. Solo cambiando la lunghezza e gli amminoacidi di questi anelli sono in grado di generare diversi siti di legame per diversi antigeni senza cambiare la struttura di base.

#### 4.2.6 La costante di equilibrio misura la forza di legame

Le molecole che collidono con superfici che si accoppiano male formano pochi legami deboli e si disassociano rapidamente mentre se ne formano troppi l'associazione può persistere nel tempo. Associazioni forti accadono quando una funzione biologica richiede che le molecole rimangano associate per molto tempo. Eventualmente ogni complesso di ligandi e proteine arriva ad un equilibrio in cui i tassi di distruzione e associazione delle molecole si eguaglia. Si può pertanto calcolare la costante di equilibrio  $K$  per misurare la forza dell'associazione ed è anche una misura diretta della differenza di energia libera tra lo stato legato e libero.

#### 4.2.7 Gli enzimi sono catalizzatori potenti e molto specifici

Molte proteine possono svolgere la loro funzione semplicemente legandosi con altre molecole, ma per alcune questo è solo un necessario primo passo, come per gli enzimi. Gli enzimi sono molecole che causano le trasformazioni chimiche che creano e rompono legami covalenti nelle cellule. Legano dei ligandi detti substrati e li convertono in prodotti con grande rapidità. Gli enzimi velocizzano le reazioni di molti ordini di grandezza senza modificarsi: agiscono da catalizzatori. È infatti la catalisi di un insieme di reazioni chimiche da parte degli enzimi che permette alla cellula di rimanere in vita. Gli enzimi si possono raggruppare in classi funzionali che possono svolgere reazioni chimiche simili. Ogni tipo di enzima all'interno della classe è altamente specifico e catalizza un singolo tipo di reazione. Lavorano in gruppo, con il prodotto di un enzima che diventa il substrato per un altro in una rete elaborata di cammini che fornisce alla cellula energia e materiali.

#### 4.2.8 Il legame con il substrato è il primo passo nella catalisi enzimatica

Per un enzima il legame di ogni substrato alla proteina è un inizio necessario. Denotando un enzima  $E$ , il substrato  $S$  e il prodotto  $P$  la reazione più semplice è:



C'è un limite a quanto substrato un enzima può processare nel tempo, nonostante un aumento della concentrazione di enzima questa raggiunge un massimo in cui la molecola di enzima è saturata con il substrato e il tasso della reazione dipende su quanto rapidamente l'enzima può processarlo. Questo tasso massimo è detto numero di turnover, solitamente di 1000 molecole di substrato al secondo per molecola di enzima. L'altro parametro cinetico è  $K_m$ , la concentrazione del substrato che permette la reazione di procedere a metà del tasso massimo. Un valore di  $K_m$  basso vuol dire che l'enzima raggiunge il tasso catalitico massimo ad una bassa concentrazione di substrato e che si legano molto strettamente.

#### 4.2.9 Gli enzimi velocizzano la reazione stabilizzando selettivamente stati transitivi

Gli enzimi raggiungono tassi di reazioni elevatissimi grazie a molte ragioni. In primo luogo quando due molecole devono reagire l'enzima aumenta la concentrazione locale di queste molecole al sito catalitico, mantenendole nell'orientazione corretta. Oltre a questo dell'energia di legame contribuisce alla catalisi. Le molecole di substrato devono passare degli stati intermedi di geometria e distribuzione di elettroni alterata prima di formare il prodotto finale. L'energia libera necessaria per ottenere lo stato più instabile o stato di transizione è detta energia di attivazione ed è il determinante maggiore per il tasso di reazione. Gli enzimi possiedono un maggiore grado di affinità per lo stato di transizione del substrato rispetto alla forma stabile. Siccome il legame stretto riduce l'energia dello stato di transizione, l'enzima accelera la reazione abbassando l'energia di attivazione richiesta.

#### 4.2.10 Gli enzimi possono usare simultaneamente catalizzatori acidi e basici

Gli enzimi non solo si legano strettamente ad uno stato di transizione ma contengono atomi posizionati in modo da alterare la distribuzione degli elettroni che partecipano nella creazione e rottura del legame covalente. I legami polipeptidici possono essere idrolizzati nell'assenza di un enzima

esponedoli a un acido o base forte. Sono gli unici a poter usare catalisi acide o basiche simultaneamente in quanto la struttura rigida impedisce che si combinino. L'adattamento stretto tra enzima e substrato deve essere preciso: un piccolo cambio, anche di  $1\text{\AA}$  può ridurre di l'attività dell'enzima di un migliaio di volte.

##### 4.2.11 Il lisozima illustra come un enzima funziona

Per dimostrare come gli enzimi catalizzano le reazioni chimiche si analizza un enzima antibiotico che si trova in secrezioni come saliva e lacrime. Il lisozima catalizza il taglio di catene polisaccaridi nella parete cellulare dei batteri. Il lisozima catalizza un'idrolisi: aggiunge una molecola d'acqua ad un singolo legame nella catena causandone la rottura. La rottura è energeticamente favorevole in quanto l'energia richiesta per la rottura del legame è minore di quella della catena intatta, ma vi si trova una barriera energetica e la molecola d'acqua che collide solo se il polisaccaride è distorto nello stato di transizione, stato irraggiungibile in condizioni normali. Quando il polisaccaride si lega al lisozima il sito attivo dell'enzima è una lunga fessura che lega sei zuccheri legati alla volta. Appena il polisaccaride si lega forma un complesso e l'enzima lo taglia aggiungendo una molecola d'acqua lungo uno dei legami zucchero-zucchero. Le catene prodotte sono rilasciate, lasciando libero l'enzima per una successiva interazione. Il tasso di idrolisi aumenta perchè le condizioni causate nel sito attivo riducono l'energia di attivazione necessaria distorcendo uno degli zuccheri. Il legame da rompere è tenuto vicino da due amminoacidi acidi (acido glutamico e aspartico) che partecipano nella reazione. Nelle reazioni che coinvolgono due o più reagenti il sito attivo agisce come stampo che unisce i substrati nella corretta orientazione. Passi successivi nella reazione portano le catene laterali nello stato originale permettendo il riutilizzo dell'enzima.

##### 4.2.12 Molecole strettamente legate aggiungono altre funzioni alle proteine

Ci sono molte istanze in cui è necessario aggiungere altre molecole alle proteine affinché svolgano la loro funzione. Tali molecole si trovano nei siti attivi e sono dette coenzimi che sono spesso vitamine o loro derivati. In alcuni casi possono formare legami covalenti che le rendono parte della proteina stessa.

##### 4.2.13 Complessi multienzima aiutano ad aumentare il tasso del metabolismo cellulare

L'efficienza degli enzimi è cruciale per il mantenimento della vita in quanto il tasso di reazioni desiderabili deve essere maggiore del tasso delle reazioni avverse che avvengono naturalmente. Si può misurare il tasso del metabolismo attraverso il consumo di ATP. Una cellula mammifera consuma la propria capacità di ATP ogni paio di minuti, ovvero circa  $10^7$  molecole al secondo. Tale velocità è dovuta all'efficienza degli enzimi, alcuni hanno raggiunto la massima velocità possibile e sono limitati unicamente dalla frequenza delle collisioni (reazione limitata dalla diffusione). La quantità del prodotto di un enzima dipende dalla sua concentrazione e quella del substrato. Se una sequenza deve avvenire rapidamente deve essere presente ogni intermedio metabolico ed enzima in grande concentrazione. Ci sono limiti alla concentrazione massima. La maggior parte dei metaboliti sono presenti in concentrazioni micromolari e gli enzimi con concentrazioni ancora più basse. Il tasso metabolico viene mantenuto pertanto grazie all'organizzazione spaziale della cellula che può aumentare il tasso della reazione spostando enzimi formando grandi proteine dette complessi multienzima. Essendo organizzato in modo che il prodotto di un enzima sia passato direttamente all'enzima successivo la

concentrazione del substrato non deve essere limitante. La maggior parte degli enzimi inoltre hanno evoluto siti di legame con particolari regioni della cellula. Anche i sistemi di membrane intracellulari sono utilizzati a questo scopo in quanto possono sequestrare particolari substrati e i relativi enzimi.

### 4.2.14 La cellula regola le attività catalitiche dei propri enzimi

Una cellula vivente contiene migliaia di enzimi, molti dei quali operano contemporaneamente e vicini tra di loro, generando una complessa rete di cammini metabolici composti da catene di reazioni chimiche. In questi cammini ci sono molti punti in cui diversi enzimi competono per lo stesso substrato. Sono pertanto necessari controlli per regolare quando e quanto rapidamente ogni reazione accade. Questa regolazione avviene a molti livelli: viene controllato il numero di enzimi prodotti regolando l'espressione genica. Gli enzimi vengono inoltre confinati in compartimenti subcellulari come membrane intracellulari o scaffold proteici (o concentrandoli su essi). Gli enzimi possono anche essere modificati covalentemente per disattivarli. Il tasso della distruzione delle proteine è un altro metodo di regolazione, ma il principale è il cambio di attività derivato dal legame con una molecola particolare. Quest'ultimo accade quando un enzima si lega con una molecola (non substrato) a un sito regolatorio speciale, alterando il tasso con cui l'enzima converte il substrato in prodotto. Nel processo di feedback inhibition un prodotto successivo del cammino chimico inibisce un enzima precedente, pertanto quando la reazione produce grandi quantità di prodotto rallenta da sola. Quando i cammini si intersecano o si ramificano ci sono diversi punti di controllo alla fine del cammino che controllano la propria sintesi. Quando la concentrazione di prodotto diminuisce il processo ritorna a velocità normali. Un altro tipo di regolazione è positiva: il prodotto stimola l'azione dell'enzima, questo accade principalmente quando un prodotto di un ramo della rete chimica stimola l'attività di un enzima in un altro cammino.

### 4.2.15 Gli enzimi allosterici possiedono due o più siti di legame che interagiscono

Si nota nelle regolazioni a feedback che la molecola regolatoria possiede una forma diversa rispetto al substrato o all'enzima. Per questo l'effetto sull'enzima è detto allosteria. Gli enzimi che partecipano in questo processo possiedono due siti di legame sulla loro superficie: uno attivo per il legame con il substrato e uno regolatorio per il legame con la molecola regolatoria. Questi due siti comunicano in modo da influenzare gli eventi catalitici. L'interazione è detta cambio conformazionale della proteina: il legame in uno dei siti attivi cambia leggermente la forma della proteina stessa: durante l'inibizione quando la molecola regolatoria si lega cambia la forma del sito attivo incapacitandolo. La maggior parte delle proteine sono allosteriche e possono adottare due o più conformazioni differenti, cambiando tra esse in base a quale ligando si lega. Oltre agli enzimi sono allosterici i recettori, le proteine strutturali e quelle motrici.

### 4.2.16 Due ligandi i cui siti di legame sono accoppiati devono avere effetto reciprocamente sul legame dell'altro

Gli effetti di un legame del ligando segue un principio chimico detto linkage: se una proteina che lega il glucosio lega un'altra molecola su un sito attivo distante, se tale sito attivo cambia forma a causa del cambio di conformazione causato dal glucosio i due siti sono detti accoppiati. Quando due ligandi preferiscono il legame con la stessa conformazione di una proteina allosterica ogni ligando deve aumentare l'affinità dell'altro per la proteina. In maniera inversa il linkage opera negativamente

quando i due ligandi si vogliono legare a conformazioni diverse della molecola e devono competere per il legame.

#### 4.2.17 Insiemi di proteine simmetrici producono transizioni allosteriche cooperative

Un singolo feedback negativo non basta per l'ottimale regolazione della cellula e pertanto la maggior parte degli enzimi che partecipano nel feedback sono insiemi simmetrici di subunità identiche. Con questo ordinamento il legame di una molecola ad un ligando su un sito su una subunità promuove un cambio allosterico nell'intero insieme (i legami con i ligandi delle altre subunità avviene più facilmente). Avviene pertanto una transizione allosterica cooperativa che permette un cambio nella concentrazione dei ligandi per cambiare l'insieme da uno stato attivo a uno inattivo. Questo principio è comune a proteine che non sono enzimi.

#### 4.2.18 Molti cambi nelle proteine sono guidati dalla fosforilazione proteica

Le proteine sono regolate anche dall'addizione di un gruppo fosfato. Un evento di fosforilazione può influenzare la proteina: può cambiare la conformazione attraendo un insieme di amminoacidi carichi positivamente e cambiando la forma del sito attivo. Quando un secondo enzima rimuove il gruppo fosfato la proteina ritorna alla conformazione originale. In secondo luogo il gruppo fosfato può formare parte di una struttura che altre proteine riconoscono. Infine la sua addizione può mascherare un sito di legame che tiene unite due proteine, rompendo il legame. La fosforilazione proteica reversibile controlla l'attività, struttura, e localizzazione di enzimi e altre proteine.

#### 4.2.19 Una cellula eucariotica contiene un gran numero di proteina chinasi e proteina fosfatasi

La fosforilazione di una proteina coinvolge il trasferimento catalizzato dalla proteina chinasi del gruppo fosfato terminale di una molecola di ATP al gruppo idrossile di un amminoacido serina, treonina o tirosina. Tale reazione è tipicamente unidirezionale a causa della grande quantità di energia libera rilasciata quando viene rotto il legame fosfato-fosfato nell'ATP. La proteina fosfatasi catalizza l'azione inversa di rimozione del gruppo fosfato o defosforilazione. Esistono un gran numero di questi enzimi, ognuno responsabile di poche proteine o specifici substrati da subunità regolatori. Lo stato di fosforilazione dipende dall'attività relativa della proteina chinasi e fosfatasi che la modificano. Questi enzimi fanno parte di una grande famiglia di enzimi che condivide una sequenza catalitica di circa 290 amminoacidi. Le differenze permettono la specificità.

#### 4.2.20 La regolazione della proteina chinasi Src rivela come una proteina può funzionare come un microprocessore

Le migliaia di diverse proteine chinasi sono organizzate in una rete complessa di cammini segnalatori che aiutano a coordinare le attività della cellula, guidare il ciclo cellulare e ritrasmettere segnali nella cellula dall'esterno. Proteine chinasi individuali svolgono il ruolo di dispositivi di input/output nel processo di integrazione. Una parte importante dell'input a queste proteine di processamento dei segnali viene dal controllo eseguito dai fosfati aggiunti e rimossi da essi dalle proteine chinasi e fosfatasi. La famiglia Src delle proteine chinasi hanno questo comportamento e contengono una corta regione N<sup>+</sup> terminale che si lega covalentemente a un acido grasso fortemente idrofobico che

la ancora alla faccia citoplasmatica della membrana plasmatica. Dopo il gruppo terminale si trovano due domini SH3 e SH2 seguiti da il dominio catalitico della chinasi. Normalmente esistono nella versione inattiva, ma transizionano in quella attiva quando viene rimosso il fosfato C<sup>-</sup> terminale al legame del dominio SH3 da una proteina di attivazione. QUando accade l'attivazione segnala il completamento di un insieme di eventi.

#### 4.2.21 Le proteine che si legano e regolano il GTP sono onnipresenti regolatori

Il controllo dell'attività di una cellula attraverso addizione e rimozione di un gruppo fosfato avviene anche quando non viene direttamente attaccato a essa ma è parte del nucleotide guanina GTP che si lega molto strettamente con un gruppo di proteine dette GTP-leganti. Queste proteine si trovano nella conformazione attiva quando formano il legame con il GTP. La perdita del gruppo fosfato avviene quando il legame GTP è idrolizzato in GDP da una reazione catalizzata dalla proteina stessa. Tali proteine, dette anche GTPasi costituiscono una grande famiglia di proteina con variazioni dello stesso dominio globulare dove avviene il legame con il GTP. Quando il GTP legato si idrolizza il dominio cambia conformazione e disattiva la proteina.

#### 4.2.22 Le proteine regolatorie GAP e GEF controllano l'attività delle proteine GTP-leganti determinando se è legato GTP o GDP

Le proteine GTP-leganti sono controllate da proteine regolatorie che determinano se a esse è legato GTP o GDP. Sono disattivate da una proteina GTPasi-attivante che si lega a essa e idrolizza il GTP legato in GDP che rimane legato e fosfato inorganico che viene rilasciato. La proteina rimane in forma inattiva fino a che non incontra un fattore di scambio di nucleotide guanina (GEF) che si lega alla proteina con il GDP e causa il suo rilascio. Il sito è immediatamente riempito da una molecola di GDP, presenti in eccesso. Il GEF attiva la proteina aggiungendo il gruppo fosfato indirettamente.

#### 4.2.23 Le proteine possono essere regolate dall'aggiunta covalente di altre proteine

Le cellule contengono una famiglia speciale di piccole proteine i cui membri sono attaccati a molte altre proteine in modo da determinarne l'attività. In ogni caso la loro estremità carbossilica si lega all'amminoacido lisina di una proteina obiettivo attraverso un legame isopeptidico. Questa addizione porta ad un cambio di conformazione reversibile (attraverso reazioni catalizzate da enzimi). La proteina più comune di questo tipo è la ubiquitina che si lega alle proteine in vari modi, ognuno con un significato diverso. Può formare catene di poliubiquitina con diverse modifiche alle funzioni della proteina che legano.

#### 4.2.24 Un elaborato sistema di coniugazione di ubiquitina è utilizzato per marcare le proteine

Per selezionare le proteine target per l'addizione di ubiquitina si deve attivare la terminazione carbossilica di quest'ultima attraverso un enzima ubiquitina-attivante che utilizza l'idrolisi dell'ATP per ricavare l'energia necessaria ad attaccare l'ubiquitina con se stessa attraverso un legame covalente ad alta energia. La proteina passa a enzimi di coniugazione di ubiquitina che la uniscono con le ubiquitina ligasi che si legano a speciali segnali di degradazione detti degrons nei substrati aiutando il secondo insieme di enzimi a formare una catena di poliubiquitina collegata ad una lisina. La catena

sulla proteina obiettivo viene riconosciuta da un particolare recettore nel proteosoma, causano la distruzione di essa.

##### 4.2.25 I complessi proteici con parti interscambiabili fanno un efficiente uso dell'informazione genica

Il SCF ubiquitina ligasi è un complesso proteico che lega proteine obiettivo aggiungendogli una catena di poliubiquitina. Ha una struttura a forma di C composta da cinque subunità proteiche, la più grande serve come impalcatura per le altre. Ad un'estremità si trova il secondo enzima del processo di coniugazione di ubiquitina, all'altra parte un braccio che lega il substrato o proteina F-box. Quando il complesso è attivato F-box lega il sito attivo alla proteina obiettivo e la sposta nel centro in modo che sue lisine entrino in contatto con la poliubiquitina. Catalizza poi una ripetuta addizione di polipeptidi ubiquitina alle lisine marcando la proteina per distruzione rapida nel proteosoma. Specifiche proteine sono marcate per distruzione in risposta a segnali specifici, aiutando a guidare il ciclo della cellula. Il tempismo della distruzione richiede la creazione di specifici pattern di fosforilazione della proteina richiesto per il riconoscimento dalla subunità F-box e l'attivazione dell'ubiquitina ligasi SCF. Molti dei bracci che legano il substrato sono interscambiabili nel complesso proteico e richiedono più di 70 geni. Esistono molti tipi di proteine F-box e una famiglia di SCF-like ubiquitina ligasi. Una proteina con parti interscambiabili fa uso economico dell'informazione genetica nella cellula.

##### 4.2.26 Una proteina GTP-legante mostra come grandi movimenti proteici siano possibili

La proteina EF-Tu fornisce un buon esempio di come cambi allosterici nella conformazione della proteina producono movimenti amplificando un cambio conformazionale locale. Tale proteina è un fattore di allungamento nella sintesi proteica, caricando ogni amminoacil-tRNA sul ribosoma in quanto quest'ultimo forma un forte complesso con la forma legata al GTP. Questa molecola di tRNA può trasferire il suo amminoacido alla catena polipeptidica crescente solo dopo che il GTP è idrolizzato. Essendo che tale reazione è fatta partire da un giusto adattamento alla molecola di mRNA nel ribosoma EF-Tu serve come fattore discriminante tra coppie errate di mRNA e tRNA. Confrontando la conformazione della proteina nelle forme legate a GTP e GDP si osserva come il riposizionamento del tRNA avviene. La dissociazione del gruppo fosfato causa un cambio nel sito che lega il GTP che causa una propagazione di cambio di movimento lungo un  $\alpha$ -elica detta switch helix che serve come aggancio che aderisce a un sito specifico in un altro dominio della molecola, mantenendola in una conformazione chiusa. Il cambio conformazionale causa la separazione della switch helix permettendo a domini separati di aprirsi rilasciando la molecola di tRNA permettendo l'utilizzo dell'amminoacido.

##### 4.2.27 Le proteine motrici producono grandi movimenti nelle cellule

Lo scopo principale delle proteine motrici è muovere altre molecole. Sono responsabili della contrazione muscolare e dei movimenti delle cellule, oltre a permettere movimenti intracellulari come il movimento dei cromosomi in parte opposte della cellula durante la mitosi, per muovere gli organelli e gli enzimi con strand di DNA durante la sintesi del DNA. Senza nessuna guida questi movimenti sono reversibili e la proteina si sposterebbe casualmente lungo un filamento. Per rendere i cambi conformazionali, e pertanto i movimenti, unidirezionali si deve accoppiare il movimento con l'idrolisi di molecole di ATP legate alla proteina. Essendo che l'idrolisi dell'ATP rilascia molta energia



è molto improbabile che avvenga il movimento inverso. Il legame di ATP sposta la proteina in una conformazione e l'idrolizzazione cambia la conformazione e il rilascio dell'ADP la cambia in una terza. L'energia rilasciata dall'idrolisi rende i cambi conformazionali irreversibili e il ciclo va in una direzione, spostando sempre la proteina lungo la stessa direzione.

#### **4.2.28 I trasportatori di membrana imbrigliano energia per pompare le molecole attraverso le membrane**

Le proteine allosteriche possono imbrigliare l'energia derivata dall'idrolisi dell'ATP, dal gradiente ionico o dal trasporto di elettroni. I trasportatori ACO costituiscono una classe di proteine pompe legate alla membrana. La loro funzione principale è trasportare molecole idrofobiche dal citoplasma, rimuovendo molecole tossiche. Tali trasportatori contengono un paio di subunità che attraversano la membrana legate a una coppia di subunità che legano ATP sotto la membrana plasmatica. L'idrolisi dell'ATP guida un cambio conformazionale trasmettendo una forza che causa il movimento da parte delle subunità legate alla membrana della molecola legata attraverso il doppiostrato lipidico. Alcune di queste proteine sono pompe rotanti che accoppiano l'idrolisi dell'ATP con il trasporto di ioni  $H^+$  e sono utilizzate per acidificare l'interno del lisozoma. Possono funzionare al contrario catalizzando la reazione di sintesi dell'ATP se il gradiente è abbastanza ripido.

#### **4.2.29 Le proteine formano grandi complessi che funzionano come macchinari proteici**

Grandi proteine possono svolgere funzioni elaborate. La maggior parte delle reazioni sono catalizzate da un insieme altamente coordinato e collegato di queste macchine proteiche che utilizzano una reazione energeticamente favorevole come l'idrolizzazione dell'ATP per causare una serie di cambi conformazionali che causano il movimento coordinato dell'insieme.

#### **4.2.30 Impalcature concentrano insiemi di proteine che interagiscono**

Le proteine sono localizzate in specifici siti della cellula e sono assemblate e attivate solo quando necessario. Questo meccanismo coinvolge impalcature proteiche, proteine con multipli siti di legame con altre proteine che servono per unire proteine interagenti e per posizionarle in parti specifiche della cellula.

#### **4.2.31 Molte proteine sono controllate da modifiche covalenti che le dirigono verso siti specifici all'interno della cellula**

Un gran numero di proteine sono modificate su più amminoacidi con eventi regolatori producendo pattern diversi. Queste modifiche covalenti possono essere considerate come un codice regolatorio combinatorio: gruppi modificanti specifici sono aggiunti o rimossi in risposta a segnali cambiando l'attività e la stabilità della proteina, pertanto la molecola è in grado di rispondere velocemente e con grande versatilità a cambi nella sua condizione o ambiente.

#### **4.2.32 Una complessa rete di interazioni proteiche sottosta alle funzioni della cellula**

Per comprendere le funzioni di ogni complesso proteico si deve ricostituirlo dalle parti proteiche purificate in modo da poter studiare dettagliatamente i modi di operazioni in condizioni controllate.

Esistono un gran numero di interazioni proteina-proteina e possono essere studiate attraverso mappe di interazione tra proteine, utili per identificare la funzione probabile di proteine non caratterizzate in base alla posizione relativa alle altre. Queste reti devono essere analizzate con cura perchè la stessa proteina può essere utilizzata in complessi diversi e avere diverse funzioni. Nei confronti incrociati le proteine con pattern di interazione simile probabilmente hanno la stessa funzione della cellula.

## Parte II

# Meccanismi genetici basici

## Capitolo 5

# DNA, cromosomi e genomi

La vita dipende dall'abilità della cellula di conservare, recuperare e tradurre le istruzioni genetiche richieste per creare e mantenere un organismo vivente. Queste informazioni ereditarie sono passate da una cellula a una sua figlia durante la divisione cellulare e da una generazione di organismi all'altra attraverso le cellule riproduttive, sono salvate come geni. Le informazioni genetiche consistono principalmente su istruzioni per la costruzione delle proteine, macromolecole versatili che svolgono la maggior parte delle funzioni della cellula. Le informazioni genetiche sono trasportate su cromosomi, strutture a filo nel nucleo che diventano visibili durante la divisione e composti di DNA (acido desossiribonucleico) e proteine in egual misura. La determinazione della struttura a doppia elica del DNA ha risolto il problema di come le informazioni nel DNA sono replicate e come una molecola di DNA utilizza la sequenza dei suoi monomeri per produrre le proteine.

### 5.1 La struttura e la funzione del DNA

Negli anni 50 la struttura del DNA è stata determinata grazie a diffrazione a raggi X, che indicarono la composizione a due fili avvolti in un'elica fornendo un grande indizio a Watson-Crick e al loro modello.

#### 5.1.1 Una molecola di DNA consiste di due catene di nucleotidi complementari

Una molecola di acido desossiribonucleico consiste di due catene polinucleotide lunghe note come strand e composte da quattro tipi di subunità. Le catene sono antiparallele tra di loro e tra le basi dei nucleotidi si formano legami a idrogeno che le tengono unite. I nucleotidi sono composti da zuccheri a 5 atomi di carbonio a cui sono attaccati dei gruppi fosfati e una base contenente dell'azoto. Nel caso dei nucleotidi del DNA lo zucchero è il desossiribosio attaccato a un singolo gruppo fosfato (da cui il nome) e la base può essere adenina (A), citosina (C), guanina (G) o timina (T). I nucleotidi sono legati covalentemente in una catena attraverso gli zuccheri e i fosfati che formano un backbone di legami zucchero-fosfato-zucchero-fosfato alternati. I legami dei nucleotidi danno al filo del DNA una polarità chimica. Il gruppo 5' fosfato si lega con il gruppo 3' idrossile di un altro monomero e pertanto tutte le subunità hanno lo stesso orientamento. Inoltre le due terminazioni dello strand sono facilmente distinguibili e ci si riferisce ad esse come alla terminazione 3' e 5', indicando la polarità. Grazie alla direzionalità e linearità del filo di DNA può essere letto facilmente. La doppia elica si genera dalla struttura chimica delle catene polinucleotidiche in quanto sono mantenute insieme

da legami a idrogeno tra le basi sui fili diversi tutte le basi si trovano all'interno della doppia elica con i backbones verso l'esterno. In ogni caso una base a due anelli (una purina) è sempre legata con una ad anello singolo (una pirimidina): A si accoppia con T e G con C. Questo accoppiamento complementare delle basi permette a coppie di basi di essere ammassate nell'ordinamento più energeticamente favorevole all'interno della doppia elica in quanto ogni coppia possiede la stessa lunghezza mantenendo i backbones a distanza costante. Per massimizzare l'efficienza spaziale i backbones si avvolgono formando una doppia elica destra con un giro completo ogni 10 paia. I membri di ogni base possono adattarsi insieme solo se i fili sono antiparalleli. Ogni strand della molecola contiene una sequenza complementare a quella dell'altro strand.

### 5.1.2 La struttura del DNA mette a disposizione un meccanismo per l'ereditarietà

La struttura a polimero lineare composto da quattro monomeri permette al DNA di trasportare informazioni in forma chimica, mentre la natura a doppio filamento permette la sua duplicazione. Essendo ogni filamento complementare all'altro può agire come stampo per la sintesi di un nuovo filamento complementare. Nominando i filamenti  $S$  e  $S'$ ,  $S$  agisce da stampo per la formazione di  $S'$  e  $S'$  per  $S$ , permettendo l'accurata copia dell'informazione nel DNA. Questa capacità permette alla cellula di replicare il proprio genoma e passarlo ai discendenti. Il DNA permette la codifica di proteine (le cui funzioni sono determinate dalla forma e pertanto in ultimo dalla sequenza di amminoacidi) attraverso una corrispondenza esatta tra i quattro nucleotidi e i venti amminoacidi. L'intero gruppo di informazioni mantenute dal DNA di un organismo è detto genoma e specifica tutte le molecole di RNA e proteine che l'organismo sintetizzerà.

### 5.1.3 Negli eucarioti il DNA è racchiuso nel nucleo cellulare

Circa tutti il DNA della cellula si trova isolato nel nucleo, che occupa circa il 10% del volume totale, delimitato da un involucro nucleare formato da due bistrati lipidici concentrici. Queste membrane sono punte a intervalli da pori nucleari in cui le molecole si trasferiscono dal nucleo al citosol. Tale involucro è connesso a un sistema di membrane intracellulari detto reticolo endoplasmatico che si estende nel citoplasma ed è meccanicamente supportato da una rete di filamenti intermedi detti lamine nucleari. L'involucro permette alle proteine che agiscono sul DNA di essere concentrate dove necessario e mantiene gli enzimi nucleari e citosolici separati.

## 5.2 DNA cromosomico e il suo confezionamento nella fibra cromatina

La funzione più importante del DNA è quella di trasportare i geni, informazioni che specificano tutte le molecole di RNA e le proteine che compongono l'organismo. Il DNA nucleare degli eucarioti è diviso in cromosomi.

### 5.2.1 Il DNA eucariotico è confezionato in un insieme di cromosomi

Ogni cromosoma consiste di una singola molecola lineare di DNA estremamente lunga associata con proteine che piegano e impacchettano il filamento in una struttura compatta. Oltre a queste proteine si trovano altre proteine e molecole di RNA che si occupano dell'espressione genica e della riparazione e replicazione del DNA. Il complesso formato dal DNA e le proteine strettamente legate prende il

nome di cromatina. Nei batteri d'altra parte il DNA si trova in forma circolare. Con l'eccezione dei gameti, alcune cellule specializzate che non possono moltiplicare o a cui manca il DNA e altre che replicano il loro DNA senza completare la divisione cellulare ogni nucleo cellulare umano contiene due coppie di ogni cromosoma, uno ereditato dal padre e uno dalla madre. I cromosomi di tale coppia sono detti omologhi, l'unica coppia non omologa nei maschi è quella dei cromosomi Y (dal padre) e X (dalla madre). Ogni uomo contiene 46 cromosomi: 22 paia comuni tra i sessi e una coppia detta dei cromosomi sessuali che possono essere distinti attraverso colorazione basata sull'ibridazione del DNA in cui un piccolo filamento marcato con un colore fluorescente agisce da sonda che si lega alla sequenza complementare illuminando il cromosoma obiettivo. L'insieme dei 46 cromosomi durante la mitosi è detto cariotipo.

### 5.2.2 I cromosomi contengono lunghe stringhe di geni

I cromosomi trasportano geni, definiti come segmenti di DNA che contengono le istruzioni per sintetizzare una particolare proteina (o una famiglia) o molecole di RNA significativamente funzionali. Oltre ai geni il genoma di organismi multicellulari e di molti altri eucarioti contiene grandi segmenti sparsi la cui funzione non è compresa. Alcuni di questi segmenti sono fondamentali per la regolazione dell'espressione genica. Questi segmenti generano grandi diversità nella dimensione genomica quando si comparano le specie e creano sequenze molto diverse per organismi strettamente imparentati, anche se contengono gli stessi geni. La divisione del genoma in cromosomi è unico per la specie, con numero e dimensione dei cromosomi diversa.

### 5.2.3 La sequenza nucleotidica del genoma umano mostra come i geni sono ordinati

Si nota come solo una piccola percentuale del genoma umano codifica le proteine e circa metà è composta da pezzi mobili che si sono inseriti nei cromosomi durante l'evoluzione. La dimensione media di un gene è circa 27000 coppie di basi. Siccome solo circa 1300 paia sono richieste per codificare una proteina di dimensione media (430 amminoacidi negli umani) il resto della sequenza consiste di lunghezze di DNA non codificante che interrompe quello codificante. Le sequenze codificanti sono dette esoni, mentre quelle non codificanti introni. Il DNA umano consiste di una lunga stringa di esoni e introni (che sono la maggioranza) che si alternano. Oltre a introni e esoni ogni gene è associato con sequenze di DNA regolatorio responsabile che il gene sia espresso al momento e a livello giusto. Negli esseri umani occupano decine di migliaia di coppie di basi. Oltre ai geni codificanti le proteine ne esistono altri che codificano molecole di RNA con funzione propria.

### 5.2.4 Ogni molecola di DNA che forma un cromosoma lineare deve contenere un centromero, due telomeri e un'origine di replicazione

Per formare un cromosoma funzionale una molecola di DNA deve poter essere abile di replicarsi e i replicati devono essere separati e partizionati in cellule figlie. Questo processo accade attraverso una serie ordinata di passaggi detti ciclo cellulare che fornisce la separazione temporale tra la duplicazione dei cromosomi e la segregazione in due cellule figlie. Durante una lunga interfase i geni sono espressi e i cromosomi sono replicati rimanendo in un paio di cromatidi sorelle. In questo momento i cromosomi sono estesi in modo che la maggior parte della cromatina esista come lunghi fili e i cromosomi individuali non siano distinguibili. In un momento successivo si condensano in modo da separare le cromatidi sorelle e distribuiti alle cellule figlie. I cromosomi altamente condensati sono dette cromosomi mitotici. Ogni cromosoma opera come un'unità strutturale indipendente. Le

operazioni di replicazione sono controllate da tre tipi di sequenze nucleotidiche nel DNA, ognuna delle quali si lega a proteine specifiche che guidano la macchina che replica e segrega i cromosomi. Un tipo di sequenza nucleotidica agisce come origine della replicazione del DNA, ovvero il luogo dove la duplicazione inizia, ne esistono diverse per assicurarsi che l'intero cromosoma venga replicato rapidamente. Dopo la replicazione del DNA le due cromatidi sorelle rimangono attaccate a una fine mentre vengono condensate producendo cromosomi mitotici. Un'ulteriore sequenza specializzata detta centromero permette a una copia di ogni cromosoma duplicato e condensato di essere portata nella cellula figlia quando si divide. Un complesso proteico detto cinetocoro si forma al centromero e attacca i cromosomi al mandrino mitotico, permettendo la loro separazione. La terza sequenza di DNA specializzata forma i telomeri, le terminazioni dei cromosomi. Permettono una replicazione efficiente e formano strutture che proteggono la fine del cromosoma da essere confusa con una molecola di DNA in necessità di riparazione.

### 5.2.5 Le molecole di DNA sono altamente conservate nei cromosomi

Tutti gli organismi eucariotici possiedono metodi per impacchettare il DNA in cromosomi. Questa compressione viene svolta da proteine che avvolgono e piegano il DNA in livelli sempre più alti di organizzazione. Nonostante molto meno condensati rispetto a cromosomi mitotici sono comunque compressi. La struttura dei cromosomi è dinamica: regioni specifiche del interfase dei cromosomi decondensano per permettere l'accesso di sequenze specifiche del DNA per l'espressione genica, riparazione e replicazione ricondensandosi dopo. La condensazione avviene pertanto in modo da permettere accesso rapido e localizzato al DNA su richiesta.

### 5.2.6 I nucleosomi sono unità base della struttura del cromosoma eucariote

Le proteine che si legano al DNA per formare i cromosomi sono divise in istoni e proteine non-istoniche che contribuiscono con la stessa massa ad un cromosoma. Il complesso di entrambe le classi di proteine è noto come cromatina. Gli istoni sono responsabili per il primo e basilico livello di condensazione cromosomica detto nucleosoma, un complesso DNA-proteina. Si nota come il nucleosoma sia formato da una stringa di DNA con parti di esso avvolte intorno a istone. Ogni nucleo istonico del nucleosoma consiste di un complesso di 8 proteine istoniche e DNA a doppio filamento lungo 147 coppie di basi. Questo ottamero forma un nucleo proteico lungo il quale si lega. La regione di DNA linker che separa ogni nucleo di nucleosoma può essere lunga fino a 80 basi. In media i nucleosomi si ripetono ogni circa 200 coppie di basi.

### 5.2.7 La struttura del nucleo del nucleosoma rivela come il DNA è condensato

La forma a disco del nucleo istonico in cui il DNA è avvolto in una bobina sinistra di 1.7 giri è composta da quattro istoni, proteine che condividono una piega istonica formata da tre  $\alpha$ -eliche connesse da due anelli. Durante l'assemblamento del nucleosoma le pieghe istoniche si legano prima tra di loro formando dimeri, poi tetrameri e infine ottameri. L'interfaccia tra il DNA e l'istone forma 142 legami a idrogeno in ogni nucleosoma. Circa metà di questi legami derivano dal backbone amminoacido e quello del DNA, le interazioni idrofobiche e saline mantengono il DNA e la proteina insieme nel nucleosoma. Un quinto del nucleo istonico è formato da lisina o da arginina (amminoacidi basici) e le loro cariche positive possono neutralizzare il backbone del DNA. Queste interazioni spiegano perché il legame con l'istone è indipendente dalla sequenza delle basi. Oltre alla piega

istonica ogni nucleo istonico possiede una coda N-terminale che si estende all'esterno del nucleo DNA-istone che subisce cambi covalenti che controllano aspetti della struttura cromatina e della funzione. Gli istoni sono proteine altamente conservate.

#### **5.2.8 I nucleosomi hanno strutture dinamiche e sono soggetti a cambi catalizzati da complessi rimodellanti a cromatina ATP-dipendenti**

Il DNA in un nucleosoma isolato si disavvolge quattro volte al secondo, rimanendo esposto per intervalli di tempo, rendendolo disponibile per il legame con altre proteine. Dalla cromatina nella cellula è richiesto un ulteriore allentamento è necessario in quanto le cellule contengono una varietà di complessi di rimodellamento a cromatina ATP-dipendenti. Questi complessi includono una subunità che idrolizza ATP che lega il nucleo proteico del nucleosoma e il DNA legato a esso. Utilizzando l'energia fornita dall'ATP muove il DNA cambiando la struttura del nucleosoma temporaneamente allentando il legame tra il DNA e il nucleo istonico. Attraverso ripetuti cicli di idrolisi dell'ATP il complesso può catalizzare uno scorrimento del nucleosoma in modo da riposizionarlo esponendo specifiche regioni di DNA rendendole disponibili ad altre proteine. Inoltre cooperando con altre altre proteine che si legano agli istoni e servono come accompagnatrici di istoni questi complessi sono capaci di rimuovere parte o tutto il nucleo del nucleosoma. Un tipico nucleosoma è sostituito ogni ora o due all'interno della cellula. Questi complessi esistono in molte varianti con ruoli diversi. La maggior parte sono proteine con 10 o più subunità altri creano specifici cambi sugli istoni. Quando i geni sono attivati o disattivati questi complessi vengono portati a regioni specifiche del DNA dove agiscono sulla struttura cromatinica. La maggior influenza sul posizionamento del nucleosoma sembra essere altre proteine strettamente legate al DNA, che possono favorire la presenza di un nucleosoma adiacente a loro o forzare il suo allontanamento.

#### **5.2.9 I nucleosomi sono tipicamente condensati in una fibra cromatinica**

I nucleosomi sono tipicamente condensati uno sull'altro, generando vettori in cui il DNA è altamente denso, pertanto la cromatina forma una fibra. Questa vicinanza è causata da legami tra nucleosomi che coinvolgono la coda degli istoni e un istone addizionale che svolge la funzione di linker (H1) che si lega a ogni nucleosoma e cambia il percorso del DNA quando esce dal nucleosoma.

### **5.3 La struttura e la funzione della cromatina**

I meccanismi che creano le strutture cromatiniche in diverse regioni della cellula hanno diverse funzioni e possono essere ereditate attraverso ereditarietà epigenetica, ovvero ereditarietà superimposta sull'ereditarietà genetica del DNA.

#### **5.3.1 L'eterocromatina è altamente organizzata e limita l'espressione genica**

Esistono due tipi di cromatina nell'interfase dei nuclei di molte cellule eucariotiche: una forma altamente condensata detta eterocromatina e una meno condensata detta eucromatina. L'eterocromatina è altamente concentrata in certe regioni specializzate: i centromeri e nei telomeri e presente in luoghi che variano in base allo stato fisiologico della cellula. Contiene tipicamente pochi geni e quando parti eucromatiche vengono convertite in eterocromatiche i rispettivi geni vengono disattivati. Descrive domini di cromatina compatti che sono resistenti all'espressione genica.



#### 5.3.2 Lo stato eterocromatico è autopropagante

Attraverso la rottura e la riunione dei cromosomi una parte eucromatica può essere trasformata in una eterocromatica in un effetto detto di posizione. Riflette l'allargamento dello stato eterocromatico ed è il modo in cui l'eterocromatina è creata e mantenuta. QUando una condizione eterocromatica si stabilisce in una zona viene ereditata dalle cellule figlie nella variegazione dell'effetto di posizione. Si nota pertanto come l'eterocromatina genera sè stessa in un feedback positivo che si espande spazialmente e temporalmente. Esistono geni che aumentano o diminuiscono questo processo che codificano proteine cromosomali non istoni che interagiscono con gli istoni coinvolti nella modifica o mantenimento della struttura cromatinica.

#### 5.3.3 I nuclei istonici sono modificati covalentemente a molti siti diversi

Le catene laterali degli amminoacidi dei quattro istoni del nucleosoma sono soggette a molti cambi covalenti, come l'acetilazione della lisina, la mono-, di- e trimetilazione delle lisine e la fosforilazione delle serine. Un gran numero di queste modifiche avvengono sulle code istoniche N<sup>+</sup> terminali che protrudono dal nucleosoma. Sono tutte reversibili e catalizzate da enzimi altamente specifici. Ogni enzima è reclutato su siti specifici della cromatina in tempi determinati in base alle proteine regolatorie della trascrizione o fattori di trascrizione che riconoscono e si legano a specifiche sequenze di DNA. Sono prodotte in tempi e luoghi diversi, determinando dove e quando tali enzimi agiscono. Il DNA determina pertanto come gli istoni sono modificati, ma in alcuni casi le modifiche covalenti possono persistere creando una memoria della storia di sviluppo della cellula che può essere trasmessa ereditariamente. Differenti gruppi di nucleosomi presentano diverse modifiche, che vengono controllate e hanno importanti conseguenze. L'acetilazione delle lisine allenta la struttura cromatinica in quanto il gruppo acetile rimuove la carica positiva della lisina, riducendone l'affinità con le code dei nucleosomi adiacenti, ma gli effetti più importanti sono generati dall'abilità di reclutare altre proteine in queste lunghezze di cromatina alterata. La trimetilazione di una specifica lisina sulla coda H3 attrae la proteina HP1 eterocromatina-specifica contribuendo alla stabilizzazione ed espansione della cromatina. Le proteine reclutate agiscono con gli istoni modificati determinando dove e quando i geni saranno espressi. Si nota pertanto come la struttura di ogni dominio cromatinico governa la lettura delle informazioni genetiche e infine struttura e funzioni della cellula.

#### 5.3.4 La cromatina acquisisce varietà addizionale attraverso inserzioni specifiche al sito di un insieme di varianti di istoni

Nella cellula sono contenuti alcune varianti di istoni che si possono assemblare in nucleosomi. Gli istoni principali sono sintetizzati primariamente durante la fase S del ciclo cellulare e si assemblano in nucleosomi sulle eliche di DNA figlie dietro la forcella di replicazione, mentre le varianti sono sintetizzate nell'interfase e sono inseriti in cromatine già formate con un processo di cambio istonico catalizzato dal complesso di rimodellazione della cromatina ATP-dipendente che contiene subunità che causano il suo legame a siti specifici della cromatina e a accompagnatori di istoni che trasportano una particolare variante, inserendoli pertanto in luoghi con alta specificità.

#### 5.3.5 Le modifiche covalenti e le varianti istoniche agiscono insieme per controllare le funzioni dei cromosomi

Le modifiche degli istoni avvengono in insiemi ordinati. Alcune combinazioni determinano come e dove il DNA condensato possa essere acceduto e manipolato, creando un insieme di marcature che

determinano se la lunghezza di cromatina è stata appena replicata, o è danneggiata o come e dove l'espressione genica deve accadere. Molte proteine regolatorie contengono piccoli domini che si legano a marcature specifiche come la lisina trimetilata sull'istone H3. Questi domini sono legati insieme in una grande proteina o complesso che riconosce una combinazione specifica di modifiche all'istone. Si crea pertanto un complesso lettore che permette a particolari combinazioni di marcature di attrarre proteine per eseguire le funzioni biologiche nel momento corretto. Queste marcature sono dinamiche e si creano sulle code istoniche che si trovano sull'esterno del nucleo del nucleosoma.

#### **5.3.6 Un complesso di proteine lettrici e scrittrici possono espandere modifiche alla cromatina lungo un cromosoma**

Il fenomeno dell'effetto di posizione richiede che forme modificate di cromatina abbiano l'abilità di espandersi per distanze sostanziale lungo il cromosoma. Si nota come gli enzimi che aggiungono o rimuovono modifiche agli istoni fanno parte di complessi che possono essere trasportati dai regolatori di trascrizione su lunghezze specifiche di DNA, ma dopo che l'enzima marca i nucleosomi può avvenire una reazione a catena in cui lavora insieme a una proteina di lettura nello stesso complesso che riconosce la marcatura e si lega a tale nucleosoma, attivando l'enzima di scrittura e posizionandolo su un nucleosoma vicino. Attraverso i cicli di lettura-scrittura la proteina può trasportare l'enzima di scrittura lungo il DNA espandendo la marcatura lungo il cromosoma. Un processo simile avviene per rimuovere le modifiche agli istoni.

#### **5.3.7 Barriere nella sequenza di DNA bloccano l'espansione dei complessi di lettura-scrittura separando domini di cromatina vicini**

Nella cellula esistono sequenze di DNA che marcano i confini di domini di cromatina e li separano. Questa sequenza di barriera contiene un insieme di siti di legame per gli enzimi istoni acetilasi. Siccome la lisina acetilata è incompatibile con la metilazione richiesta per l'espansione dell'eterocromatina si blocca la sua espansione. Esistono comunque molte altre modifiche alla cromatina che raggiungono questo scopo.

#### **5.3.8 La cromatina nei centromeri rivela come le varianti degli istoni possono creare strutture speciali**

Le varianti di istoni che trasportano nucleosomi possono produrre marcature nella cromatina persistenti. Un esempio di questo è la formazione e ereditarietà di una struttura di cromatina specializzata al centromero, la regione necessaria per l'attacco al mandrino mitotico e alla segregazione ordinata delle copie del genoma durante la divisione cellulare. Ogni centromero è contenuto da una cromatina centromerica che persiste durante l'interfase che contiene una variante dell'istone H3 nota come CENP-A (centromere protein-A) e altre proteine che condensano il nucleosoma in ordinamenti particolarmente densi e formano il cinetocoro, la struttura richiesta per l'attacco al mandrino mitotico. I centromeri consistono di sequenze di DNA corte e ripetute dette DNA alfa satellite. La presenza di queste sequenze in altre parti di DNA suggerisce come da sole non siano sufficienti per la formazione del centromero. Altri centromeri sono stati osservati su cromosomi frammentati e inizialmente eucromatici. I centromeri sono pertanto definiti da un insieme di proteine, piuttosto che da una sequenza di DNA specifica.

### 5.3.9 Alcune strutture cromatiniche possono essere ereditate direttamente

I cambi nell'attività del centromero devono essere trasmessi alle generazioni successive. Si nota come la formazione di un centromero *de novo* richiede un inizio evento di inseminazione con strutture specifiche al DNA che contengono nucleosomi con CENP-A, che avviene più rapidamente lungo il DNA alfa satellite. Essendo che i tetrameri di H3-H4 sono ereditati si nota come un nucleosoma contenente CENP-A è stato assemblato su una lunghezza di DNA è banale come un nuovo centromero possa essere generato nello stesso luogo, assumendo che la presenza di un istone CENP-A recluti selettivamente più istoni CENP-A nei suoi vicini. Si nota inoltre come la generazione del centromero sia un processo altamente cooperativo.

### 5.3.10 L'attivazione e la repressione di strutture di cromatina può essere ereditato epigeneticamente

L'ereditarietà epigenetica è centrale nella creazione di organismi multicellulari in quanto permettono la creazione di tessuti diversi. Si nota come le strutture specifiche della cromatina tendono a persistere e a essere trasmesse nei cicli di divisione.

### 5.3.11 Le strutture di cromatina sono importanti per le funzioni del cromosoma eucariotico

La condensazione del DNA in nucleosomi è fondamentale per l'evoluzione di organismi multicellulari in quanto cellule si devono specializzare cambiando l'accessibilità e l'attività di centinaia di geni, processo che richiede sulla memoria cellulare che si trova in parte nella struttura cromatinica che crea strutture che silenziano geni temporaneamente o in maniera persistente.

## 5.4 La struttura globale dei cromosomi

Come una fibra a  $30nm$  il nucleo avrebbe una lunghezza di  $1mm$  e sarebbe troppo grande per essere contenuto nel nucleo. Si rende pertanto necessario un secondo livello di piegamento che coinvolge il piegamento della cromatina in una serie di anelli e bobine in maniera dinamica.

### 5.4.1 I cromosomi sono piegati in grandi anelli di cromatina

I cromosomi, accoppiati in preparazione per la meiosi si presentano in una serie di grandi anelli di cromatina che si emanano da un asse cromosomico lineare. In queste circostanze un anello contiene sempre la stessa sequenza di DNA che rimane estesa nello stesso modo. Questi cromosomi producono grandi quantità di RNA e la maggior parte dei geni negli anelli sono espressi. La maggior parte del DNA si trova comunque condensata nell'asse dove i geni non sono espressi.

### 5.4.2 Cromosomi politene sono utili per visualizzare le strutture di cromatina

Alcune tipi di cellule diventano anormalmente grandi attraverso multipli cicli di sintesi del DNA senza divisione cellulare. Tali cellule sono dette poliploidi e tutte le copie di ogni cromosoma sono allineate creando giganti cromosomi politene organizzati in bande e interbande, dove il DNA è più concentrato nelle prime. Queste strutture permettono di esaminare l'organizzazione della cromatina su larga

scala. Si nota come insiemi specifici di proteine non istoni si assemblano sui nucleosomi influenzando le funzioni biologiche. Il reclutamento di queste proteine può avvenire su grandi lunghezze di DNA, generando strutture cromatiniche simili a grandi tratti del genoma, separate dai domini vicini da proteine di barriera. Il cromosoma interfase può essere pertanto considerato come un mosaico di strutture cromatiniche contenenti modifiche del nucleosoma associate con un particolare insieme di proteine non istoni.

##### 5.4.3 Esistono multiple forme di cromatina

L'analisi della locazione degli istoni e delle proteine non istone nella cromatina può essere mappata lungo tutta la sequenza di DNA di un organismo. Tre tipi di cromatina repressiva dominano nell'organismo della *Drosophila* e due tipi di cromatina su geni attivamente trascritti. Ognuno di questi tipi è associato con un complesso diverso di proteine non istone. L'eterocromatina classica contiene più di 6 proteine come la proteina eterocromatina 1 (HP1), mentre le forme a polycomb contengono un simile numero di proteine PcG. Oltre ai cinque principali sono presenti forme minori ognuna delle quali diversamente regolata e con ruoli distinti. L'insieme delle proteine legate ad un luogo varia in base al tipo della cellula e al suo stato di sviluppo.

##### 5.4.4 Gli anelli di cromatina decondensano quando i geni al loro interno sono espressi

Quando la cellula si sviluppa distinti puffs cromosomici si formano e i vecchi recedono in cromosomi politene quando nuovi geni sono espressi e i vecchi sono disattivati. Dall'ispezione di un puff si nota come la maggior parte si generino dalla decondensazione di una singola banda cromosomica. Quando i geni nell'anello non sono espressi questo assume una struttura spessa, estendendosi durante la loro espressione. Si nota come la cromatina a lato del loop decondensato appaia più compatta in quanto un anello costituisce un dominio funzionale distinto della struttura cromatinica. Anelli di cromatina altamente condensata si espandono quando il gene è espresso. Ognuno dei 46 cromosomi dell'interfase tendono a occupare il proprio territorio discreto nel nucleo e non sono impigliati tra di loro estensivamente. Le regioni eterocromatiniche sono associate con la lamina nucleare. La maggior parte dei cromosomi umani sono piegati in una conformazione detta globulo frattale: un ordinamento senza nodi che facilita un impacchettamento denso che preserva l'abilità della cromatina di spiegarsi e ripiegarsi.

##### 5.4.5 La cromatina può muoversi a siti specifici nel nucleo per alterare l'espressione dei geni

La posizione dei geni nel nucleo cambia quando diventa altamente espresso, pertanto una regione che diventa attivamente trascritta si espande fuori dal territorio del suo cromosoma in un anello esteso in quanto l'assemblamento delle più di 100 proteine richieste per iniziare l'espressione genica viene facilitato in regioni del nucleo ricche di queste. Si nota come il nucleo è eterogeneo con regioni con funzioni diverse in cui porzioni del cromosoma si possono muovere quando sono soggette a processi biochimici.

### 5.4.6 Reti di macromolecole formano un insieme di ambienti biochimici distinti nel nucleo

Uno dei subcompartimenti più grossi del nucleo è il nucleolo, una struttura dove si forma la subunità del ribosoma e altre reazioni specializzate. Consiste di una rete di RNA e proteine concentrate intorno a geni di RNA ribosomiale che sono attivamente trascritti, presenti in multiple copie nel genoma che sono raggruppate in un singolo nucleolo nonostante si trovino su cromosomi diversi. Sono presenti anche diversi organelli come corpi di Cajal e granuli intercromatinici, composti da proteine e RNA che si legano creando reti altamente permeabili a altre proteine e molecole di RNA nel nucleoplasma circostante. Queste strutture creano ambienti biochimici distinti immobilizzando gruppi di macromolecole distinte permettendo a altre molecole in essi di essere efficientemente processate attraverso cammini di reazione complessi impartendo molti dei vantaggi cinetici della compartimentalizzazione pur essendo incapaci di concentrare o escludere piccole molecole in quanto non possiedono il bistrato lipidico. Questi subcompartimenti si formano solo quando necessario creando alte concentrazioni locali di enzimi e molecole di RNA necessarie per un processo particolare. In maniera analoga quando il DNA è danneggiato da radiazioni l'insieme di enzimi necessario alla riparazione del DNA si unisce in foci discrete nel nucleo creando fabbriche di riparazione che sintetizzano DNA o RNA. Queste entità si formano attraverso lunghe catene polipeptidiche o molecole di RNA non codificanti intervallati con siti di legame specifici che concentrano multiple proteine e altre molecole necessarie. Esiste anche una struttura analoga al citoscheletro in cui i cromosomi e gli altri componenti del nucleo sono organizzati detta matrice nucleare.

### 5.4.7 I cromosomi mitotici sono altamente condensati

I cromosomi di quasi tutte le cellule eucariotiche diventano visibili durante la mitosi, quando si arrotolano per formare una struttura altamente condensata con una lunghezza di un decimo rispetto a un cromosoma interfase. Le due molecole di DNA prodotte durante la replicazione del DNA durante la replicazione sono piegate separatamente producendo due cromatidi sorelle unite al centromero, coperti da grandi quantità di complessi RNA-proteine e altre molecole. Una volta che si toglie questa copertura ogni cromatide si organizza in anelli di cromatina che si estendono da un'impalcatura centrale. Le caratteristiche visibili del cromosoma mitotico sono nello stesso ordine rispetto alla molecola di DNA. La condensazione di cromosomi mitotici è il livello finale di impacchettamento dei cromosomi. Questo processo di compattazione è dinamico e altamente organizzato che permette lo sgarbugliamento delle cromatidi sorelle in modo da permettere la loro separazione e protegge le molecole di DNA fragili dalla rottura quando sono tirate nelle cellule figlie. La condensazione comincia nella fase *M* ed è connessa con la progressione del ciclo cellulare. Durante questa fase si blocca l'espressione dei geni e sono fatte specifiche modifiche agli istoni in modo da riorganizzare la cromatina mentre si compatta.

## 5.5 Come si evolvono i genomi

Geni omologhi (con uno stesso antenato) possono essere riconosciuti in lunghe distanze filogenetiche. Il riconoscimento di questa similitudine è utile per inferire la funzione di proteine e geni. La sequenza dei geni individuali è più conservata rispetto alla struttura genomica.

### 5.5.1 Il confronto tra i genomi rivela sequenze di DNA attraverso la loro evoluzione del throughput di evoluzione

Le regioni di progeine che codificano le sequenze di amminoacidi delle proteine (esoni) si trovano in segmenti corti (145 paia di nucleotidi), dispersi in area in cui la sequenza nucleotidica ha poche conseguenze. Questo ordinamento rende difficile identificare tutti gli esoni e inizio e fine di un gene. Un approccio è cercare per sequenze di DNA con una funzione simile e sono più probabilmente conservate rispetto a quelle senza funzione. Tali regioni con sequenze simili sono dette regioni conservate. Oltre a rivelare le sequenze che codificano esoni importanti e molecole di RNA contengono sequenze di DNA regolatorio con funzione sconosciuta. Le regioni non conservate riflettono il DNA la cui sequenza è molto meno probabilmente critica per la funzione. Si possono ottenere risultati più importanti includendo diverse specie. Circa il 5% del genoma consiste di sequenze conservate multispecie, di cui solo un terzo codifica proteine, mentre il resto è regolatorio e cluster di siti di legami per proteine e per la produzione di molecole di RNA con altro scopo.

### 5.5.2 L'alterazione dei genomi è causata da fallimento del normale meccanismo per la copia e il mantenimento del DNA e da elementi del DNA trasponibili

L'evoluzione dipende da incidenti e errori seguiti da sopravvivenza non casuale. La maggior parte dei cambi genetici sono il risultato di fallimenti nel meccanismo in cui i genomi sono copiati e riparati quando danneggiati insieme al movimento di parti del DNA trasponibili. Il meccanismo di replicazione compie un errore in un nucleotide su mille ogni milione di anni. Nonostante questo in una popolazione di 10·000 diploidi individuali in un milione di anni ogni possibile sostituzione nucleotidica sarà stata provata circa 20 volte. Gli errori nella replicazione, ricombinazione o riparazione del DNA possono portare a cambi locali nella sequenza dette mutazioni puntuali o a riordinamenti a larga scala come eliminazioni, duplicazioni, inversioni e traslocazioni. Oltre a questi fallimenti il genoma contiene elementi di DNA mobile che sono fonte di cambi genomici, questi trasposoni sono sequenze parassitiche che si diffondono nel genoma che colonizzano, distruggendo la funzione o alterando la regolazione dei geni. Questo processo ha modificato profondamente i genomi.

### 5.5.3 La sequenza genomica di due specie differisce in proporzione al tempo in cui si sono evolute separatamente

Le differenze tra i genomi delle specie si sono accumulate lungo 3 miliardi di anni. Il processo di evoluzione genomica si può ricostruire tra confronti dettagliati dei genomi degli organismi contemporanei. La struttura di base per questo lavoro è l'albero filogenetico, costruito secondo il confronto dei geni o sequenze proteiche. Per organismi distanti la conservazione delle sequenze è dovuta a una selezione purificativa in cui vengono eliminate le mutazioni che interferiscono con le funzioni genetiche importanti.

### 5.5.4 L'albero filogenetico costruito dal confronto delle sequenze di DNA traccia le relazioni di tutti gli organismi

Gli alberi filogenetici basati sulle sequenze molecolari possono essere confrontati con i record di fossili in modo da avere un'idea ancora più chiara del processo evolutivo in quanto il secondo fornisce date assolute, nonostante tempi di divergenza precisi siano difficili da stabilire. Gli alberi costruiti con questo aiuto suggeriscono che i cambi nella sequenza di particolari geni o proteine avvengono ad un

tasso costante che dà un orologio molecolare per l'evoluzione unico per differenti categorie di sequenze di DNA. Questo orologio scorre più rapidamente per sequenze non soggette a selezione purificativa: porzioni di introni che non hanno splicing o segnali regolatori, la terza posizione in codoni sinonimi e geni che sono stati inattivati irreversibilmente da mutazioni (pseudogeni) e più lentamente per sequenze con vincoli funzionali critici. Occasionalmente cambi rapidi possono accadere in sequenze altamente conservate e si pensa riflettano periodi di selezione positiva per mutazioni vantaggiose. La velocità di questo orologio molecolare è determinata dal tasso di selezione purificante e dal tasso di mutazione. Si nota come il tasso di mutazione in mitocondri animali sia estremamente alto. Le categorie per cui l'orologio è veloce sono le più informative per eventi evolutivi recenti. Questi orologi presentano inoltre una risoluzione temporale maggiore rispetto ai record fossili e sono una guida affidabile.

### 5.5.5 Un confronto tra cromosomi umani e dei topi mostra come la struttura dei genomi diverge

I lignaggi umani e dei topi si sono evoluti separatamente per circa 80 milioni di anni e i secondi hanno orologi molecolari veloci, divergendo più rapidamente. Se il genoma è organizzato in cromosomi quasi identicamente la sua organizzazione diverge significativamente. Continuano ad esistere blocchi di DNA in cui l'ordine dei geni è lo stesso, dette regioni di sintenia. Da queste osservazioni si è notato come piccole parti di DNA sono eliminate e aggiunte a velocità elevate, anche se guadagni di sequenza da duplicazione e moltiplicazione di trasposoni ha compensato per queste eliminazioni, pertanto la dimensione del genoma è rimasta incambiata dall'ultimo comune antenato di topi e umani per gli umani, mentre è diminuita per i topi. Prove di questo possono essere ottenute dal confronto dettagliato delle regioni di sintenia. Il DNA è aggiunto ai genomi attraverso la duplicazione spontanea di segmenti cromosomici e dall'inserzione di nuove coppie di trasposoni attivi essendo gli eventi di trasposizione duplicativi.

### 5.5.6 La dimensione del genoma dei vertebrati riflette i tassi relativi di addizione e perdita del DNA in un lignaggio

In vertebrati imparentati lontanamente la dimensione del genoma può cambiare considerevolmente senza effetto sull'organismo o sul numero dei geni. Questo avviene in quanto tutti i vertebrati sono soggetti di un continuo processo di addizione e perdita del DNA e pertanto la dimensione del genoma dipende dall'equilibrio di questi processi. Il numero di geni è invece mantenuto attraverso selezione purificante.

### 5.5.7 Si può inferire la sequenza di alcuni genomi antichi

I genomi di organismi ancestrali possono essere inferiti ma non osservati direttamente. La selezione deve aver mantenuto proprietà funzionali critiche, pertanto il confronto tra organismi odierni mostra che la frazione del genoma soggetta a selezione purificante è piccola e pertanto la sequenza di DNA varia grandemente. Per antenati con un gran numero di discendenti si può inferire la loro sequenza confrontando la sequenza di questi ultimi.

### **5.5.8 Il confronto tra sequenze multispecie identifica sequenze di DNA conservato di funzione sconosciuta**

La massa di sequenze di DNA nei database mette a disposizione una risorsa utile a determinare i cammini evolutivi e come le cellule e gli organismi funzionano. Si è notata una grande quantità di sequenza di DNA che non codifica proteine conservata durante l'evoluzione dei mammiferi, chiaramente rivelata analizzando blocchi di DNA sintonici da specie diverse, determinando le sequenze conservate multispecie. La maggior parte di queste sequenze non codificanti sono corte per la maggior parte e la loro conservazione implica una funzione importante che è stata mantenuta dalla selezione purificante. Molte di queste sequenze producono molecole di RNA non tradotte come i lunghi RNA non codificanti lncRNA che si pensa abbiano un'importante funzione nella regolazione dell'espressione genica.

### **5.5.9 Cambi in sequenze precedentemente conservate possono aiutare a decifrare passi critici dell'evoluzione**

L'analisi delle differenze tra le specie viene effettuata attraverso le sequenze conservate multispecie che rappresentano le sequenze di DNA probabilmente funzionalmente importanti. Queste sequenze non sono conservate perfettamente e si trovano delle differenze che in una piccola percentuale rappresentano segni di scatti evolutivi. Le regioni accelerate umane (HARs) si pensa riflettano funzioni importanti che ci differiscono dalle altre specie. Un quarto dei 50 siti individuali si trovano vicino a geni associati con lo sviluppo neuronale. Un altro approccio consiste nel ricercare mutazioni importanti attraverso le sequenze di DNA conservate e concentrarsi sui siti cromosomici con eliminazioni. Solo una di queste eliminazioni trovate negli umani si trovano in regioni codificanti.

### **5.5.10 Mutazioni nella sequenza di DNA che controlla l'espressione genica ha guidato molti dei cambi evolutivi nei vertebrati**

Si può tentare di tracciare le origini degli elementi regolatori del DNA che hanno avuto ruoli critici nell'evoluzione dei vertebrati. Si inizia identificando i 3 milioni di sequenze non codificanti che sono state conservate in evoluzioni recenti dei vertebrati che rappresenta un'innovazione caratteristica di un ramo dell'albero dei vertebrati e consistono di DNA regolatorio che regola il gene vicino. Si possono identificare i geni più vicini a esse e si può stimare quando ogni elemento regolatorio è arrivato in esistenza. Gli elementi regolatori conservati appaiono associati con geni che codificano proteine per la regolazione della transizione per proteine per lo sviluppo embrionale, successivamente si trovano le regolazioni per geni codificanti i recettori per segnali extracellulari e l'ultima innovazione sembra riguardare le proteine che modificano altre proteine dopo la traduzione.

### **5.5.11 La duplicazione genica fornisce una fonte di novità genica durante l'evoluzione**

L'evoluzione dipende dalla creazione di un nuovo gene e dalla modifica di quelli preesistenti. I geni senza omologhi sono scarsi e si trovano spesso famiglie di geni con differenti membri in specie diverse per la cui creazione è stato necessario la duplicazione ripetuta di geni le cui copie diversero per svolgere funzioni diverse. La duplicazione dei geni avviene ad alti tassi in tutti i lignaggi, contribuendo al processo di addizione del DNA.



### 5.5.12 Geni duplicati divergono

Inizialmente non è presente alcuna selezione per mantenere lo stato duplicato, pertanto molti eventi di duplicazione possono essere seguiti da mutazioni che causano la perdita di funzione sui geni. Questo ciclo eliminerebbe l'effetto della duplicazione. Passando il tempo la somiglianza tra tali pseudogeni disattivati e l'origine della duplicazione viene erosa dalle mutazioni con la relazione omologa non più identificabile. Un'altra possibilità è che entrambe le copie rimangano funzionali mentre divergono, prendendo ruoli diversi e spiega la presenza delle famiglie geniche e ha un ruolo critico nell'evoluzione di complessità biologica crescente. Una duplicazione dell'intero genoma può avvenire nel caso in cui avvenga la replicazione del DNA senza divisione cellulare. Il risultato con il passare del tempo crea altro materiale genico con possibilmente funzioni diverse.

### 5.5.13 L'evoluzione del gene globina mostra come la duplicazione del DNA contribuisce all'evoluzione degli organismi

La somiglianza delle proteine globina odierne in struttura e sequenza amminoacida indica che devono derivare da un gene antenato comune. Si possono ricostruire eventi che hanno prodotto i vari tipi di molecole di emoglobina considerando le diverse forme che assume nell'albero della vita. Questa molecola ha permesso agli organismi di crescere in dimensione e molecole che legano l'ossigeno si possono trovare in piante, funghi e batteri. Negli animali la proteina più primitiva è una catena polipeptidica di circa 150 amminoacidi, ma si presenta in forma più complessa in vertebrati complessi, composta da due catene di globina. Durante una serie di mutazioni e duplicazioni hanno stabilito un gene globina per nel genoma in cui ognuno codificante due catene  $\alpha$ - e  $\beta$ - che si associano per formare una molecola di emoglobina che consiste delle quattro catene. I quattro siti che si legano all'ossigeno nella molecola  $\alpha_2\beta_2$  interagiscono, permettendo un cambio allosterico cooperativi quando lega e rilascia l'ossigeno che permette uno scambio più efficiente. Durante l'evoluzione dei mammiferi il gene per la catena  $\beta$  si è duplicato ed è stato modificato dando origine a una catena simile alla  $\beta$  presente nel feto, con una molecola con un'affinità più alta rispetto a quella presente negli adulti. Il nuovo gene si è duplicato di nuovo e mutato producendo due nuovi geni  $\varepsilon$  e  $\gamma$ , con la prima prodotta nelle prime fasi dello sviluppo per formare  $\alpha_2\varepsilon_2$  rispetto alla catena fetale  $\gamma$  che produce  $\alpha_2\gamma_2$ . Durante l'evoluzione dei primati una nuova duplicazione del gene  $\beta$  ha prodotto un gene per la  $\delta$ -globina producendo una forma minore di emoglobina  $\alpha_2\delta_2$  presente unicamente nei primati adulti. Ognuno di questi geni duplicati è stato modificato da mutazioni puntiformi che influiscono sulle proprietà della molecola di emoglobina finale e cambi nelle regioni regolatorie che determinano la temporizzazione e il livello di espressione del gene e pertanto ogni globina è creata in diverse quantità e a tempi diversi dello sviluppo dell'uomo. Nel genoma umano i geni generati dal  $\beta$  gene originale sono ordinati in una serie di sequenze di DNA omologhe che si trovano a 50'000 basi di distanza su un singolo cromosoma. Un singolo cluster di geni  $\alpha$ -globina si trova su un cromosoma separato. Esistono duplicazioni di sequenze di pseudo geni globina disattivati da mutazioni che impediscono la loro espressione come proteine funzionali.

### 5.5.14 I geni che codificano nuove proteine possono essere creati ricombinando gli esoni

La duplicazione del DNA può agire su una scala più piccola unendo segmenti di DNA corti duplicati. Le proteine codificate da geni creati in questa maniera possono essere riconosciute dalla presenza di domini simili legati covalentemente l'uno all'altro in serie. Le immunoglobuline e le proteine fibrose sono un esempio di questo. Ogni esone separato codifica un'unità di proteina pieghevole o dominio.

Si crede che l'organizzazione del DNA come una serie di esoni separati da introni ha facilitato l'evoluzione di nuove proteine: la duplicazione necessaria per formare un singolo gene codificante per una proteina con domini ripetuti può avvenire rompendo e riunendo il DNA lungo l'introne a un lato di un esone e pertanto gli introni aumentano la probabilità di un evento di duplicazione favorevole. Vari parti dei geni sono servite come elementi modulari che si sono duplicati e spostati lungo il genoma per formare una grande diversità delle cose viventi.

### 5.5.15 Mutazioni neutrali si diffondono diventando fissate in una popolazione, con probabilità dipendente dalla dimensione della popolazione

Confrontando due specie che si sono separate fa poca differenza quali individui sono confrontati. Nonostante questa ogni diversità fissa comincia come una nuova mutazione in un individuo singolo. Se la dimensione della popolazione riproducibile in cui è avvenuta la popolazione è  $N$ , la frequenza dell'allele iniziale per una nuova mutazione è  $\frac{1}{2N}$  per un organismo diploide. La fissazione nella popolazione dipende dalle conseguenze funzionali della popolazione: una mutazione con un evento deleterio viene eliminata dalla selezione purificante e non si fissa, mentre le mutazioni che danno un vantaggio riproduttivo su individui si possono diffondere rapidamente. La mutazione selezionata lungo una quantità di sequenze vicine sarà una parte di un grande mosaico. Le mutazioni neutrali possono diffondersi e fissarsi nella popolazione e contribuiscono al cambio evolutivo dei genomi, creando la maggiore differenza tra scimmie e esseri umani. La diffusione delle mutazioni neutrali dipende su una variazione casuale del numero di progenie che porta la mutazione prodotta da ogni individuo che porta la mutazione, causando cambi nella relativa frequenza dell'allele mutante nella popolazione. Questo processo di cammino casuale può causare l'estinzione della mutazione o la sua fissazione. Assumendo una popolazione di dimensione costante e con accoppiamenti casuali e selezione neutrale per la popolazione, quando una nuova mutazione avviene in una popolazione di dimensione  $N$ , la probabilità che essa si fissi è circa  $\frac{1}{2N}$  in quanto ci sono  $2N$  copie di geni nella popolazione diploide e ognuna di esse ha una probabilità uguale di diventare la versione dominante. Per le mutazioni che si fissano il tempo di fissazione è circa  $4N$  generazioni.

### 5.5.16 Molto può essere imparato dall'analisi della variazione tra gli umani

Esiste un grande numero di varianti che si incontrano. Da un confronto dettagliato di sequenze di DNA tra un gran numero di umani in tutto il globo è possibile determinare quante generazioni sono passate dall'origine di una mutazione neutrale. Da questi dati è possibile determinare le migrazioni degli esseri umani primitivi. Un'altra fonte di variazione si trova nella duplicazione e eliminazione di grandi blocchi di DNA. Confrontando il genoma di individuo con il modello si possono trovare circa 100 differenze in blocchi di sequenza. Alcune di queste copy number variations (CNVs) sono molto comuni, mentre altre sono presenti in poche minoranze. Le variazioni intraspecie sono caratterizzate come polimorfismi a singolo nucleotide (SNPs) che sono punti in cui la sequenza genomica in una grande frazione della popolazione possiede un nucleotide, mentre un'altra un secondo. Per qualificarsi come un polimorfismo la variante deve essere comune abbastanza che due individui randomici siano diversi sia almeno dell'1%. Queste variazioni possono utili per analisi della mappatura genetica in cui si tenta di associare tratti specifici con sequenze specifiche, pur non avendo effetto sul fitness umano. Si trovano anche sequenze con tassi di mutazioni estremi come le ripetizioni CA, onnipresenti nel genoma umano e degli altri eucarioti. Se sequenze con il motif  $(CA)_n$  sono replicate con bassa fedeltà a causa di un errore tra lo stampo e un nuovo filamento di DNA appena sintetizzato, pertanto

il valore  $n$  varia grandemente tra un genoma e il prossimo. Questa ripetizione crea un ottimo marcatore. Nonostante si trasmetta abbastanza fedelemente tra i figli (una lunghezza per la madre e una per il padre) i cambi sono abbastanza frequenti da permettere alti livelli di eterozigosità nella popolazione. Un sottoinsieme di questi mutazioni sono responsabili per gli aspetti ereditabili degli individui.

## Capitolo 6

# Replicazione, riparazione e ricombinazione del DNA

L'abilità della cellula di mantenere un alto grado di ordine dipende dall'accurata duplicazione di grandi quantità di informazioni genetiche trasportate in forma chimica come DNA. Il processo della replicazione del DNA deve avvenire prima che una cellula possa produrre due cellule figlie geneticamente indipendenti. Il mantenimento dell'ordine richiede anche una continua sorveglianza e riparazione dell'informazione genetica, continuamente danneggiata dall'ambiente chimicamente, attraverso radiazioni, calore e molecole attive generate nella cellula. Questi processi vengono svolti da proteine che catalizzano processi rapidi e accurati che avvengono all'interno della cellula. Se la sopravvivenza a breve termine di una cellula dipende dalla prevenzione dei cambi nella sequenza, quella a lungo termine richiede che una sequenza di DNA cambi lungo le generazioni in modo da permettere adattamento evolutivo a un ambiente dinamico. Nonostante gli sforzi della cellula avvengono dei cambi nel DNA che possono mettere a disposizioni varianti che la selezione spinge durante l'evoluzione.

### 6.1 Il mantenimento della sequenza di DNA

La sopravvivenza dell'individuo richiede un alto grado di stabilità genetica. Solo raramente il processo di mantenimento del DNA fallisce causando cambi permanenti nel DNA o mutazioni, che possono distruggere un organismo se avvengono in posizioni vitali della sequenza del DNA.

#### 6.1.1 I tassi di mutazione sono estremamente bassi

Il tasso di mutazione può essere determinato direttamente attraverso esperimenti con batteri come l'*Escherichia coli* che si divide ogni 30 minuti e una cellula singola può generare una grande popolazione in meno di un giorno. In tale popolazione è possibile individuare una piccola frazione di batteri con una mutazione dannosa in un gene particolare se non è necessario alla sopravvivenza. Tale frazione è una sottostima delle mutazioni in quanto ne esistono silenti. Dopo aver corretto per queste mutazioni silenti si trova che un singolo gene per una proteina di dimensione media ( $10^3$  paia di nucleotidi) accumula una mutazione una volta ogni  $10^6$  generazioni: il tasso di mutazione è pertanto di tre cambi nucleotidici per  $10^{10}$  nucleotidi per generazione di cellule. Recentemente è stato possibile misurare il tasso di mutazione direttamente in organismi più complessi e la riproduzione sessuale. In

questo caso la sequenza genomica di una famiglia si sequenzia e una comparazione determina circa 70 mutazionidi singoli nucleotidi in ogni discendente. Normalizzando alla dimensione del genoma umano il tasso di mutazione di un nucleotide cambia per  $10^8$  nucleotidi per generazione. Questa è una sottostima in quanto non considera le mutazioni letali che non sarebbero presenti nella progenie. Circa 100 divisioni accadono tra il concepimento e il tempo di produzione di uova e sperma che producono una nuova generazione, pertanto il tasso umano di mutazione è di 1 mutazione per  $10^{10}$  divisioni cellulari. In entrambi gli esperimenti si nota come i tassi di mutazioni sono estremamente bassi e con un fattore di tre l'uno dall'altro: sono infatti preservati i meccanismi base che garantiscono questi tassi bassi e sono stati conservati da cellule ancestrali molto antiche.

### 6.1.2 I tassi di mutazione bassi sono necessari per la vita come la conosciamo

Essendo che la maggior parte delle mutazioni sono deleterie, nessuna specie può permettersi di accumularle. Pur essendo il tasso di mutazioni basso si pensa limiti il numero di proteine che può dipendere a 30'000 in quanto per un numero maggiore la probabilità che una componente chiave sia danneggiata da una mutazione diventa troppo elevata. Le cellule di un organismo che si riproduce sessualmente sono di due tipi: le cellule germinali e quelle somatiche. Le prime trasmettono l'informazione tra genitore e prole, mentre le seconde formano il corpo dell'organismo. Le cellule germinali devono essere protette contro alti tassi di mutazione per mantenere la specie, cosa che deve avvenire anche nelle cellule somatiche per la corretta formazione di un corpo organizzato. I cambi nucleotidici nelle cellule somatiche possono creare cellule varianti, alcune delle quali attraverso selezione naturale locale proliferano rapidamente. Nei casi estremi si genera un cancro, la cui probabilità aumenta linearmente con il tasso di mutazione. Pertanto sia la perpetuazione di una specie con un gran numero di geni e la prevenzione di cancro che risultano da mutazioni delle cellule somatiche dipendono dall'alta fedeltà con cui le sequenze di DNA sono replicate e mantenute.

## 6.2 Meccanismi di replicazione del DNA

Tutti gli organismi duplicano il proprio DNA con estrema accuratezza prima di ogni divisione cellulare.

### 6.2.1 L'accoppiamento delle basi sottostà la replicazione e riparazione del DNA

Il meccanismo che usa la cellula per copiare la sequenza di DNA è il DNA templating che richiede la separazione dell'elica di DNA in due filamenti stampi e il riconoscimento di ogni nucleotide nel filamento stampo da parte del nucleotide libero complementare. La separazione dell'elica espone i gruppi donatori e accettori con legami a idrogeno per ogni base, allineandolo per la polimerizzazione catalizzata da enzimi in una nuova catena di DNA. Il primo di tali enzimi è detto DNA polimerasi: i nucleotidi liberi che formano il substrato sono trifosfati deossiribonucleidici e la loro polimerizzazione in DNA richiede un singolo filamento come stampo.

### 6.2.2 La forcella di replicazione del DNA è asimmetrica

Durante la replicazione ognuno dei due filamenti originali viene utilizzato come stampo per la formazione di un nuovo filamento. Siccome ognuna delle due cellule figlie eredita una nuova doppia elica

contenente un filamento originale e uno nuovo quest'ultima è detta replicata semiconservativamente. Si nota una regione di replicazione localizzata che si muove lungo la doppia elica parentale. A causa della struttura a forma di Y tale regione è detta forcella di replicazione e vi si trova un complesso multienzimico che contiene la DNA polimerasi che sintetizza il DNA per entrambi i filamenti figli. Il meccanismo di replicazione del DNA sembra una continua crescita di entrambi i filamenti, ma a causa dell'orientamento antiparallelo il meccanismo richiederebbe la polimerizzazione da 5'-3' per un filamento e da 3'-5' per l'altro e pertanto due tipi diversi di enzimi DNA polimerasi che in realtà riescono a sintetizzare unicamente nella direzione 5'-3'. Il filamento nella direzione opposta viene sintetizzato grazie all'esistenza di segmenti di DNA detti frammenti di Okazaki (1000 – 2000 basi) transitori che si trovano alla forcella di replicazione. Questi sono polimerizzati unicamente nella direzione 5'-3' e sono uniti dopo la sintesi per creare lunghe catene di DNA. La forcella ha una struttura simetrica e il filamento figlio sintetizzato continuamente è detto il filamento principale la cui sintesi dipende leggermente dal filamento in ritardo la cui sintesi è discontinua e ha direzione opposta alla crescita della sintesi del DNA.

### 6.2.3 L'alta fedeltà del meccanismo di replicazione del DNA richiede molti meccanismi di correzione

Le coppie complementari standard non sono le uniche possibili: con piccoli cambi nella geometria dell'elica è possibile formare legami tra G e T e esistono rare forme tautomeriche di C che si accoppiano con A. L'alta fedeltà della duplicazione richiede meccanismi di controllo sequenziali che correggono ogni accoppiamento iniziale errato. Il primo passo è fatto dalla DNA polimerasi quando un nuovo nucleotide è aggiunto covalentemente alla catena: quello corretto ha un'affinità più alta con la polimerasi che si muove. Dopo il legame, ma prima dell'addizione covalente alla catena l'enzima subisce un cambio conformazionale in cui si stringe lungo il sito attivo. Lo stringimento avviene più facilmente per le basi incorrette. La seconda reazione di correzione è detta correzione esonucleica avviene quando un nucleotide è aggiunto covalentemente alla catena. La DNA polimerasi è altamente discriminante nelle catene di DNA che allungano: richiedono una base 3' – OH accoppiata con un filamento primario. Queste molecole con un nucleotide sbagliato a tale terminazione del primer non sono efficaci come stampi. Le molecole di DNA polimerasi correggono il primer attraverso un separato sito catalitico (in una diversa subunità o dominio). Questo 3'-5' esonucleasi di correzione elimina ogni residuo non accoppiato o malaccoppiato alla terminazione del primer, continuando fino a che abbastanza nucleotidi sono stati rimossi per generare una terminazione correttamente accoppiata che può iniziare la sintesi. Queste proprietà di autocorrezione della DNA polimerasi dipendono dalla richiesta di un primer a terminazione perfettamente accoppiata, cosa non inclusa nell'RNA polimerasi. La frequenza di errore è di 1 ogni  $10^4$  eventi di polimerizzazione nella sintesi e traduzione dell'RNA.

### 6.2.4 Solo la replicazione nella direzione 5'-3' permette una correzione efficiente

La necessità di accuratezza spiega perché la replicazione del DNA avviene solo nella direzione 5'-3': se una DNA polimerasi aggiungesse trifosfati deossiribonucleici nella direzione opposta la terminazione 5' dovrebbe fornire il trifosfato necessario per il legame covalente e gli sbagli non potrebbero essere eliminati attraverso idrolizzazione in quanto la terminazione 5' senza tale gruppo terminerebbe la sintesi. Nonostante tutti questi meccanismi la DNA polimerasi può creare degli errori, che possono essere individuati da un processo detto di riparazione direzionata al filo.

### 6.2.5 Un enzima che polimerizza i nucleotidi sintetizza corti primer di RNA sul filamento in ritardo

Per il filamento principale è necessario un solo primer all'inizio della replicazione, mentre sul filamento in ritardo ogni volta che la DNA polimerasi completa un corto frammento di Okazaki deve iniziare a sintetizzare un nuovo frammento a un sito più in avanti. Un meccanismo produce il primer necessario e dipende dall'enzima detto DNA primasi, che usa trifosfati ribonucleici per sintetizzare corti primer a RNA sul filamento in ritardo. Si noti come un filamento di RNA può formare legami con uno di DNA, generando una doppia elica ibrida se le sequenze sono complementari e pertanto lo stesso principio di sintesi del DNA guida la sintesi dei primer a RNA. Essendo che un primer a RNA contiene un nucleotide appropriatamente accoppiato con una terminazione  $3' - OH$  a una fine può esser allungato dalla DNA polimerasi a questa fine per formare un frammento di Okazaki, sintesi che finisce quando la polimerasi incontra il primer a RNA attaccato alla terminazione  $5'$  del segmento precedente. Per produrre un filamento continuo di DNA si utilizza un sistema di riparazione del DNA che elimina i primer a RNA e li sostituisce con DNA. L'enzima DNA ligasi successivamente unisce la terminazione  $3'$  del nuovo segmento con quella  $5'$  del vecchio. Il primer a RNA è necessario per mantenere basso il tasso di mutazioni in quanto marca i frammenti come copie sospette.

### 6.2.6 Proteine speciali aiutano l'apertura della doppia elica sopra la forcella di replicazione

Per permettere la sintesi la doppia elica deve essere aperta prima della forcella di replicazione in modo che i fosfati desossiribonucleici possano formare coppie di base con i filamenti. Essendo la doppia elica stabile in condizioni fisiologiche sono necessarie la DNA elicasi e un singolo filamento di DNA legato a proteine per aprire la doppia elica in tale ambiente. La DNA elicasi sono state per la prima volta isolate come proteine che idrolizzano l'ATP quando sono legate a un singolo filamento del DNA, reazione che le permette di spingersi lungo un filamento singolo. Quando incontrano una regione a doppia elica continuano a muoversi lungo il proprio filamento separandola a 1000 nucleotidi al secondo. Esistono elicasi che lavorano in entrambe le direzioni della polarità. Proteine che si legano a singoli filamenti di DNA o proteine di destabilizzazione dell'elica si legano strettamente e cooperativamente per esporre singoli filamenti di DNA senza coprire le basi aiutando l'elicasi stabilizzando la conformazione a singolo filamento e impedendo la formazione di corte eliche a forcella nel filamento in ritardo.

### 6.2.7 Un anello che scivola mantiene una DNA polimerasi in movimento sul DNA

Da sole le molecole di DNA polimerasi sintetizzerebbero una piccola stringa di nucleotidi prima di separarsi dallo stampo. Questa tendenza permette a una polimerasi che ha sintetizzato un frammento di Okazaki di separarsi e essere riciclata velocemente ma rende difficile la sintesi di sequenze lunghe se non fosse per una proteina detta PCNA che funziona come un morsetto scorrevole che mantiene la polimerasi fermamente sul DNA mentre si muove e la rilascia appena incontra una regione a doppia elica. Tale proteina forma un grande anello intorno alla doppia elica. Una faccia si lega al retro della DNA polimerasi, mentre l'intero anello scorre liberamente lungo il DNA. L'assemblaggio di tale proteina richiede un idrolisi dell'ATP da parte di un complesso proteico detto caricatore di morsetto che idrolizza l'ATP mentre carica il morsetto su una giunzione primer. Sul filamento principale la DNA polimerasi è strettamente legata al morsetto e i due rimangono associati per molto tempo. Sul filamento in ritardo ogni volta che la polimerasi arriva alla terminazione  $5'$  del frammento di Okazaki

si rilascia dal morsetto e si dissocia dallo stampo. Questa molecola si associa successivamente con un nuovo morsetto assemblato sul primer a RNA del successivo frammento.

### 6.2.8 Le proteine a una forcella di replicazione cooperano per formare una macchina di replicazione

Le proteine coinvolte nella replicazione sono ordinate in un complesso multienzima che si mantiene stazionario rispetto all'ambiente con il DNA che si muove al suo interno. All'inizio della forcella di replicazione la DNA elicasi apre l'elica dove due molecole di DNA polimerasi lavorano sul filamento principale e quello in ritardo, la prima in maniera continua la seconda in piccoli intervalli. L'associazione stretta di queste proteine aumenta l'efficienza della replicazione ed è permessa da un piegamento all'indietro del filamento in ritardo. Questo ordinamento facilita il caricamento del morsetto polimerasi ogni volta che un frammento di Okazaki è sintetizzato. Le proteine di replicazione sono legate insieme in una singola grande unità ( $> 10^6$  dalton) permettendo una sintesi efficiente e coordinata.

### 6.2.9 Un sistema di riparazione direzionato al filamento rimuove gli errori di replicazione che sfuggono alla macchina di replicazione

Una classe di mutanti possiede alterazioni nei geni mutatori che aumentano il tasso di mutazioni spontanee: alcuni di essi possiedono forme difettive di esonucleasi di correzione. Questo studio ha scoperto un meccanismo che rimuove errori di replicazioni e sfuggiti all'esonucleasi detto sistema di strand-directed mismatch repair che individua potenziali distorsioni nell'elica del DNA e corregge uno dei due nucleotidi scorretti: per essere corretto deve essere in grado di distinguere e rimuovere gli error solo nei nuovi filamenti. Nei procarioti viene utilizzato un meccanismo di distinzione dei filamenti che dipende dalla metilazione di residui di A nel GATC gruppi metile sono aggiunti in tutti i residui nella sequenza GATC, ma non fino a che è passato abbastanza tempo: pertanto le uniche sequenze GATC non metilate si trovano unicamente nel filamento dietro la forcella di replicazione. Il riconoscimento di GATC non metilati permette a nuovi filamenti di DNA di essere distinti dai vecchi. Il processo per questa correzione coinvolge pertanto il riconoscimento del filamento, l'escissione della porzione che contiene la mancata corrispondenza e la risintesi del segmento escisso con il vecchio stampo. Questo sistema riduce il numero di errori di un fattore tra 100 e 1000. Nelle cellule eucariotiche viene utilizzato un altro metodo per riconoscere il nuovo filamento: i nuovi filamenti in ritardo contengono transientemente nicks o single-strand breaks che forniscono il segnale che direziona il sistema di correzione. Questa strategia richiede che anche il filamento principale sia nicked.

### 6.2.10 La DNA topoisomerasi impedisce l'ingarbugliamento del DNA durante la replicazione

Mentre la forcella di replicazione si muove lungo il DNA crea il problema dell'avvolgimento: due filamenti genitori devono essere srotolati in modo da permettere la replicazione. In principio questo srotolamento può essere ottenuto rapidamente ruotando il cromosoma lungo la forcella ma è sfavorevole per cromosomi lunghi, pertanto solo il DNA davanti alla forcella viene srotolato e il sovraarrotolamento è rilassato dalla DNA topoisomerasi, associabile a una nucleasi reversibile che si aggiunge al backbone fosfato rompendo i legami fosfodiersteri nel filamento che si riforma quando la proteina va via. La topoisomerasi 1 produce una rottura a filamento singolo transiente che permette a due sezioni dell'elica del DNA tra le due parti del nick di ruotare liberamente tra di loro utilizzano



### 6.3. INIZIALIZZAZIONE E COMPLETAMENTO DELLA REPLICAZIONE DEL DNA NEI CROMOSOMI

---

i legami fosfodiesteri come pivot. Tensione nell'elica guida la rotazione nella direzione che la elimina: la replicazione avviene con la rotazione di piccole sezioni della sequenza. Siccome il legame con la proteina mantiene l'energia del legame fosfodiesteri la ricreazione di esso è rapida e non richiede energia esterna. La topoisomerasi II forma un legame covalente con entrambi i filamenti del DNA, creando una rottura in entrambi transiente. Questi enzimi sono attivati dai siti sul cromosoma dove le doppie eliche si incrociano come quelle generate da un sovraavvolgimento in una forcella di replicazione. Quando una molecola di topoisomerasi II si lega in un sito di incrocio usa l'idrolisi dell'ATP per rompere una doppia elica reversibilmente creando un gate di DNA, causa la seconda vicina doppia elica di passare nell'apertura e successivamente riunisce l'elica rotta e si dissocia dal DNA. Il passaggio della doppia elica nel gate avviene nella direzione che causa l'eliminazione del sovraavvolgimento. Può anche separare due cicli di DNA interbloccati.

#### 6.2.11 La replicazione del DNA è fondamentalmente simile in eucarioti e batteri

Molto di quello che è conosciuto dell'enzimologia della replicazione del DNA negli eucarioti è stato conservato durante il processo evolutivo che ha separato batteri ed eucarioti. Pur essendoci più proteine per i secondi le funzioni di base sono le stesse.

## 6.3 Inizializzazione e completamento della replicazione del DNA nei cromosomi

### 6.3.1 La sintesi del DNA inizia alle origini di replicazione

Essendo la doppia elica stabile per iniziare la replicazione deve essere aperta e i due filamenti separati, processo svolto da proteine iniziatrici che si legano alla doppia elica e separano i due filamenti rompendo i corrispondenti legami a idrogeno. Queste regioni in cui la doppia elica viene aperta sono dette origini di replicazione, specificate da sequenze di nucleotidi che attraggono proteine iniziatrici e altre lunghezze facili da aprire: A e T sono mantenute insieme da meno legami a idrogeno rispetto a C e G. Il processo alla base è lo stesso per batteri ed eucarioti, ma differisce nel modo in cui il processo avviene e come è regolato.

### 6.3.2 I cromosomi batterici hanno un'origine singola per la replicazione

Il genoma dell'E. coli è contenuto in una molecola di DNA circolare. La replicazione inizia ad un singolo sito e le due forcelle di replicazione procedono in direzioni opposte fino a che si incontrano. L'unico punto in cui la replicazione può essere controllata è l'inizio. Il processo inizia quando le proteine iniziatrici legate a ATP si legano in copie multiple in siti specifici del DNA avvolgendolo formando un complesso DNA-proteine che destabilizza la doppia elica vicina. Questo complesso attrae due DNA elicasi, legate a un caricatore elicasi, analogo al caricatore di morsetti ma che mantiene l'elicasi in una forma inattiva fino a che è propriamente caricata sulla forcella di replicazione nascente. Una volta che è caricata si dissocia e l'elicasi inizia a svolgere il DNA esponendo i filamenti in modo che la DNA primasi possa sintetizzare il primo primer a RNA che porta all'assemblaggio delle proteine necessarie per creare due forcelle di replicazione che continuano a sintetizzare il DNA fino a che tutto lo stampo è stato replicato. L'interazione tra le proteine iniziatrici con l'origine di replicazione è regolata, permettendo l'inizio del processo solo se ci sono nutrienti sufficienti per completarlo e che solo una replicazione avvenga per divisione cellulare. Dopo che la replicazione è iniziata la proteina iniziatrice è disattivata dall'idrolisi dell'ATP e l'origine di replicazione ha

### 6.3. INIZIALIZZAZIONE E COMPLETAMENTO DELLA REPLICAZIONE DEL DNA NEI CROMOSOMI

---

un periodo refrattario causato da un ritardo della metilazione dei nuovi nucleotidi A nell'origine. L'iniziazione non può accadere fino a che le A sono metilate e la proteina iniziatrice è riportata allo stato legato all'ATP.

#### 6.3.3 I cromosomi eucariotici contengono multiple origini di replicazione

Le forcelle di replicazione eucariotiche si muovono di circa 50 nucleotidi al secondo a causa dello stato più denso della cromatina. Si è notato come sono presenti da 30·000 a 500·000 origini di replicazione per la divisione di una cellula umana in modo che una cellula possa coordinare le regioni attive con altre caratteristiche dei cromosomi come i geni che sono espressi. Le forcelle di replicazione sono formate in coppie e creano una bolla di replicazione che si muove in direzioni opposte da un punto di origine comune, che si fermano quando collidono con un'altra forcella che si muove nella direzione opposta. Molte forcelle di replicazione operano indipendentemente su ogni cromosoma e formano due eliche di DNA figlie complete.

#### 6.3.4 Negli eucarioti la replicazione del DNA avviene durante un'unica parte del ciclo vitale della cellula

Se i batteri replicano il DNA quasi continuamente negli eucarioti questo avviene solo durante la fase S o di sintesi del DNA che in una cellula mammifera dura circa 8 ore, al termine delle quali ogni cromosoma è stato completamente replicato in due coppie che rimangono unite al centromero fino alla fase M.

#### 6.3.5 Diverse regioni dello stesso cromosoma si replicano a tempi distinti durante la fase S

Le origini di replicazione non sono tutte attivate allo stesso momento, ma in cluster di circa 50 adiacenti, ognuna delle quali è replicata durante una piccola parte della fase S. L'ordine di attivazione dipende in parte dalla struttura cromatinica: l'eterocromatina nell'ultima parte della fase in quanto il ritardo potrebbe essere in relazione con la decondensazione della cromatina. Le forcelle di replicazione si muovono ad una velocità costante durante la fase S.

#### 6.3.6 Un grande complesso multisubunità si lega alle origini di replicazione eucariotiche

La maggior parte delle sequenze di DNA che servono come origini di replicazione contengono un sito di legame per una proteina iniziatrice formata da molte subunità detta ORC (origin recognition complex), una lunghezza di DNA ricca in A e T e almeno un sito di legame per proteine che facilitano il legame dell'ORC modificando la struttura cromatinica. Per garantire che tutto il DNA sia copiato una e una sola volta le elicasi replicative vengono caricate sequenzialmente sulle origini e attivate per iniziare la replicazione del DNA. Durante la fase  $G_1$  le elicasi replicative sono caricate sul DNA vicino all'ORC per creare un complesso prereplicativo. Successivamente proteine chinasi specializzate attivano l'elicasi. L'apertura della doppia elica permette il caricamento delle proteine di replicazione rimanenti come la DNA polimerasi. La chinasi che inizia la replicazione previene l'assemblaggio di nuovi complessi prereplicativi fino alla fase M successiva che resetta l'intero ciclo. Lo fanno fosforilando l'ORC rendendolo incapace di accettare nuove elicasi. Pertanto il complesso prereplicativo può formarsi unicamente nella fase  $G_1$ , mentre può essere attivato e disassemblato unicamente nella fase S, due fasi mutualmente esclusive.

### 6.3.7 Nuovi nucleosomi sno assemblati in coda alla forcella di replicazione

La duplicazione dei cromosomi richiede la sintesi e assemblaggio di nuove proteine cromosomiche sul DNA in coda alla forcella. La cellula richiede un gran numero di nuove proteine istoni in massa uguale al nuovo DNA sintetizzato per creare nuovi nucleosomi in ogni ciclo della cellula, pertanto gli organismi possiedono multiple copie del gene per ogni istone. Gli istoni sono sintetizzati principalmente nella fase S, quando il livello di mRNA istone aumenta di 50 volte e degradati alla fine della fase. Le proteine istone invece sopravvivono per l'intera vita della cellula. Il collegamento tra sintesi del DNA e degli istoni riflette un meccanismo di feedback che monitora il livello degli istoni liberi per garantire che le quantità siano uguali. Quando la forcella di replicazione avanza e passa attraverso nucleosomi del genitore e pertanto è richiesto un complesso di rimodellamento della cromatina che destabilizza le interfacce DNA-istone in modo che la forcella possa superarle efficientemente. Mentre la forcella di replicazione passa attraverso la cromatina gli istoni sono spostati. Quando un nucleosoma viene attraversato l'ottamero istonico viene rotto in un tetramero H3-H4 e due dimeri H2A-H2B. Il primo rimane associato debolmente con il DNA ed è distribuito a caso tra uno dei due figli, mentre i dimeri sono completamente rilasciati. I tetrameri H3-H4 recentemente sintetizzati sono aggiunti per riempire i buchi e i dimeri, metà vecchi e metà nuovi sono aggiunti a caso per completare il nucleosoma. La lunghezza dei frammenti di Okazaki è determinata dal punto in cui la DNA polimerasi è bloccata da un nuovo nucleosoma. Pertanto i frammenti hanno la stessa lunghezza della lunghezza di ripetizione dei nucleosomi. L'aggiunta di tetrameri e dimeri richiede accompagnatori istoni o fabbriche di assemblaggio di cromatina, complessi a multisubunità che si legano agli istoni altamente basici e li rilasciano solo nel contesto appropriato e diretti verso il DNA appena replicato attraverso un'interazione specifica con il morsetto PCNA che rimangono sul DNA abbastanza a lungo per permettere ai primi di completare il loro compito.

### la telomerasi replica la fine dei cromosomi

Il meccanismo di replicazione del filamento in ritardo incontra problemi alla fine del cromosoma lineare: l'RNA primer finale non può essere sostituito dal DNA in quanto non c'è una fine 3' – OH disponibile per la polimerasi di riparazione. I batteri risolvono il problema con DNA circolare, mentre gli eucarioti attraverso sequenze specializzate dette telomeri che contengono molte ripetizioni di sequenze corte (negli umani GGGTTA), riconosciute dall'enzima telomerasi che le rifornisce ogni volta che la cellula si divide. La telomerasi riconosce la fine di un telomero e la allunga nella direzione 5'-3' utilizzando uno stampo a RNA componente dell'enzima stesso per sintetizzare nuove copie della ripetizione. Dopo l'estensione del filo genitore della telomerasi la replicazione del filamento in ritardo può essere completata dalla DNA polimerasi standard che usa queste estensioni per sintetizzare il filamento complementare.

### 6.3.8 I telomeri sono condensati in strutture specializzate che proteggono la fine dei cromosomi

I telomeri devono essere in grado di distinguere tra le rotture accidentali del cromosoma: una nucleasi specializzata si attacca alla fine 5' dei cromosomi lasciando protrudere un singolo filamento che in combinazione con le ripetizioni di GGGTTA attrae un gruppo di proteine che formano un cappuccio detto shelterin che nasconde i telomeri dai rivelatori di danni che monitorano il DNA: la fine protrudente del DNA si infila nella sequenza ripetuta del telomero. Questi anelli a T sono regolati dal shelterin e proteggono ulteriormente la fine del cromosoma.

### 6.3.9 La lunghezza dei telomeri è regolata dalla cellula e dagli organismi

Essendo i processi che fanno crescere e riducono ogni sequenza di telomeri sono bilanciati approssimativamente una fine cromosomica contiene un numero variabile di ripetizioni telomeriche. Molte cellule hanno un numero di meccanismi omeostatici che mantengono il numero di queste ripetizioni in intervallo. Nella maggior parte delle divisioni cellulari i telomeri si accorciano gradualmente per limitare la proliferazione di cellule ribelli nei tessuti adulti. Le ripetizioni telomeriche sono erose in quantità diverse in tipi diversi di cellule grazie a un enzima che non può tenere il passo con la duplicazione. Dopo molte generazioni le cellule discendenti avranno cromosomi senza funzione telomerica e attivano una risposta ad danni del DNA che causa la fine del ciclo cellulare e della duplicazione in un processo detto replicative cell senescence.

## 6.4 Riparazione del DNA

Mantenere la stabilità genetica per la vita richiede meccanismi per riparare le lesioni accidentali che continuano ad avvenire. La maggior parte dei cambi nel DNA sono immediatamente corretti da un insieme di proteine dette DNA repair. Solo lo 0.02% dei cambi del DNA si accumula come permutazioni permanenti. Una grande parte della capacità di codifica dei genomi è utilizzata per queste proteine.

### 6.4.1 Senza la riparazione del DNA, danni spontanei cambierebbero rapidamente la sequenza del DNA

Nonostante il DNA sia un materiale altamente stabile è una molecola organica complessa suscettibile a cambi spontanei che porterebbero a mutazioni se lasciate non riparate. Il DNA della cellula umana perde circa 18'000 basi purine ogni giorno a causa dei legami N-glicosilici al deossiribosio si idrolizzano nella depurinazione. Similmente una deaminazione della citosina nell'uracile accade a circa 100 basi per cellula al giorno. Le basi del DNA possono essere danneggiate da una collisione con metaboliti reattivi prodotti nella cellula come forme reattive di ossigeno o il donatore di metile S-adenosilmetionine o da radiazioni ultraviolette. Se lasciate non corrette queste modifiche porterebbero alla cancellazione di basi o a sostituzioni durante la replicazione con conseguenze disastrose.

### 6.4.2 La doppia elica del DNA è prontamente riparata

La struttura a doppia elica del DNA è adatta alla riparazione in quanto porta due copie dell'informazione genetica, pertanto quando un filamento è danneggiato quello complementare mantiene una copia intatta della stessa informazione che viene utilizzata durante la riparazione. I tipi di processi di riparazione presentati non possono operare su DNA o RNA a singolo filamento.

### 6.4.3 Danni al DNA possono essere rimossi attraverso molti cammini

Le cellule hanno diversi modi per riparare il DNA utilizzando diversi enzimi che agiscono in base alla lesione. Nei due cammini più comuni il danno è asportato e la sequenza originale è ripristinata dalla DNA polimerasi utilizzando il filamento non danneggiato come stampo. Differiscono riguardo a come eliminano il danno. Nel primo cammino detto riparazione tramite asportazione della base coinvolge un insieme di enzimi detti DNA glicosilasi che possono riconoscere un tipo specifico di base nel DNA e catalizzare la rimozione idrolitica. Ne esistono al meno sei tipi, come quelli che rimuovono

C e A deaminate, diversi tipi di basi alcalinate o ossidate, basi con anelli aperti e quelle in cui un doppio legame carbonio-carbonio è stato convertito in un legame singolo. Una base alterata viene riconosciuta in un processo con un passo base in cui avviene un flip-out del nucleotide danneggiato dall'elica mediato da un enzima che permette alla DNA glicosilasi di controllare tutte le parti della base per danni. Questi enzimi si muovono lungo tutto il DNA alla ricerca di danni. Quando viene trovata una base danneggiata la rimuove dallo zucchero. Il buco creato dalla DNA glicosilasi è riconosciuto da un enzima detto AP endonucleasi che taglia il backbone a fosfodiester e il vuoto lasciato è riparato. La depurinazione lascia uno zucchero con una base mancante e sono pertanto riparate direttamente con un AP endonucleasi. Il secondo cammino principale è detto riparazione a asportazione del nucleotida. Questo meccanismo può riparare danni causati da grandi cambi nella struttura della doppia elica. Un grande complesso multienzimico scansiona il DNA alla ricerca di una distorsione della doppia elica e la DNA elicasi rimuove il singolo filamento contenente la lesione. Il grande gap prodotto è riparato dalla DNA polimerasi e dalla DNA ligasi. Un'alternativa a questi meccanismi è la reversione chimica diretta, utilizzata selettivamente per la rapida rimozione di lesioni altamente mutageniche o citotossiche.

##### **6.4.4 L'accoppiamento di riparazione a asportazione di nucleotidi con la trascrizione garantisce che il DNA più importante è riparato efficientemente**

Tutto il DNA è sotto costante sorveglianza per i danni, ma le cellule possiedono modi per direzionare la riparazione a sequenze che sono richieste urgentemente. Lo fanno legando la polimerasi al cammino di riparazione a asportazione di nucleotidi. La RNA polimerasi stalla a queste lesioni e attraverso proteine di accoppiamento direziona il complesso di riparazione a questi siti. La RNA polimerasi viene poi fatta riprendere da dove si era fermata con una reazione complessa.

##### **6.4.5 La chimica delle basi del DNA facilita l'individuazione dei danni**

La doppia elica del DNA sembra ottima per la riparazione e la natura delle quattro basi rende chiara la distinzione tra quelle danneggiate e non danneggiate. Ogni possibile deaminazione presente nel DNA presenta una base innaturale che può essere riconosciuta e rimossa da una DNA glicosilasi specifica. L'uracile nell'RNA è stato sostituito dalla timina in quanto se no il sistema di riparazione non sarebbe stato in grado di distinguere una C deaminata da una U naturale.

##### **6.4.6 Speciali translesion della DNA polimerasi sono usate durante emergenze**

Se il DNA soffre di danni pesanti si rende necessaria una strategia diversa: la replicazione della DNA polimerasi si ferma quando incontra DNA danneggiato e in emergenze vengono utilizzate altre polimerasi, versatili ma meno accurate dette translesion polimerasi per replicare attraverso il danno. Tali polimerasi possono riconoscere un tipo specifico di danno e aggiungere i nucleotidi necessari per ripristinare la sequenza iniziale. Altri fanno congetture educate. Non hanno un meccanismo di proofreading esonucleico e sono meno discriminanti nella scelta dell'incorporazione del nucleotide e possono pertanto aggiungere solo pochi nucleotidi alla volta. Queste polimerasi portano rischi alla cellula a causa della loro imprecisione.

### 6.4.7 Rotture nel doppio filamento sono efficientemente riparate

Un tipo di danno al DNA pericoloso avviene quando entrambi i filamenti della doppia elica sono rotti, lasciando nessun filamento stampo capace di riparare il danno accuratamente. Due meccanismi vengono utilizzati per riparare a questi danni. Il primo è l'unione delle terminazioni non omologo in cui le terminazioni rotte sono riunite da DNA ligation con la perdita dei nucleotidi nel sito dell'unione. Questo meccanismo è comune delle cellule somatiche dei mammiferi. Ha come risultato una mutazione e può inoltre unire due parti del cromosoma non inizialmente insieme. Il secondo tipo è detto ricombinazione omologa. Entrambi vengono utilizzati, il secondo solo durante e poco dopo la replicazione quando cromatidi sorelle sono disponibili come stampo.

### 6.4.8 Danno al DNA ritarda la progressione del ciclo cellulare

Le cellule possono riconoscere molti tipi di danni e gli enzimi sono rendono efficiente la riparazione ritardando la progressione del ciclo cellulare fino a che la riparazione è completa. Questi ritardi facilitano la riparazione rendendo disponibile il tempo necessario affinché i meccanismi si completino. I danni risultano anche in una maggiore sintesi di enzimi di riparazione attraverso speciali proteine di segnale che percepiscono i danni e regolano l'espressione genica.

### 6.4.9 Riparazione omologa

Un ulteriore meccanismo di riparazione del DNA è detta ricombinazione omologa, la cui caratteristica fondamentale è uno scambio di filamenti di DNA tra un paio di duplex omologhi della sequenza di DNA, segmenti di doppia elica che sono molto simili nella sequenza nucleotidica. Questo scambio permette a una lunghezza di DNA duplex di essere stampo per ripristinare informazioni perse o danneggiate. Essendo non legata al filamento complementare al danno è molto versatile. È il modo in cui vengono riparate rotture a doppio filamento. Questi danno la maggior parte delle volte nascono quando la forcella di replicazione si stalla o viene rotta indipendentemente. La ricombinazione omologa corregge questi errori ed è fondamentale per ogni cella proliferante. È il meccanismo più versatile. Inoltre durante la meiosi catalizza lo scambio di informazioni genetiche tra cromosomi omologhi materni e paterni creando nuove combinazioni di sequenza di DNA passata ai discendenti.

### 6.4.10 L'accoppiamento delle basi del DNA guida la ricombinazione omologa

La ricombinazione omologa avviene unicamente tra duplex di DNA che hanno regioni di sequenza estensiva omologhe. I due duplex che stanno svolgendo ricombinazione omologa testano la sequenza dell'altro da un duplex e un singolo filamento dell'altro. La corrispondenza non deve essere perfetta ma deve essere molto simile. L'interazione di accoppiamento di basi può essere imitato permettendo a una doppia elica di riformarsi dai singoli filamenti nel processo di DNA rinaturazione o ibridazione che avviene quando rare collisioni giustappungono nucleotidi complementari su due filamenti singoli, permettendo la creazione di corte lunghezze di doppia elica. Questa parte lenta è seguita da un zippering in cui la regione di doppia elica si estende per massimizzare il numero di regioni che accoppiano le basi. L'ibridazione può creare regioni di doppia elica consistenti di filamenti che si sono originati da molecole duplex diverse se sono similmente complementari. La ricombinazione omologa avviene attraverso un insieme di reazioni che permettono a due duplex del DNA di campionare le rispettive sequenze senza dissociarsi in singoli filamenti.

##### 6.4.11 La ricombinazione omologa può riparare perfettamente rotture a doppio filamento nel DNA

La ricombinazione omologa può riparare doppi filamenti accuratamente, senza alcuna perdita o alterazione al sito di riparazione. Per fare questo il DNA danneggiato deve essere avvicinato ad un DNA omologo che serve da stampo per la riparazione. Per questa ragione il processo avviene dopo la replicazione, che può creare rischi di incidenti che richiedono questa riparazione. Il cammino più semplice per la riparazione mostra come il duplex di DNA danneggiato e l'omologo si ingarbugliano in modo che uno dei filamenti danneggiati può usare il complementare del duplex intatto come stampo. Le terminazioni del DNA danneggiato sono resecati da nucleasi per produrre terminazioni 3' a filamento singolo protrudenti. Successivamente avviene uno scambio di filamenti o strand invasion in cui una terminazione a filamento singolo 3' dalla molecola danneggiata entra nel duplex stampo e cerca per la sequenza omologa attraverso accoppiamento delle basi. Dopo che questo è avvenuto una polimerasi estende il filamento invasore utilizzando l'informazione fornita dallo stampo ripristinando l'informazione. Infine i filamenti si separano, avviene altra sintesi di riparazione e la legatura ripristina le doppie eliche originali e completano il processo di riparazione.

##### 6.4.12 Il cambio di filamenti avviene grazie alla proteina RecA/Rad51

La proteina che compie lo scambio di filamenti avviene grazie alla proteina RecA in *E. coli* e Rad51 negli eucarioti. Per catalizzare lo scambio la proteina si lega cooperativamente al filamento invasore, facendo in modo che l'insieme formi i DNA in gruppi di tre nucleotidi consecutivi mantenuti come in una doppia elica convenzionale ma tra triplette adiacenti il backbone è svolto e allungato. Questo filamento si lega al duplex DNA in modo da allungarlo, destabilizzandolo e rendendo la separazione dei filamenti semplice. Il filamento invasore può poi campionare la sequenza attraverso accoppiamento di basi, che avviene in blocchi di triplette: se una corrispondenza è trovata si prova la prossima tripletta e così via. RecA idrolizza l'ATP e i passaggi precedenti richiedono che ogni monomero di RecA siano legati con esso, ma la ricerca non richiede l'idrolisi, avvenendo per collisione molecolare. Una volta che la reazione è completa l'idrolisi è necessaria per disassemblare RecA dal complesso delle molecole di DNA e la DNA polimerasi e DNA ligasi possono completare il processo di riparazione.

##### 6.4.13 La ricombinazione omologa può salvare forcelle di replicazione danneggiate

Il ruolo più importante della ricombinazione omologa è che salva forcelle di replicazione in satello o danneggiato. Una causa possono essere intervalli vuoti nell'elica di DNA sopra la forcella che quando sono raggiunti causano la sua rottura.

##### 6.4.14 Le cellule regolano l'uso della ricombinazione omologa durante la riparazione del DNA

La ricombinazione omologa presenta dei rischi alla cellula in quanto può riparare i danni usando bit di genoma sbagliati come stampo. Un cromosoma umano danneggiato può essere riparato usando l'omologo da altri parenti invece che dalla cromatide sorella, convertendo la sequenza del DNA riparato dalla sequenza materna a quella paterna o viceversa. Questo causa una perdita di eterozigotità, passaggio critico nella formazione di molti tumori. Per minimizzare il rischio di errori ogni passaggio del processo è regolato. Il processamento delle terminazioni danneggiate è coordinato con

la divisione cellulare: gli enzimi di nucleasi che causano questo processo sono attivati in parte da fosforilazione solo nelle fasi S e  $G_2$  quando un duplex filgio può essere utilizzato come stampo per la riparazione aiutato dalla prossimità. Il caricamento di RecA e Rad52 è controllato attraverso una serie di proteine accessori come Rad52 in modo da rendere il processo accurato ed efficiente. Sono create ad alto livello negli eucarioti e trasportate nel nucleo in una forma inattiva. Quando avviene un danno convergono rapidamente nel sito e attivano.

##### 6.4.15 La ricombinazione omologa è cruciale per la meiosi

La ricombinazione omologa viene usata come metodo per trasportare nuove combinazioni di geni come risultato di scambio di materiale genetico tra diversi cromosomi, parte necessaria per la meiosi. Diventa parte integrale del processo dove cromosomi sono separati a cellule germinali producendo cromosomi attraverso crossing-over e conversione dei geni, creando cromosomi ibridi che contengono informazioni sia materne che paterni. Il processo di base è lo stesso.

##### 6.4.16 La ricombinazione meiotica comincia con una rottura a doppio filamento programmata

La ricombinazione meiotica inizia con una proteina specializzata che rompe entrambi i filamenti dell'elica in un cromosoma ricombinante e come una topoisomerasi rimane legata covalentemente al DNA rotto. Una nucleasi specializzata degrada la terminazione legata dalla proteina, rimuovendola e lasciando una terminazione 3' protrudente. Molte proteine specifiche alla meiosi direzionano le altre in modo da svolgere il processo in maniera leggermente diversa. Inoltre la ricombinazione avviene tra cromosomi materni e paterni.

##### 6.4.17 Le giunzioni di Holliday sono formate durante la meiosi

Nella meiosi un intermedio detto giunzione di Holliday o scambio cross-strand può adottare multiple conformazioni e un insieme speciale di proteine di ricombinazione che vi si legano e stabilizzano l'isomero simmetrico aperto. Le proteine che legano le giunzioni possono catalizzare una reazione di branch migration dove il DNA è avvolto attraverso la giunzione rompendo e riformando accoppiamenti di basi. Le giunzioni utilizzano l'idrolisi dell'ATP per espandere le regioni di DNA eteroduplex inizialmente create dalla reazione di scambio di filamento. Tipicamente si formano in coppie.

##### 6.4.18 La ricombinazione omologa produce crossover e non-crossover durante la meiosi

Durante la meiosi ci sono due risultati della ricombinazione omologa: negli umani il 90% delle rotture del doppio filamento si risolvono come non-crossover: i due duplex si separano in una forma non alterata ad eccezione di una regione eteroduplex vicino al sito della rottura. L'altro risultato è dovuto alla formazione e taglio di una doppia giunzione di Holliday con un enzima specializzato per la formazione di un crossover: le due porzioni originali di ciascun cromosoma sono scambiate, creando due cromosomi che hanno fatto crossover. I crossover che si formano sono distribuiti nel cromosoma in modo che il crossover in una posizione inibisce crossing-over nelle regioni vicine. Il meccanismo regolatorio associato non è capito ma viene detto crossover control garantisce la distribuzione costante dei punti di crossover e assicura che ogni cromosoma svolga un crossover ogni meiosi. Circa due crossover avvengono per cromosoma e svolgono importante ruolo meccanico nella segregazione dei cromosomi durante la meiosi. La macchina di ricombinazione lascia una regione



eteroduplex con un singolo filamento con la sequenza di DNA dell'omologo paterno accoppiato con un filamento dell'omologo materno. Queste regioni possono tollerare un piccolo numero di coppie non corrispondenti, marcando siti di conversione genica potenziale.

### 6.4.19 La ricombinazione omologa solitamente causa conversione genica

È una regola fondamentale della genetica che a parte il DNA mitocondriale, ereditato dalla madre, ogni genitore faccia una contribuzione genetica uguale alla discendenza. Un insieme completo di geni è ereditato dalla madre e uno dal padre. Sottostante a questa legge si trova l'accurata separazione di cromosomi a cellula germinali durante la meiosi. Quando una cellula diplode in un genitore fa la meiosi per produrre quattro cellule germinali aploidi metà dei geni dovrebbero essere materne e l'altra metà paterne. Versioni alternative dello stesso gene sono dette alleli e la divergenza dalla distribuzione aspettata durante la meiosi è detta conversione genica, che avviene per una piccola parte del genoma e in molti casi parte di un gene viene cambiata. Molti cammini possono causare la conversione genica, ma uno dei più importanti nasce come conseguenza della ricombinazione durante la meiosi: se i due filamenti parte di una regione eteroduplex non possiedono sequenze identiche si formano coppie di basi errate e pertanto si utilizza a caso lo stampo materno o paterno per la riparazione e pertanto un allele sarà perso e l'altro duplicato.

## 6.5 Trasposizione e ricombinazione specifica al sito conservativa

L'ordine dei geni sui cromosomi che subiscono la ricombinazione omologa è lo stesso. La trasposizione e ricombinazione conservativa specifica al sito sono ricombinazioni che non richiedono regioni di DNA omologo sostanziale. Queste due reazioni di ricombinazioni possono alterare l'ordine dei geni in un cromosoma e tipi di mutazioni che introducono nuovi blocchi di DNA nel genoma. Sono dedicate a muovere una grande varietà di segmenti di DNA specializzati o elementi genetici mobili da una posizione all'altra. Spesso uno di questi geni codifica un enzima che catalizza il movimento dell'elemento rendendo questa ricombinazione possibile. Tutte le cellule contengono elementi genetici mobili che hanno avuto un ruolo evolutivo profondo nella formazione del genoma odierno. Sono considerati parassiti molecolari che persistono in quanto la cellula non può eliminarli, ma possono anche creare benefici a essa. Il loro movimento produce molte varianti genetiche da cui dipende l'evoluzione in quanto possono riordinare sequenze vicine dell'ospite.

### 6.5.1 Attraverso la transizione gli elementi genetici mobili possono inserirsi in ogni sequenza di DNA

Gli elementi che si muovono per trasposizione sono detti trasposoni. In questo processo un enzima specifico codificato dal trasposone detto trasposasi agisce su una sequenza di DNA specifica a ogni fine del trasposone causando il suo inserimento in un nuovo sito di DNA obiettivo. La maggior parte sono poco selettivi e possono inserirsi in locazioni diverse nel genoma. La maggior parte si muovono raramente. I trasposoni possono essere classificati in base a struttura e meccanismo di trasposizione in trasposoni solo a DNA retrotrasposoni simili ai retrovirus e retrotrasposoni non retrovirali.

### 6.5.2 Trasposoni solo a DNA si possono muovere attraverso un meccanismo a copia-incolla

Questi trasposoni esistono unicamente come DNA durante il loro movimento sono predominanti nei batteri e sono responsabili per la resistenza agli antibiotici. Nonostante questi elementi mobili possono traspirarsi solo all'interno della cellula si possono muovere tra cellule attraverso il trasferimento di geni orizzontale. Una volta introdotti nella nuova cellula si possono inserire nel genoma ed essere passati a tutta la progenie. Si possono spostare da un sito donatore a uno obiettivo attraverso una trasposizione copia-incolla dove il trasposone è asportato da un luogo del genoma e inserito in un altro. La reazione produce una corta duplicazione della sequenza obiettivo al sito di inserzione, ripetizioni che affiancano il trasposone servono come records di eventi di trasposizione precedenti. Quando il trasposone è asportato dal luogo originale lascia un buco nel cromosoma la cui lesione può essere riparata da un sistema di riparazione a doppio filamento se il cromosoma è stato appena replicato e esiste una copia dell'ospite. Alternativamente una reazione di unione delle terminazioni non omologa avviene e la sequenza del DNA a fianco del trasposone è modificata, producendo una mutazione nel sito da cui il trasposone si è asportato.

### 6.5.3 Alcuni virus usano un meccanismo di trasposizione per muoversi nel cromosoma della cellula ospite

Alcuni virus sono considerati elementi genetici mobili in quanto usano meccanismi di trasposizioni per integrare i loro genomi nella cellula ospite, codificando proteine che incapsulano la loro informazione genetica in particelle virali che possono infettare altre cellule. La trasposizione ha un ruolo fondamentale nel ciclo vitale dei virus come i retrovirus: al di fuori della cellula esistono come un singolo filamento di RNA incapsulato in un capsido proteico con un enzima trascrittasi inversa. Durante il processo di infezione l'RNA virale entra la cellula ed è convertito in un doppio filamento di DNA dall'enzima che può polimerizzare DNA su uno stampo a RNA o DNA. Il termine retrovirus si riferisce alla capacità del virus di invertire il flusso delle informazioni genetiche. Una volta che la trascrittasi inversa ha prodotto la molecola di DNA sequenze specifiche nelle sue terminazioni sono riconosciute da una trasposasi chiamata integrasi che lo inserisce nel cromosoma con un meccanismo simile dai trasposoni solo a DNA.

### 6.5.4 Retrotrasposoni simili ai retrovirus assomigliano ai retrovirus ma non possiedono un capsido

Una famiglia di trasposoni detta retrotrasposoni simili ai retrovirus si muove dentro e fuori i cromosomi con un meccanismo simile a quello dei retrovirus. Questi elementi sono presenti in organismi diversi e non hanno la capacità di lasciare la cellula residente. Il primo passo nella loro trasposizione è la trascrizione dell'intero trasposone, producendo una copia a RNA ed l'elemento che è tradotto come un RNA messaggero nella cellula codifica un enzima di trascrittasi inversa che forma una copia di DNA a doppia elica della molecola di RNA attraverso un intermedio RNA-DNA ibrido che viene poi integrato nel cromosoma con un enzima di integrasi codificato dall'elemento.

### 6.5.5 Una grande frazione del genoma umano è composta di retrotrasposoni nonvirali

Una frazione significativa di molti cromosomi dei vertebrati è composta da sequenze di DNA ripetute che, negli umani, sono principalmente versioni mutate e troncate di retrotrasposoni nonvirali.

La maggior parte sono ormai immobili, ma una piccola parte è ancora capace di muoversi. Questi trasposoni si muovono attraverso un meccanismo che richiede un complesso di endonucleasi e trascrittasi inversa. L'RNA e la trascrittasi hanno un ruolo più diretto nell'evento di ricombinazione. Alcuni non portano la propria endonucleasi o trascrittasi inversa.

### 6.5.6 La ricombinazione conservativa specifica al sito può riordinare il DNA reversibilmente

Questo tipo di meccanismo di ricombinazione riordina altri tipi di elementi di DNA mobili. In questo cammino rottura e riunione avvengono a due siti speciali, uno in ogni molecola di DNA partecipante. In base alla posizione e al relativo orientamento dei due siti può accadere integrazione, esportazione o inversione del DNA. Questo processo avviene grazie a un enzima che rompe e riunisce due doppie eliche a sequenze specifiche. Lo stesso enzima che le unisce le può rirompere, ripristinando la sequenza delle due molecole di DNA originale. Viene spesso utilizzato dai virus a DNA per muovere i loro genomi in quelli della cellula ospite. Quando è integrato il DNA virale viene replicato ed è passato a tutte le cellule discendenti. Se la cellula ospite subisce danni il virus può invertire la reazione di ricombinazione, asportare il proprio genoma e incapsularlo in una particella di virus. Sono molte le differenze con la trasposizione: richiede sequenze specializzate sia sul donatore che sul recipiente che contengono siti di riconoscimento per la ricombinasi che catalizzerà il riordinamento. I meccanismi di reazione sono fondamentalmente diversi: la ricombinasi che catalizza la reazione assomiglia alla topoisomerasi in quanto forma legami covalenti temporanei con il DNA e usa questa energia per completare il riordinamento del DNA. Tutti i legami fosfati utilizzati durante un evento di ricombinazione sono ripristinati al completamento. La ricombinazione conservativa specifica al sito viene utilizzata da molti batteri per controllare l'espressione di un gene specifico.

## Capitolo 7

# Come una cellula legge il genoma, dal DNA alle proteine

Il DNA nel genoma usa l'RNA come intermediario nella sintesi delle proteine. Quando una cellula necessita una proteina utilizza la sequenza appropriata della catena nucleotidica copiandola in RNA durante la trascrizione che direziona direttamente la sintesi della proteina durante la traduzione. Esistono varianti di questo processo in cui i trascritti a RNA vengono processati nel nucleo con processi come RNA splicing prima che posano uscire da esso. Questi cambi possono cambiare il significato di una molecola di DNA. Per molti geni inoltre il prodotto finale è RNA. I genomi di organismi multicellulari sono disordinati con corti esoni e lunghi introni. Sezioni che codificano il DNA sono separate da lunghe sequenze senza apparente significato.

### 7.1 Dal DNA all'RNA

Essendo che molte copie identiche dello stesso RNA possono essere completate dallo stesso gene ogni molecola di RNA può guidare la sintesi di molte proteine identiche, ma i geni sono trascritti e tradotti a tassi diversi, permettendo la cellula di avere vaste quantità di alcune proteine e piccole di altre. Inoltre la cellula regola l'espressione di ognuno dei suoi geni secondo i suoi bisogni, controllando la produzione del suo RNA.

#### 7.1.1 Le molecole di RNA hanno un unico filamento

Il primo passo nella lettura delle istruzioni geniche è la copia di una particolare sequenza della sequenza di nucleotidi in una a RNA. L'informazione nell'RNA è scritta nello stesso linguaggio del DNA e questo processo è pertanto detto trascrizione. L'RNA è un polimero lineare composto da quattro tipi di subunità nucleotidiche legate da legami a fosfodiesteri. Differisce dal DNA in quanto i nucleotidi nell'RNA sono ribonucleici, ovvero contengono ribosio e contiene la base uracile invece della timina che si può legare all'adenina. La struttura complessiva è molto diversa: l'RNA è a filamento singolo e una catena può piegarsi in una forma simile a una proteina permettendogli di avere precise funzioni strutturali e catalitiche.

### 7.1.2 La trascrizione produce RNA complementare a un filamento di DNA

L'RNA è sintetizzato attraverso la trascrizione del DNA, che comincia con l'apertura e lo svolgimento di una piccola porzione della doppia elica che espone le basi sui filamenti, uno dei quali agisce come stampo per la sintesi della molecola di RNA. La sequenza di nucleotidi è determinata dall'accoppiamento di basi complementari tra i nucleotidi che arrivano e lo stampo. Quando avviene una corrispondenza il ribonucleide che arriva è legato covalentemente con la catena crescente in una reazione catalizzata da enzimi. La catena è allungata un nucleotide alla volta e possiede una sequenza complementare allo stampo. Il filamento di RNA non rimane legato con lo stampo ma è separato dietro al regione dove i nucleotidi sono aggiunti causando il rilasciamento come singolo filamento. Le molecole di RNA sono inoltre molto più corte rispetto le molecole di DNA.

### 7.1.3 L'RNA polimerasi causa la trascrizione

Gli enzimi che svolgono la trascrizione sono detti RNA polimerasi e catalizzano la formazione del legame fosfodiesterico che lega i nucleotidi muovendosi lungo il DNA, svolgendo l'elica sopra il sito attivo per la polimerizzazione. La catena di RNA è estesa nella direzione 5'-3'. I substrati sono ribonucleoside trifosfato la cui idrolizzazione fornisce l'energia necessaria alla reazione. Il rilascio immediato dell'RNA significa che le copie possono essere create in poco tempo, con la sintesi di molecole addizionali che inizia prima che quelle prime siano completate. L'RNA polimerasi catalizza l'unione di ribonucleidi e può cominciare una catena di RNA senza un primer. L'RNA polimerasi fa un errore una volta ogni  $10^4$  nucleotidi e le conseguenze di tale errore sono meno significative. Inoltre la stessa RNA polimerasi che comincia una molecola di RNA deve finirla senza dissociarsi dallo stampo. L'RNA polimerasi contiene un meccanismo di proofreading: se un ribonucleotide è aggiunto la polimerasi può indietreggiare e il sito attivo svolge una reazione di asportazione dove una molecola d'acqua sostituisce il pirofosfato ed è rilasciata una molecola di monofosfato.

### 7.1.4 Le cellule producono diverse categorie di molecole di RNA

La maggior parte dei geni trasportati in un DNA della cellula specificano la sequenza di amminoacidi della proteina e le molecole di RNA che sono copiate da questi geni sono detti RNA messaggeri o mRNA. Il prodotto finale di altri geni è la molecola di RNA, detti RNA non codificanti che servono come componenti strutturali, enzimatiche e regolatore per molti processi. Molecole di RNA piccolo nucleare o snRNA direzionano lo splicing di pre-mRNA per formare mRNA, l'RNA ribosomiale o rRNA forma il nucleo del ribosoma e i transfer RNA o tRNA forma gli adattatori che selezionano gli amminoacidi e li mantengono in posizione. I microRNA o miRNA e RNA piccolo interferente siRNA servono come regolatori per l'espressione genica e RNA piwi-interagente o piRNA protegge le linee germinali dai trasposoni. I long noncoding RNA o lncRNA con funzione di impalcature e regolano diversi processi cellulari come l'inattivazione del cromosoma X. Ogni segmento di DNA trascritto è detto unità di trascrizione che tipicamente possiede le informazioni di un gene. La maggior parte dell'RNA nella cellula è rRNA/

### 7.1.5 Segnali codificati nel DNA indicano l'RNA polimerasi dove iniziare e dove finire

Per trascrivere un gene accuratamente la RNA polimerasi deve riconoscere dove iniziare e finire sul genoma. Questo avviene in maniera diversa rispetto a batteri ed eucarioti. L'iniziazione di

una trascrizione è il punto in cui la cellula regola quali proteine devono essere prodotte e a quale velocità. L'RNA polimerasi batterica è un complesso a multisubunità che sintetizza l'RNA. Una subunità addizionale detta fattore sigma  $\sigma$  associa con l'enzima nucleo e lo assiste nella lettura dei segnali nel DNA che dicono dove iniziare la trascrizione. Il fattore  $\sigma$  e l'enzima di nucleo formano un oloenzima RNA polimerasi che aderisce debolmente al DNA batterico quando collidono e scivola rapidamente lungo il DNA fino a dissociarsi. QUando l'oloenzima arriva a una sequenza speciale che indica il punto di inizio per la sintesi di RNA detto protomero si lega fortemente in quanto il fattore  $\sigma$  crea contatto specifico con i limiti della base esposti all'esterno nella doppia elica. L'oloenzima polimerasi al protomero apre la doppia elica esponendo una piccola lunghezza di nucleotidi su ogni filamento chiamata bolla di trascrizione (di circa 10 nucleotidi), stabilizzata dal legame con il fattore  $\sigma$  con le basi non accoppiate. L'altro filamento agisce come stampo per l'accoppiamento di basi con i ribonucleotidi che arrivano, uniti dalla polimerasi per iniziare la catena di RNA. I primi 10 nucleotidi sono sintetizzati attraverso un meccanismo di "scrunching" dove la RNA polimerasi rimane legata al protomero e tira il DNA nel suo sito attivo espandendo la bolla di trascrizione. Questo processo genera stress e le catene di RNA sono rilasciate e forzando la polimerasi a riiniziare la sintesi. Questo processo di iniziazione abortiva è superato e lo stress generato aiuta l'enzima a rompere l'interazione con il protomero e con il fattore  $\sigma$ . La polimerasi inizia a muoversi lungo il DNA sintetizzando l'RNA muovendosi di base in base espandendo la bolla e contraendola al retro. Si continua l'allungamento della catena fino a che l'enzima incontra un terminatore dove la polimerasi si ferma e rilascia la molecola di RNA e lo stampo a RNA. La polimerasi si riassocia con il fattore  $\sigma$  ed è libera di riiniziare un processo di trascrizione. La maggior parte dei segnali di terminazione nei batteri è formata da una stringa di coppie A-T precedute da una sequenza due volte simmetrica di DNA che quando trascritta forma una forcina attraverso l'accoppiamento di basi che aiuta il disengagement dell'RNA trascritto dal sito attivo.

### 7.1.6 I segnali di inizio e terminazione della trascrizione sono eterogenei nella sequenza nucleotidica

Le sequenze di inizio e fine sono codificate da sequenze in relazione, che riflettono aspetti del DNA che sono riconosciuti direttamente dal fattore  $\sigma$ . Queste caratteristiche formano una sequenza di nucleotidi consenzianti, una media di un gran numero di sequenze. Si possono anche riconoscere attraverso la frequenza relativa di basi in ogni posizione. Tale sequenza nei batteri varia in modo da determinare la forza dei geni (il numero di eventi di iniziazione per gene). Per i terminatori la struttura di base è quella che forma la forcina nell'RNA. Il filamento scelto per la sintesi dell'RNA dipende dall'orientamento del promotore.

### 7.1.7 L'iniziazione della trascrizione negli eucarioti richiede molte proteine

Gli eucarioti possiedono la RNA polimerasi I, II e III, simili strutturalmente e con subunità in comune, ma trascrivono diverse categorie di geni. La I e la III trascrivono geni che codificano tRNA, rRNA e piccoli RNA, la II trascrive la maggior parte dei geni, inclusi quelli che codificano le proteine. La RNA polimerasi II richiede molti fattori detti fattori di trascrizione generali e l'iniziazione avviene su DNA condensato in nucleosomi e forme superiori di struttura cromatinica.

### 7.1.8 La RNA polimerasi II richiede un insieme di fattori di trascrizione generali

I fattori di trascrizione generali aiutano a posizionare la polimerasi correttamente al promotore, a separare i due filamenti di DNA permettendo l'inizio della trascrizione e rilasciarla dal promotore per iniziare la modalità di allungamento. Sono generali in quanto richieste da tutti i promotori utilizzati dalla polimerasi II. Sono un insieme di proteine che interagiscono dette TFIIA, TFIIB, TFIIC, TFIIID e hanno funzione equivalente al fattore  $\sigma$ . Il processo di assemblaggio inizia quando TFIIID si lega a una sequenza di DNA a doppia elica detta TATA box attraverso la subunità TBP. Il legame causa una distorsione nel DNA della TATA box che crea una marcatura per la locazione di un promotore attivo. Altri fattori insieme alla RNA polimerasi II formano un complesso di iniziazione della trascrizione. Dopo che si è formato sul DNA promotore la RNA polimerasi II ottiene l'accesso al filamento stampo e TFIIH che contiene una DNA elicasi idrolizza l'ATP svolgendo il DNA. La RNA polimerasi II rimane al promotore sintetizzando corte lunghezze di RNA fino a subire una serie di cambi conformazionali che le permettono di spostarsi ed entrare nella fase di allungamento. In questa transizione viene aggiunto un gruppo fosfato alla coda della RNA polimerasi detto CTD (C-terminal domain). Durante l'iniziazione la serina nella quinta posizione della sequenza ripetuta è fosforilata da TFIIH che contiene una chinasi in una delle subunità. La polimerasi può poi disengaggiarsi dal cluster e subisce una serie di cambi conformazionali che le permettono di trascrivere per lunghe distanze senza dissociarsi dal DNA. Dopo che si è entrati nella fase di allungamento i fattori di trascrizione generali si separano per iniziare un altro processo.

### 7.1.9 La polimerasi II richiede attivatori, mediatori e proteine per la modifica della cromatina

L'inibizione della trascrizione negli eucarioti è complessa e richiede molte proteine. Innanzitutto delle proteine dette attivatori trascrizionali devono legarsi a speciali sequenze del DNA dette enhancers e aiutare ad attrarre l'RNA polimerasi II al punto d'inizio. Successivamente è necessario un complesso proteico detto mediatore che permette alle proteine attivatrici di comunicare con la Polimerasi II e con i fattori di trascrizione generali. Alla fine l'iniziazione della trascrizione richiede il reclutamento di enzimi modificatori della cromatina, complessi di rimodellizzazione e enzimi modificatori degli istoni che aumentano l'accesso al DNA nella cromatina. L'ordine di assemblaggio di queste proteine non segue un cammino preciso e differisce per gene. Per cominciare la trascrizione la RNA polimerasi II deve essere rilasciata da questo complesso attraverso proteolisi insito delle proteine attivatrici.

### 7.1.10 L'allungamento della trascrizione negli eucarioti richiede proteine accessorie

Una volta che l'RNA polimerasi ha iniziato la trascrizione si muove a scatti. RNA polimerasi allunganti sono associate con una serie di fattori di allungamento, proteine che diminuiscono la probabilità che l'RNA polimerasi si dissocia prima di aver finito la trascrizione. Si associano con la polimerasi dopo l'iniziazione. Quando l'RNA polimerasi si muove lungo un gene alcuni enzimi legati ad essa modificano gli istoni, lasciando una traccia, che potrebbe aiutare nelle trascrizioni successive e nella coordinazione dell'allungamento.

### 7.1.11 La trascrizione crea tensione superelicale

Il superavvolgimento del DNA è una conformazione che il DNA assume quando è presente tensione superelicale. Un grande superavvolgimento di DNA si forma per ogni 10 paia di nucleotidi svolti, processo energeticamente favorevole in quanto ripristina una torsione normale nella regione accoppiata rimanente. La RNA polimerasi crea tensione superelicale mentre si muove lungo una lunghezza di DNA ancorata alla terminazione, con tensione positiva davanti a lei e negativa dietro. Negli eucarioti questa tensione è rimossa dalla DNA topoisomerasi, mentre nei batteri una DNA girasi usa l'energia dell'idrolisi dell'ATP per imporre superavvolgimenti al DNA mantenendolo in costante tensione, ma in senso opposto da quello fornito dalla polimerasi.

### 7.1.12 L'allungamento della trascrizione è strettamente accoppiato con il processamento dell'RNA

Negli eucarioti la trascrizione è il primo passo per la produzione di una molecola di mRNA matura. Un passo successivo è la modificazione covalente delle terminazioni dell'RNA e la rimozione delle sequenze di introni nel processo di RNA splicing. La terminazione 5' viene incappucciata e la 3' attraverso polidenilazione che permettono alla cellula di capire se le terminazioni sono entrambe presenti prima che sia esportata dal nucleo e tradotta. L'RNA splicing permette di sintetizzare proteine diverse dallo stesso gene. La fosforilazione della coda CTD della polimerasi permette la disassociazione delle altre proteine e l'associazione di nuove. Alcune di queste si legano all'RNA che si sta sintetizzando processandolo.

### 7.1.13 L'incappucciamento dell'RNA è la prima modifica dei pre-mRNA eucariotici

Appena l'RNA polimerasi II ha prodotto circa 25 nucleotidi di RNA alla terminazione 5' della molecola viene aggiunto un cappuccio formato di una guanina modificata. Tre enzimi in successione svolgono la reazione necessaria: una fosfatasi rimuove il fosfato dalla terminazione, una guanil trasferasi aggiunge un GMP in un legame inverso e un metil trasferasi aggiunge un gruppo metile alla guanosina. Questi enzimi si trovano alla catena della RNA polimerasi fosforilata alla posizione Ser5. Questo cappuccio metile indica la terminazione 5' dell'mRNA eucariotico e lo differenzia da altri tipi di RNA.

### 7.1.14 L'RNA splicing rimuove le sequenze di introni dal pre-mRNA

Essendo i geni eucariotici dispersi in sequenze di introni che vengono trascritte insieme agli esoni si devono rimuovere i primi attraverso l'RNA splicing per produrre la proteina. Questo processo avviene alla produzione dell'mRNA. Ogni splicing rimuove un introne attraverso due reazioni di trasferimento di fosforile o transesterificazioni sequenziali che legano insieme gli esoni rimuovendo gli introni. Il macchinario che lo svolge è un complesso consistente di 5 molecole di RNA e centinaia di proteine. Idrolizza molte molecole di ATP per evento. La complessità assicura uno splicing accurato e flessibile. Oltre agli aspetti evolutivisti della divisione in domini delle proteine per la ricombinazione lo splicing permette di produrre un insieme di proteine diverse dallo stesso gene.

### 7.1.15 Sequenze di nucleotidi segnalano dove avviene lo splicing

Il macchinario dello splicing deve riconoscere tre porzioni della molecola di RNA precursore: il sito di splice 5', quello 3' e il ramo nella sequenza di introni che forma la base con il lazo. Ogni sito



possiede una sequenza di nucleotidi simile per ogni introne che fornisce indizi sul luogo di splicing. Queste sono corte e possono avere variabilità estesa e pertanto sono presenti altre informazioni per compiere la scelta ultima.

### 7.1.16 Lo splicing dell'RNA è svolto dallo spliceosoma

I passi fondamentali dello splicing sono svolti da molecole di RNA specializzate che riconoscono la sequenza che riconosce il luogo dello splicing e ne catalizza le reazioni chimiche. Sono molecole di circa 200 nucleotidi e sono U1, U2, U4, U5 e U6 e sono dette snRNA, insieme con almeno sette subunità proteiche per formare un snRNP (small nuclear ribonucleoprotein). Gli snRNP formano il nucleo dello spliceosoma, un complesso di RNA e proteine che svolge lo splicing. Durante la reazione di splicing il riconoscimento delle giunzioni 5', 3' e del sito di ramificazione avviene attraverso accoppiamento di basi tra il snRNA e le sequenze di RNA nel substrato. All'interno della cellula il complesso esiste come un assemblaggio vago di tutte le componenti che svolgono lo splicing come un'unità coordinata che continua a riordinarsi ogni volta che compie uno splice.

### 7.1.17 Lo spliceosoma usa l'idrolisi dell'ATP per produrre una serie di riordinamenti RNA-RNA

L'idrolisi dell'ATP è necessaria per l'assemblaggio e riordinamento dello spliceosoma per rompere e formare interazioni RNA-RNA. Ogni splice richiede circa 200 proteine. Questi riordinamenti permettono l'esaminazione del pre-RNA dal snRNP. U1 riconosce il sito 5' attraverso l'accoppiamento di basi e successivamente questi legami sono rotti con l'idrolisi dell'ATP e viene sostituito con U6. Questi tipi di riordinamento avvengono molte volte e permettono un controllo da parte dello spliceosoma dei segnali di splicing aumentando la precisione del processo. Questi riordinamenti avvengono anche per creare i siti attivi nello spliceosoma per le due transesterificazioni, cosa che avviene sequenzialmente dopo che i segnali di splicing sono stati analizzati più volte. I siti catalitici sono formati da proteine e molecole di RNA e le seconde catalizzano la reazione chimica. Una volta che la chimica dello splicing è completata gli snRNP rimangono legati al lazo. Il loro disassemblaggio richiede un'altra serie di riordinamenti RNA-RNA che richiedono l'idrolisi dell'ATP che permettono il ritorno di snRNA alla loro conformazione originale che ne permette il riutilizzo. Al completamento lo spliceosoma direziona un insieme di proteine per legarsi all'mRNA vicino la posizione precedentemente occupata dall'introne detto complesso di giunzione degli esoni (EJC) che marcano il sito di uno splicing riuscito e influenzano il destino dell'mRNA.

### 7.1.18 Altre proprietà del pre-mRNA e la sua sintesi spiega le scelte dei siti propri di splice

Il meccanismo di riconoscimento dello spliceosoma sfrutta due strategie aggiuntive per aumentarne l'affidabilità. Il primo è una conseguenza di essere accoppiato con la trascrizione: quando questa procede la coda fosforilata della RNA polimerasi porta varie componenti dello spliceosoma che sono direttamente trasferite dalla polimerasi all'RNA mentre emerge da essa, aiutando a tenere traccia di esoni e introni. La seconda strategia è detta definizione degli esoni: la dimensione degli esoni tende a essere più uniforme di quella degli introni e attraverso questa definizione si possono ricercare sequenze di esoni di dimensione omogenea. Mentre la sintesi dell'RNA procede un gruppo di componenti aggiuntive come proteine SR si assemblano sulla sequenza di esoni e aiutano a marcare i siti di splice. Queste proteine reclutano U1 snRNA che marca il limite dell'esone downstream e U2 che specifica l'upstream. Questi processi aumentano la precisione del deposito delle componenti

iniziali dello splicing. Negli esoni sono presenti sequenze dette splicing enhancers. Le sequenze di introni non sono rimosse dall'RNA nello stesso ordine in cui sono marcate.

### 7.1.19 La struttura cromatinica influenza l'RNA splicing

I nucleosomi tendono a essere posizionati sugli esoni e causano la proteina responsabile per la definizione degli esoni di assemblarsi all'RNA quando emerge dalla polimerasi. Cambi nella struttura cromatinica sono usati per cambiare i pattern di splicing in due modi. Siccome splicing e trascrizione sono accoppiate il tasso con cui la RNA polimerasi si muove lungo il DNA ha effetto sul tasso di splicing: minore la velocità minore il salto di esoni: l'assemblaggio dello spliceosoma iniziale può essere completato prima che una scelta alternativa di sito di splicing venga presentata. I nucleosomi in cromatina condensata possono causare una pausa nella polimerasi. Inoltre specifiche modifiche agli istoni attraggono componenti dello spliceosoma che possono essere facilmente trasferiti all'RNA emergente.

### 7.1.20 Lo splicing di RNA mostra plasticità notevole

In confronto ad altri processi nell'espressione dei geni lo splicing è flessibile. Il macchinario di splicing si è evoluto in modo da cercare il pattern migliore per le giunzioni, ma se uno di essi è stato danneggiato da una mutazione ricerca il prossimo migliore. Questa plasticità del processo suggerisce che cambi nei pattern di splicing sono stati importanti nell'evoluzione di geni e organismi e che mutazioni che riguardano lo splicing possono essere distrutte per l'organismo. La cellula inoltre può facilmente regolare i pattern di RNA splicing in modo da produrre diverse forme di una proteina a tempi e tessuti diversi.

### 7.1.21 Rna splicing catalizzato dallo spliceosoma si è probabilmente evoluto da meccanismi di autosplicing

Le cellule ancestrali utilizzavano RNA per le catalisi principali e salvavano informazioni geniche nell'RNA rispetto al DNA. Queste reazioni di splicing hanno molto probabilmente avuto un ruolo fondamentale. Come prova rimangono introni a RNA autosplicing.

### 7.1.22 Enzimi di processamento dell'RNA generano la terminazione 3' degli mRNA eucariotici

Quando la RNA polimerasi II raggiunge la fine di un gene un meccanismo garantisce che la 3' fine del pre-mRNA sia processata. La posizione della terminazione 3' è specificata da segnali codificati dal genoma, trascritti nell'RNA mentre la polimerasi si muove attraverso essi e successivamente riconosciuti da una serie di proteine che si legano all'RNA e enzimi di processamento dell'RNA. Due proteine a subunità multiple dette CstF (cleavage stimulation factor) e CPSF (cleavage and polyadenylation specificity factor) hanno grande importanza. Entrambe queste proteine viaggiano con la coda dell'RNA polimerasi e sono trasferite alla terminazione 3' quando emerge dalla polimerasi stessa. Una volta che queste proteine si legano con alle sequenze di riconoscimento altre proteine si assemblano con esse per creare la terminazione 3' dell'mRNA. L'RNA è rotto dalla polimerasi, successivamente un enzima detto poly-A polimerasi (PAP) aggiunge sequenzialmente 200 nucleotidi A alla terminazione appena prodotta. Il precursore delle addizioni è ATP. Mentre questa coda poly-A viene sintetizzata (senza stampo) le proteine che si legano ad essa sono assemblate su di essa. Dopo che la terminazione è stata separata la polimerasi continua a trascrivere e l'RNA che sintetizza

non possiede un cappuccio 5' e viene degradato da una esonucleasi trasportata lungo la coda della polimerasi che causa l'eventuale separazione della polimerasi dallo stampo e la terminazione della trascrizione.

### 7.1.23 mRNA maturi eucariotici sono esportati selettivamente dal nucleo

La sintesi e il processamento del pre-mRNA nel nucleo si svolge in maniera ordinata, ma solo una piccola percentuale di questo viene ulteriormente utilizzato dalla cellula, mentre il resto è non solo inutile ma potenzialmente dannoso. Per distinguere da questo pre-mRNA e l'mRNA maturo mentre una molecola di RNA viene processata perde delle proteine e ne acquisisce altre: la presenza di una proteina snRNP significa splicing incompleto o errato. Solo quando le proteine presenti sulla molecola di mRNA segnalano collettivamente che il suo processamento è stato completato con successo l'mRNA è esportato dal nucleo nel citosol, dove viene tradotto in proteine. I resti rimangono nel nucleo dove sono degradati dall'esosoma nucleare, un complesso proteico ricco di RNA esonucleasi. Le molecole di pre-mRNA più comuni della trascrizione sono hnRNP (heterogeneous nuclear ribonuclear proteins) e alcune di esse svolgono le eliche a forcina nell'RNA in modo da permettere una facile lettura dei segnali e uno splicing più semplice. I mRNA maturi sono guidati attraverso il complesso di pori nucleici o NPC, canali acquosi nella membrana nucleare che connettono il nucleoplasma con il citosol. Il passaggio di macromolecole richiede energia per un trasporto attivo in entrambe le direzioni attraverso il complesso. Le macromolecole sono mosse attraverso recettori di trasporto nucleari che le trasportano selettivamente. Affinchè avvenga l'esportazione dell'mRNA un recettore specifico deve essere caricato su di esso, un passo che avviene insieme alla rottura 3' e alla poliadenilazione. Una volta che l'esportazione avviene il recettore si dissocia, rientra nel nucleo e viene riutilizzato. Alcune delle proteine possono influenzare il comportamento dell'RNA successivamente all'esportazione, come la stabilità, l'efficienza della traduzione e la destituzione ultima.

### 7.1.24 Anche gli RNA non codificanti sono sintetizzati e processati nel nucleo

La maggior parte dell'RNA nella cellula ha funzioni strutturali e catalitiche. L'RNA più presente è quello ribosomiale rRNA, per circa l'80%. Questo RNA forma il nucleo del ribosoma. Gli eucarioti possiedono la RNA polimerasi I per la sintesi degli rRNA. È strutturalmente simile alla RNA polimerasi II senza una coda C-terminale e pertanto i suoi trascritti non sono nè incappucciati nè poliadenilati. Essendo le componenti a RNA del ribosoma prodotti finali dei geni in ogni cella sono presenti multiple copie dei geni rRNA. Le cellule umane ne contengono 200 per genoma aploide in piccoli cluster su 5 cromosomi diversi. Ci sono quattro tipi di rRNA eucariote, ognuno presente in una copia per ribosoma. Tre dei quattro (18S, 5.8S e 28S) sono creati modificando e rompendo un rRNA precursore, mentre il quarto (5S) sintetizzato da un cluster separato dalla polimerasi III. Le modificazioni che avvengono al precursore a 13'000 nucleotidi sono 100 metilazioni delle posizioni 2' - OH sui zuccheri nucleotidici e 100 isomerizzazioni dei nucleotidi uridina a pseudouridina. Queste modifiche aiutano il piegamento e assemblaggio dei rRNA finali o alterano leggermente la funzione dei ribosomi. Ogni alterazione è fatta a una posizione specifica determinata da RNA guida che si posizionano sul precursore tramite accoppiamento di basi e portano l'enzima modificatore alla posizione corretta. Tutti questi RNA sono detti piccoli RNA nucleari o snoRNA e svolgono le loro funzioni in sottocompartimenti del nucleo detti nucleoli. Molti sono codificati dagli introni di altri geni e sono sintetizzati da RNA polimerasi II e processati da sequenze di introni esportate.

### 7.1.25 Il nucleolo è una fabbrica di produzione di ribosomi

Il nucleolo è il sito per il processamento di rRNA e il loro assemblaggio in subunità del ribosoma. Non è confinato da una membrana ma consiste di un grande aggregato di macromolecole includenti i geni rRNA, rRNA precursori, maturi, enzimi per il loro processamento, snoRNP, fabbriche si assemblaggio come ATPasi, GTPasi, proteina chinasi e RNA elicasi, proteine ribosomiali e ribosomi parzialmente assemblati. Molti tipi di molecole di RNA svolgono un ruolo centrale nella chimica e struttura del nucleolo. I geni di rRNA, distribuiti in 10 cluster negli umani, durante l'interfase creano degli anelli che creano una parte del nucleolo, durante la fase M quando i cromosomi si condensano il nucleolo si frammenta e scompare. Nella parte di telofase della mitosi, quando i cromosomi ritornano al loro stato semi-disperso riappare. La sua dimensione dipende dal numero di ribosomi che la cellula sta producendo. L'assemblaggio del ribosoma è un processo complesso. Oltre al ruolo centrale nella biogenesi del ribosoma il nucleolo è il sito dove altri RNA non codificanti sono prodotti e complessi RNA-proteine sono assemblati. Il nucleo può pertanto essere considerato come una fabbrica in cui RNA non codificanti sono trascritti, processati e assemblati con le proteine formando una grande varietà di complessi ribonucleoproteici.

### 7.1.26 Il nucleo contiene una varietà di aggregati subnucleari

Nel nucleo sono presenti altri corpi come i corpi di Cajal senza membrana e altamente dinamici in base alle necessità della cellula. Il loro assemblaggio è mediato dall'associazione di domini proteici semplici e la loro apparenza è il risultato di associazioni strette di elementi proteici e a RNA coinvolti nella sintesi, assemblaggio e conservazione di macromolecole coinvolte nell'espressione genica. I bodi di Cajal sono siti dove i snRNP e i snoRNP svolgono i loro ultimi passi di maturazione e dove i snRNP sono riciclati e i loro RNA si resettando dopo i riordinamenti dello splicing. I cluster di granuli intercromatinici sono proposti come pile di snRNP e altre componenti di processamento di RNA che sono usate nella produzione di mRNA. Sembra che la funzione principale di questi aggregati sia concentrare i componenti in modo da velocizzare il loro assemblaggio. I siti dello splicing sono nell'ordine delle migliaia, altamente dinamici e il risultato dell'associazione di componenti di trascrizione e splicing per la creazione di piccole fabbriche, il nome dato ad aggregati specifici contenenti un'alta concentrazione di componenti selezionate che creano catene di montaggio biochimiche.

## 7.2 Da RNA a proteine

La maggior parte dei geni nella cellula producono molecole di mRNA che servono come intermediari sul cammino verso le proteine.

### 7.2.1 Una sequenza di mRNA è decodificata in insiemi di tre nucleotidi

L'informazione di un mRNA maturo è utilizzata per sintetizzare una proteina. Questa sintesi è detta traduzione. Essendoci 4 nucleotidi diversi e 20 amminoacidi la traduzione non avviene uno a uno. Le regole di traduzione sono dette codice genetico. La sequenza di nucleotidi viene letta in gruppi consecutivi di 3 nucleotidi con 64 possibili combinazioni. Il codice è ridondante e alcuni amminoacidi sono codificati da più triplette dette codoni che codificano un amminoacido o la terminazione del processo di traduzione. Questo codice è utilizzato universalmente, con piccole differenze nei mitocondri, che hanno sistemi indipendenti. Una sequenza di RNA può essere tradotta in uno di tre diversi reading frame, dipendenti dal luogo di inizio della decodifica. SOlo uno dei tre codifica la proteina richiesta.

### 7.2.2 Molecole di tRNA combinano amminoacidi ai codoni

I codoni non si legano direttamente all'amminoacido ma si richiede una molecola di adattamento che può riconoscere e legare sia il codone che l'amminoacido. Questi consistono di un insieme di piccole molecole di tRNA lunghi 80 nucleotidi piegandosi in una conformazione precisa. Quattro segmenti di tRNA sono a doppia elica producendo una molecola simile a un quadrifoglio che viene successivamente piagata in una forma a L compatta tenuta insieme da legami a idrogeno tra le diverse regioni della molecola. Due regioni di nucleotidi non accoppiati alle terminazioni della molecola sono fondamentali: una forma l'anticodone, tre nucleotidi che si accoppiano con il codone complementare e l'altra una regione a filamento singolo alla terminazione 3' dove l'amminoacido si attacca al tRNA. La ridondanza indica che esistono più di un tRNA per amminoacido e che alcuni tRNA possono legarsi a più codoni. Alcuni tRNA richiedono accoppiamento di basi accurato solo per due posizioni e possono tollerare una corrispondenza sbagliata (wobble) che spiega perchè molti codoni alternativi per un amminoacido differiscono solo nel terzo nucleotide.

### 7.2.3 tRNA sono modificati covalentemente prima che escano dal nucleo

La sintesi del tRNA avviene dalla RNA polimerasi III. Sono tipicamente sintetizzati come precursori più grandi, che sono rifilati per produrre tRNA maturo. Contengono anche introni. Lo splicing usa un meccanismo di copia-incolla catalizzato da proteine. Entrambi richiedono che il precursore sia correttamente piegato nella configurazione a quadrifoglio. Tutti i tRNA sono modificati chimicamente (1 in 10 nucleotidi è una versione alterata del ribonucleotide standard) che facilitano il riconoscimento del codone appropriato.

### 7.2.4 Specifici enzimi accoppiano amminoacidi con la molecola di tRNA appropriata

Il riconoscimento e l'attaccamento dell'amminoacido corretto dipende dall'enzima amminoacil-tRNA sintetasi che lega covalentemente ogni amminoacido con l'insieme corretto di molecole di tRNA. La maggior parte delle cellule possiedono una sintetasi diversa per ogni amminoacido. Queste reazioni attaccano alla terminazione 3' del tRNA l'amminoacido accoppiate con l'idrolisi dell'ATP producendo un legame ad alta energia tra il tRNA e l'amminoacido, energia utilizzata per legare l'amminoacido covalentemente con la catena polipeptidica crescente. Il codice genetico è tradotto pertanto da due insiemi di adattatori che agiscono sequenzialmente, ognuno dei quali corrisponde una superficie molecolare ad un'altra con grande specificità.

### 7.2.5 La modifica da tRNA sintetasi assicura accuratezza

La maggior parte delle sintetasi selezionano l'amminoacido corretto con un meccanismo a due fasi. Tale amminoacido ha la maggiore affinità per il sito attivo della sintetasi ed è favorito, ma la discriminazione tra amminoacidi simili avviene in un secondo passaggio dopo che l'amminoacido è stato legato covalentemente a AMP. Quando il tRNA si lega la sintetasi prova a forzare l'amminoacido adenilato in una seconda tasca di modifica nell'enzima la cui dimensione permette l'accesso unicamente ad amminoacidi strettamente imparentati con quello corretto. In questa tasca l'amminoacido è rimosso dall'AMP attraverso idrolisi, aumentando l'accuratezza a un errore ogni 40.000 accoppiamenti. La sintetasi deve anche essere in grado di riconoscere il corretto insieme di tRNA e la loro complementarità chimica permette di investigare diverse caratteristiche del tRNA. La maggior parte viene riconosciuta direttamente: ci sono tre tasche adiacenti di legame con i nucleotidi, ognuna delle quali complementare in forma e carica ad un nucleotide nell'anticodone.

### 7.2.6 Gli amminoacidi sono aggiunti alla terminazione C<sup>-</sup> di una catena polipeptidica crescente

La reazione fondamentale nella sintesi di una proteina è la formazione di un legame peptide tra il gruppo carbossile alla fine di una catena polipeptidica e un gruppo ammino su un amminoacido in arrivo. Una proteina è sintetizzata per passaggi dalla terminazione N<sup>-</sup> a quella C<sup>-</sup> durante l'intero processo la terminazione carbossile crescente rimane attivata da l'attacco covalente a una molecola di tRNA. Ogni addizione rompe il legame covalente ad alta energia, sostituendolo con uno uguale con l'amminoacido più recente. In questo modo ogni amminoacido trasporta con sé l'energia di attivazione necessaria per l'addizione del prossimo amminoacido.

### 7.2.7 Il messaggio di RNA è decodificato nei ribosomi

La sintesi delle proteine è guidata da molecole di mRNA. Per mantenere un reading frame corretto e garantire accuratezza la sintesi avviene nel ribosoma, un complesso di proteine ribosomiali e molecole di RNA ribosomiali rRNA. Sono presenti in milioni nel citoplasma. Le loro subunità grande e piccola sono assemblate nel nucleo, dove rRNA si associano con le proteine ribosomiali trasportate lì. Queste due subunità sono esportate nel citoplasma dove si uniscono per sintetizzare le proteine. I ribosomi eucariotici e batterici hanno strutture e funzioni simili, composti da due subunità, una grande e una piccola che si uniscono formando un ribosoma con una massa di milioni di danlton. La subunità piccola crea il framework in cui i tRNA sono corrisposti ai codoni dell'mRNA, mentre la grande catalizza la formazione dei legami peptidi tra gli amminoacidi. Quando non sono attive le subunità sono separate. Si uniscono su una molecola di mRNA vicino alla terminazione 5' per iniziare la sintesi di una proteina. L'mRNA viene poi tirato attraverso il ribosoma tre nucleotidi alla volta. Mentre il codone entra la sequenza è tradotta in amminoacidi attraverso il tRNA. Quando si incontra un codone di stop il ribosoma rilascia la proteina finita e le due subunità possono separarsi per un futuro riutilizzo. In un secondo si possono aggiungere 2 amminoacidi per quelli eucariotici, 20 per i batterici. Il ribosoma contiene quattro siti di legame per le molecole di RNA: uno per l'mRNA e tre (siti A, P e E) per i tRNA, collegate strettamente nei siti A e P solo se il suo anticodone forma accoppiamento di basi con il codone complementare nel ribosoma. I siti A e P sono abbastanza vicini da permettere la formazione tra le due molecole di tRNA di legami tra le basi con i codoni adiacenti con la molecola di mRNA. Una volta che inizia la sintesi delle proteine ogni nuovo amminoacido è aggiunto alla catena crescente in un ciclo di reazioni a quattro passaggi: il legame di tRNA, la formazione del legame peptide, la traslocazione della grande subunità e la traslocazione della piccola subunità. Come risultato dei due passi di traslocazione l'intero ribosoma si muove di tre nucleotidi lungo l'mRNA.

### 7.2.8 Fattori di allungamento portano avanti la traduzione e ne aumentano l'accuratezza

Due fattori di allungamento entrano e lasciano il ribosoma durante ogni ciclo idrolizzando GTP in GDP con conseguente cambi conformazionali. Questi fattori sono chiamati EF-Tu e EF-G nei batteri e EF1 e EF2 negli eucarioti. L'accoppiamento con questi fattori e la loro transizione di conformazione durante il ciclo velocizza e rende la sintesi più accurata. I cicli di associazione, idrolisi e disassociazione garantiscono che i cambi avvengano nella direzione corretta. EF-Tu aumenta l'accuratezza in quanto può legare contemporaneamente GTP e gli amminoacil-tRNA. In questa forma l'interazione codone-anticodone avviene nel sito A. A causa dei cambi di energia liberati associati la corrispondenza corretta si lega più strettamente, ma con differenze troppo lievi per garantire

accuratezza. Per aumentare l'accuratezza della reazione il ribosoma e il EF-Tu lavorano insieme: i 16 rRNA nella piccola subunità determinano la correttezza della corrispondenza codone-anticodone piegandosi intorno ad esso e controllando i dettagli molecolari. Quando si trova una corrispondenza corretta il rRNA si chiude strettamente intorno alla coppia causando un cambio conformazionale al ribosoma che causa l'idrolisi del GTP dall'EF-Tu. Solo quando il GTP viene idrolizzato l'EF-Tu rilascia la stretta sul amminoacil-tRNA e gli permette di essere utilizzato nella sintesi. Se la corrispondenza non avviene i tRNA escono dal ribosoma prima che possano essere utilizzati. Dopo che il GTP viene idrolizzato e l'EF-Tu si disassocia c'è un ritardo mentre l'amminoacido si muove in posizione, impedendo ai tRNA incorretti di proseguire la sintesi in quanto si disassociano troppo rapidamente. Un'interazione codone-anticodone al sito P che avviene dopo l'incorporazione di un amminoacido errato causa un aumento del tasso di letture errate al sito A. Eventi scorretti successivi causano una terminazione prematura causata dai fattori di rilascio, che rilasciano la proteina errata per la degradazione.

### 7.2.9 Molti processi biologici superano le limitazioni intrinseche all'accoppiamento di basi complementari

Altri meccanismi vengono utilizzati per aumentare la specificità della sintesi. Il primo è l'adattamento indotto: l'interazione codone-anticodone è controllata dall'accoppiamento di basi e dal piegamento del ribosoma, che dipende dalla correttezza della corrispondenza. Un secondo principio è il proofreading cinetico. L'idrolizzazione del GTP dopo l'accoppiamento iniziale crea un passaggio irreversibile e inizia un tempo su un delay durante il quale l'amminoacil-tRNA si muove nella posizione corretta per la catalisi. Durante questo ritardo le coppie incorrette hanno una probabilità maggiore di dissociarsi in quanto la relazione con il tRNA è più debole e il ritardo maggiore.

### 7.2.10 L'accuratezza nella traduzione richiede una spesa di energia libera

La traduzione è un compromesso tra velocità e accuratezza, inoltre la sintesi delle proteine richiede più energia libera di tutti gli altri processi biosintetici. Quattro legami fosfati ad alta energia sono rotti per ogni nuovo legame peptidico: due per caricare il tRNA con l'amminoacido e due per il ciclo di reazioni durante la sintesi nel ribosoma. Altra energia viene consumata ogni volta che un tRNA incorretto entra nel ribosoma, comincia l'idrolisi del GTP ed è rifiutato.

### 7.2.11 Il ribosoma è un ribozima

Il ribosoma è un grande complesso formato per due terzi da RNA e per un terzo da proteine. Gli rRNA sono responsabili per la sua struttura, la sua abilità di posizionare i tRNA sui mRNA e l'attività catalitica. Gli RNA ribosomiali sono piegati in strutture tridimensionali precise e compatte che formano il nucleo del ribosoma e ne determinano la forma. Le proteine si trovano generalmente sulla superficie e riempiono i vuoti e fessure dell'RNA piegato. Alcune di queste mandano fuori regioni di catena polipeptidica che possono penetrare in buchi del nucleo, il loro ruolo principale sembra quello di stabilizzare il nucleo permettendo i cambi conformazionali necessari. Aiutano inoltre nell'assemblaggio degli rRNA che costituiscono il nucleo. Anche il sito catalitico per la formazione del legame peptidico è costituito da RNA nonostante non contenga gruppi funzionali facilmente ionizzabili. Si crede che la struttura dei 23S rRNA formi una tasca altamente strutturata che orienta precisamente i due reagenti attraverso una rete di legami a idrogeno. Il tRNA nel sito P contribuisce con un gruppo OH al sito attivo e partecipa direttamente nella catalisi, assicurando

che la reazione avvenga solo quando il tRNA è posizionato correttamente. Le molecole di RNA con attività catalitica sono dette ribozimi.

### 7.2.12 La sequenza nucleotidica negli mRNA segnala dove iniziare la sintesi delle proteine

L'iniziazione e la terminazione della traduzione condividono caratteristiche con il ciclo di allungamento. Il sito di inizio della sintesi della proteina sull'mRNA è cruciale in quanto stabilisce il reading frame per l'intera lunghezza del messaggio. Un errore di uno dei nucleotidi a questo stato causerebbe ogni codone seguente di essere letto male e creando una proteina non funzionale ed è l'ultimo punto in cui la cellula può decidere se l'mRNA verrà tradotto. Il tasso di questo passaggio è determinante del tasso in cui una proteina viene sintetizzata. La traduzione inizia con il codone AUG e un tRNA iniziatore che trasporta l'amminoacido metionina, in ogni terminazione N<sup>+</sup> della proteina che viene rimosso da una proteasi specifica. Questo tRNA è riconosciuto da fattori di iniziazione in quanto ha una sequenza nucleotidica distinta da quello che usualmente trasporta la metionina. Il complesso iniziatore tRNA-metionina è prima caricato nella piccola subunità ribosomiale con i fattori eucariotici di iniziazione o eIF. Di tutti gli amminoacil-tRNA nella cellula solo l'iniziatore è in grado di legare la piccola subunità ribosomiale senza che sia presente il ribosoma completo e si lega direttamente al sito P. Successivamente la piccola subunità ribosomiale si lega alla terminazione 5' di un mRNA, riconosciuta grazie al cappuccio e poi si muove in avanti alla ricerca del primo AUG. Altri fattori di iniziazione agiscono come elicasi energizzate dall'ATP facilitandone il movimento. Quando trova AUG i fattori di iniziazione si disassociano permettendo alla subunità di associarsi con il complesso e completo ribosoma. Il tRNA iniziatore rimane al sito P, lasciando il sito A vuoto e la sintesi è pronta a cominciare. I nucleotidi che circondano il sito di inizio influenzano l'efficienza del riconoscimento dell'AUG durante lo scan. Se il sito di riconoscimento è diverso sostanzialmente dalle regioni adiacenti lo scan può volte ignorare l'AUG e passare ad un altro. Questo fenomeno detto leaky scanning viene utilizzato per produrre la stessa proteina con e senza una sequenza di segnale attaccata. Il meccanismo per selezionare il codone di inizio nei batteri è diverso: ogni mRNA contiene uno specifico sito di legame al ribosoma locato pochi nucleotidi prima dell'AUG dove inizia la traduzione. I ribosomi batterici possono assemblare direttamente su un codone di inizio interno all'mRNA che pertanto può essere utilizzato per codificare più proteine e si dice policistronico.

### 7.2.13 I codoni di fine marcano la fine della traduzione

La fine del messaggio di codifica è segnalata dalla presenza di uno dei tre codoni di fine UUA, UAG o UGA che non sono riconosciuti da un tRNA e non specificano un amminoacido, ma segnalano al ribosoma di terminare la traduzione. I fattori di rilascio legano ogni ribosoma con un codone di fine posizionato al sito A, forzando la transferasi peptidica a catalizzare l'addizione di una molecola di acqua invece di un amminoacido al peptidil-tRNA. Questa reazione libera la fine carbossilica della catena polipeptidica dalla molecola di tRNA e il suo rilascio nel citoplasma. Il ribosoma poi rilascia la molecola di mRNA legata e si separa nelle due subunità che si possono poi assemblare su un'altra molecola di mRNA. Durante la traduzione il polipeptide nascente si muove attraverso un tunnel pieno d'acqua nella subunità grande del ribosoma. Le pareti del tunnel composte dai 23S rRNA sono un insieme di piccole superfici idrofobiche incastrate in una superficie idrofila più estensiva. Questa struttura non è complementare a nessun peptide e mette a disposizione un incapsulamento in cui il polipeptide può sciogliersi. Le proteine sono senza struttura mentre passano attraverso il ribosoma, anche se si possono creare delle regioni a  $\alpha$  elica. Mentre lascia il ribosoma la proteina deve piegarsi nella conformazione utile alla cellula.



### 7.2.14 Le proteine sono create su poliribosomi

La sintesi della maggior parte delle proteine avviene tra i 20 secondi e molti minuti. Durante questo periodo avvengono molte iniziazioni su ognuno dei mRNA tradotti. Appena il ribosoma precedente ha tradotto abbastanza sequenza nucleotidica per muoversi la terminazione 5' viene messa in un nuovo ribosoma. Le molecole di mRNA che sono tradotte si trovano nella forma di poliribosomi: assemblaggi citoplasmatici grandi composti da molti ribosomi vicini al massimo 80 nucleotidi lungo un singolo mRNA.

### 7.2.15 Ci sono piccole variazioni nel codice genetico standard

Il codice genetico possiede delle rare eccezioni: nei mitocondri dei mammiferi AUA è tradotta in metionina, mentre nel citosol della cellula come isoleucina. Questo tipo di deviazione è specifica nell'organismo o nell'organello in cui avviene. In molte cellule avviene una recodifica di traduzione in cui altra informazione di sequenza nucleotidica presente nell'mRNA può cambiare il significato del codice genetico ad un sito particolare dell'mRNA in quanto esiste un ventunesimo amminoacido, la selenocisteina, essenziale per molte funzioni enzimatiche che contiene un atomo di selenio al posto dello zolfo della cisteina. È prodotta enzimaticamente da una serina attaccata a una molecola speciale di tRNA che si accoppia con il codone UGA. L'mRNA per le proteine che la contengono possiedono una sequenza nucleotidica vicina che causa l'evento di ricodifica.

### 7.2.16 Meccanismi di controllo della qualità agiscono per prevenire la traduzione di mRNA danneggiati

Gli mRNA possono lasciare il nucleo danneggiati o danneggiarsi nel processo. Sono presenti molti meccanismi per impedire la loro traduzione come il riconoscimento del cappuccio 5' e la catena poli-A prima della traduzione. Il meccanismo più potente è il decadimento di mRNA mediato da nonsense che elimina gli mRNA difettivi prima che si sposino dal nucleo. Questo viene chiamato quando la cellula determina che una molecola di mRNA possiede un codone di terminazione nonsense nel posto errato. Avviene principalmente in una molecola con uno splice errato. Questo meccanismo inizia quando una molecola di mRNA viene trasportata nel citosol. Mentre la terminazione 5' emerge dal poro l'mRNA si incontra con un ribosoma che comincia la traduzione. Mentre questa procede il complesso di giunzione degli esoni legati all'mRNA a ogni sito di splice sono mossi dal ribosoma. Il codone di fine normale si trova dentro l'ultimo esone e quando il ribosoma arriva ad esso e si ferma nessun EJC sarà legato all'mRNA. In questo caso l'mRNA passa l'ispezione e può essere tradotto. Altrimenti la molecola viene degradata. Il primo passo di traduzione permette alla cellula di verificare la capacità della molecola di mRNA di produrre la proteina corretta mentre esce il nucleo.

### 7.2.17 Alcune proteine cominciano a piegarsi mentre vengono sintetizzate

La catena polipeptidica per essere utile alla cellula deve piegarsi nella conformazione, legare ogni cofattore necessario alla sua attività, essere modificata dalla chinasi o altri enzimi e assemblarsi con altre subunità proteiche con cui funziona. Queste informazioni sono contenute nella sequenza di amminoacidi. Quando una proteina si piega la parte idrofobica si trova in un nucleo interno e un gran numero di interazioni non covalenti si formano tra varie parti della molecola. L'insieme di queste interazioni energeticamente favorevoli determina la conformazione finale della catena. Per alcune proteine il ripiegamento inizia appena la catena esce dal ribosoma cominciando dalla terminazione

$N^-$ . In questi casi si forma in pochi secondi una struttura compatta che contiene la maggior parte delle strutture secondarie. Per alcuni domini proteici si crea uno stato flessibile detto il molten globule, il processo di inizio verso l'arrivo alla conformazione corretta.

### 7.2.18 Accompagnatori molecolari aiutano a guidare il piegamento della maggior parte delle proteine

La maggior parte delle proteine non si ripiegano correttamente durante la loro sintesi e richiedono proteine dette accompagnatori molecolari, utili in quanto stabiliscono il cammino di piegamento che la proteina deve compiere. Riconoscono specificatamente configurazioni inforrette dall'esposizione delle superfici idrofobiche causato da legami reciproci. Gli accompagnatori evitano questo legandosi con tali superfici.

### 7.2.19 La cellula utilizza diversi tipi di accompagnatori

Molti accompagnatori sono detti proteine a shock termico (hsp) in quanto sono sintetizzati in quantità maggiore dopo brevi esposizioni della cellula a temperature elevate che riflette all'operazione di un sistema di feedback che risponde a un aumento in proteine malformate aumentando la sintesi degli accompagnatori che aiutano a ripiegarsi. Ci sono diverse famiglie di accompagnatori e diversi membri funzionano in diversi organelli. Le proteine hsp60 e hsp70 lavorano con il loro piccolo insieme di proteine associate, hanno un'affinità per superfici idrofobiche sposte e idrolizzano l'ATP legandosi e rilasciando il substrato proteico ad ogni ciclo di idrolisi. hsp70 lavora con cellule appena sintetizzate con ogni suo monomero legandosi a una stringa di quattro o cinque amminoacidi idrofobici. Sull'ATP legante rilascia la proteina da una struttura a basile che agisce dopo che la proteina si è completamente sintetizzata. Detto chaperonina forma una camera di isolamento per il processo di piegamento. Per entrare la camera la proteina substrato deve essere catturata attraverso l'entrata idrofobica e viene rilasciata nella camera piena di superfici idrofiliche e la camera è chiusa attraverso idrolisi dell'ATP. Qui il substrato si piega nella conformazione finale in isolamento. Quando l'ATP viene idrolizzato il coperchio della camera si separa e la proteina substrato esce dalla camera. L'energia dell'idrolisi viene utilizzata per movimenti meccanici che convertono gli accompagnatori dalla forma di cattura a quella di rilascio.

### 7.2.20 Regioni idrofobiche esposte forniscono segnali critici per il controlli di qualità delle proteine

L'azione della proteina hsp70 inizia quando una proteina sta venendo ancora sintetizzata. La cellula riconosce le proteine piegate male che richiedono turni addizionali di ripiegamento catalizzati dall'ATP. Se una proteina ha una superficie di amminoacidi idrofobici è anormale e può essere pericolosa per la cellula. Le proteine che si piegano velocemente da sole non mostrano questi pattern e possono evitare gli accompagnatori, mentre le altre sono riparate da essi. Quando questo non funziona un meccanismo la distrugge completamente. Questo cammino comincia con il riconoscimento di una superficie idrofobica anormale e finisce con la consegna della proteina a un complesso di proteasi detto proteosoma che la distrugge.

### 7.2.21 Il proteosoma è una proteasi compartimentalizzata con siti attivi reclusi

Il macchinario proteolitico e gli accompagnatori competono per il riconoscimento di una proteina piegata male. Se si ripiega rapidamente solo una piccola frazione viene degradata. L'apparato che distrugge le proteine errate è il proteosoma, una proteasi dipendente dall'ATP. È presente in molte copie disperse nel citosol e nel nucleo e distrugge anche le proteine che sono entrate nel reticolo endoplasmatico. In questo caso sono riconosciute da un sistema di sorveglianza e le retrotrasloca nel citosol per la degradazione da parte del proteosoma. Ogni proteosoma consiste di un cilindro centrale vuoto formato da multiple subunità proteiche che si assemblano come uno stack di quattro anelli eptamerici. Alcune delle subunità sono proteasi i cui siti attivi si trovano all'interno della camera impedendo la loro azione incontrollata. Ogni fine del cilindro è associata a un complesso proteico che contiene sei anelli attraverso i quali le proteine obiettivo sono portate verso il nucleo dove sono degradate. La reazione di importazione è guidata dall'idrolisi dell'ATP e svolge la proteina obiettivo mentre si muove lungo il cappuccio esondandola alla proteasi nel nucleo. Le proteine appartengono alle unfoldases o proteina AAA che funzionano come esameri e hanno caratteristiche comuni alla DNA elicasi. Una proprietà cruciale del proteosoma è la processività del meccanismo: il substrato rimane legato fino a che non è completamente convertito in peptidi corti. I 19S cappucci del proteosoma agiscono come cancelli all'entrata del nucleo proteolitico interno e solo le proteine marcate per la distruzione possono passarci altradetto. La marcatura è il legame con l'ubiquitina, una piccola proteina che in questo caso è legata in una catena alla lisina 48. Un insieme speciale di molecole E3 è responsabile per l'ubiquitilazione di proteine denaturate, mal piegate o contenenti amminoacidi anormali o ossidati attraverso una superficie idrofobica esposta che agisce come segnale per queste molecole. In ogni caso si deve distinguere tra proteine completate con la conformazione errata e proteine che si stanno formando.

### 7.2.22 Molte proteine sono controllate da distruzione regolata

Un'altra funzione di questi cammini proteolitici è di conferire corte vite a proteine specifiche la cui concentrazione deve cambiare rapidamente con stati alterati della cellula. Alcune di queste sono degradate velocemente sempre, altre solo in certe condizioni. Questo viene controllato in una classe di meccanismi dall'attività di un ubiquitina ligasi che viene accesa da fosforilazione dell'E3 o da una transizione allosterica dell'E3 causata dal suo legame con una molecola. Il complesso di promozione dell'anafase (APC) è una ligasi a più subunità che è attivata da addizione di subunità durante la mitosi temporizzato dalla cellula. In risposta ad altri segnali si può creare un segnale di degradazione nella proteina causando una rapida ubiquitilazione. Un modo comune è la fosforilazione di un sito specifico che mostra un segnale di degradazione o una disassociazione di una subunità. Segnali di degradazione potenti possono essere creati rompendo un legame peptidico se questo crea una nuova terminazione N<sup>-</sup> riconosciuta da una specifica proteina E3 come un residuo destabilizzante.

## Capitolo 8

# Controllo dell'espressione genica

### 8.1 Una panoramica del controllo dei geni

I differenti tipi di cellula in un organismo cellulare differiscono per struttura e funzione a causa delle diverse proteine che sintetizzano. Non sono differenze nella loro sequenza genica a determinarne la differenziazione ma i cambi nell'espressione dei geni.

#### 8.1.1 Diversi tipi di cellule sintetizzano diversi insiemi di RNA e proteine

Molti processi sono comuni a tutte le cellule e due di esse in un singolo organismo condividono molti prodotti dei geni come proteine strutturali e cromosomi, RNA e DNA polimerasi, enzimi di riparazione del DNA, proteine ribosomiali e RNA, gli enzimi che catalizzano le reazioni centrali del metabolismo e molte proteine che formano il citoscheletro. Alcuni RNA e proteine sono abbondanti in cellule specializzate e non sono individuati in altri luoghi. Studi dei numeri degli RNA diversi suggeriscono che una tipica cellula umana ad ogni momento esprime tra il 30 e il 60% dei suoi geni. Il livello di espressione di quasi tutti i geni varia tra un tipo di cellula e l'altro. Le differenze tra gli mRNA sottostimano le differenze finali nelle proteine in quanto ci sono diversi passaggi dopo la produzione dell'RNA con cui l'espressione viene regolata.

#### 8.1.2 Segnali esterni possono causare il cambio dell'espressione dei geni di una cellula

Ogni cellula è capace di alterare l'espressione dei geni in risposta a indizi extracellulari. Se una cellula del fegato è esposta all'ormone glucocorticoide si aumenta la produzione di energia dagli amminoacidi e altre piccole molecole inducendo l'enzima tirosina amminotrasferasi. Altre cellule rispondono diversamente o non rispondono affatto. Altre caratteristiche dell'espressione genica non cambiano e danno alla cellula le sue caratteristiche distintive.

#### 8.1.3 L'espressione dei geni può essere regolata a molti dei passaggi nel cammino da DNA a RNA a proteine

La cellula può controllare la proteina che produce controllando quanto e quanto spesso viene trascritto il gene (controllo trascrizionale), controllando lo splicing e il processamento dei trascritti di RNA (controllo del processamento dell'RNA), selezionando quale mRNA completo viene esportato

al citosol e determinando dove viene localizzato (controllo di trasporto e localizzazione dell'RNA), selezionando quale mRNA nel citoplasma viene tradotto dal ribosoma (controllo traduzionale), destabilizzando certe molecole di mRNA nel citoplasma (controllo della degradazione dell'mRNA) e attivando, disattivando, degradando o localizzando selettivamente proteine dopo che sono state create (controllo dell'attività proteica).

## 8.2 Controllo della trascrizione da parte di proteine leganti a specifiche sequenze

Un gruppo di proteine detti regolatori di trascrizione riconoscono specifiche sequenze di DNA dette sequenze cis-regolatorie in quanto devono essere sullo stesso cromosoma del gene che regolano. I regolatori di trascrizione si legano a queste sequenze che sono disperse attraverso il genoma e il legame è messo in movimento da una serie di reazioni che specificano i geni da trascrivere e il tasso di trascrizione. La trascrizione di ogni gene è controllata da una collezione di sequenze cis-regolatorie che si trovano tipicamente vicino al gen, tipicamente a monte del punto di inizio di trascrizione. La maggior parte ha un complesso ordinamento di tali sequenze, ognuna riconosciuta da una proteina diversa che determinano il tempo e il luogo di trascrizione di ogni gene.

### 8.2.1 La sequenza dei nucleotidi nella doppia elica di DNA può essere letta da proteine

I regolatori di trascrizione riconoscono sequenze cis-regolatorie corte e specifiche nella doppia elica attraverso informazioni nella sua parte esterna: il lato di ogni coppia presenta un pattern di donatori e accettori di legami a idrogeno e parti idrofobiche sia nella scanalatura maggiore che minore. Tutti i regolatori di trascrizione fanno contatto con la scanalatura maggiore.

### 8.2.2 I regolatori di trascrizione contengono motivi strutturali che possono leggere sequenze di DNA

Il riconoscimento molecolare dipende da un adattamento esatto tra le superfici di due molecole: un regolatore di trascrizione riconosce una sequenza cis-regolatoria specifica in quanto la superficie della proteina è estensivamente complementare alle caratteristiche della doppia elica che presenta tale sequenza. Ogni regolatore di trascrizione fa una serie di contatti con il DNA attraverso legami ionici, a idrogeno e interazioni idrofobiche che in grande numero legano le due parti in una relazione stretta e specifica. Molti regolatori di trascrizione possiedono motivi strutturali che usano  $\alpha$ -eliche o  $\beta$ -foglietti per legarsi alla scanalatura maggiore del DNA. Le catene laterali amminoacide che si estendono fanno contatti specifici con il DNA.

### 8.2.3 La dimerizzazione dei regolatori di trascrizione aumenta la loro affinità e specificità per il DNA

Un monomero di un regolatore di trascrizione tipico riconosce circa 6-8 coppie di nucleotidi, ma riconoscono un intervallo di sequenze strettamente imparentate con l'affinità della proteina variando in base a quanto la corrispondenza è corretta. Le sequenze cis-regolatorie mostrano l'intervallo di sequenze riconosciute da un regolatore di trascrizione particolare. La sequenza di DNA riconosciuta da un monomero non ha informazioni sufficienti per essere scelta da tutti le sequenze randomiche nel genoma e si rendono necessarie altre contribuzioni per aumentare la specificità. Molti regolatori

### 8.3. I REGOLATORI DI TRASCRIZIONE ATTIVANO E DISATTIVANO I GENI

---

di trascrizione formano dimeri con entrambi i monomeri facendo contatti molto simili con il DNA. In questo modo si raddoppia la lunghezza della sequenza riconosciuta e si aumenta l'affinità e la specificità del legame dei regolatori. Spesso si formano eterodimeri tra due regolatori diversi con più di una proteina in modo da creare diverse specificità diverse.

#### 8.2.4 I regolatori di trascrizione si legano cooperativamente al DNA

Nel caso più semplice l'insieme dei legami covalenti che tiene insieme i dimeri è sufficiente per formare queste strutture obbligatoriamente e impedendo la loro separazione, caso in cui l'unità di legamento del dimero e la curva di legame per il regolatore di trascrizione ha una forma esponenziale standard. In molti casi i dimeri e gli eterodimeri sono tenuti insieme molto debolmente. Le proteine legano il DNA cooperativamente con una curva sigmoidale. Questo vuol dire che per un intervallo di concentrazioni del regolatore di trascrizione il legame è più un fenomeno binario che per legami non cooperativi, ovvero alla maggior parte delle concentrazioni la sequenza cis-regolatoria è o quasi vuota o quasi completamente occupata.

#### 8.2.5 La struttura del nucleosoma promuove legami cooperativi di regolatori di trascrizione

Esiste un meccanismo secondario e indiretto per favorire i legami cooperativi che nasce dalla struttura nucleosomica dei cromosomi eucariotici. In generale i regolatori di trascrizione si legano al DNA nei nucleosomi con affinità minore rispetto al DNA da solo in quanto la superficie potrebbe essere contro il nucleosoma non disponibile o i cambi che i regolatori compiono al DNA sono contrastati dal legame stretto del DNA lungo il nucleo istonico. Anche senza rimodellazione i regolatori possono avere accesso limitato nel DNA che alla fine di un nucleosoma si espone temporaneamente e permette il legame dei regolatori. Queste proprietà del nucleosoma promuovono legami cooperativi in quanto se un regolatore entra nel DNA di un nucleosoma previene il suo restringimento aumentando l'affinità per un secondo regolatore. Se i due interagiscono tra di loro l'effetto cooperativo è maggiore arrivando in alcuni casi a separare l'istone. La cooperazione aumenta quando sono coinvolti complessi di rimodellazione del nucleosoma. Se un regolatore di trascrizione si lega alla sequenza cis-regolatoria e attrae un complesso di rimodellazione della cromatina, l'azione localizzata di quest'ultimo permette a un secondo regolatore di legarsi efficientemente vicino.

### 8.3 I regolatori di trascrizione attivano e disattivano i geni

#### 8.3.1 Il repressore triptofano disattiva i geni

Il genoma nel batterio *E. coli* consiste di una molecola di DNA di  $4.6 \cdot 10^6$  coppie di nucleotidi e codifica 4300 proteine. L'espressione dei geni viene regolata in base alla disponibilità di nutrienti nell'ambiente. Ci sono 5 geni che codificano per l'amminoacido triptofano in un cluster sul cromosoma e sono trascritti da un singolo promotore su una molecola di mRNA detta operone. Gli operoni sono rari negli eucarioti, dove i geni sono trascritti e regolati individualmente. Quando la concentrazione di triptofano è bassa l'operone è trascritto producendo un insieme di enzimi biosintetici che lavorano insieme per produrre triptofano da molecole più semplici. Quando il triptofano è abbondante l'amminoacido è importato nella cellula e la produzione di questi enzimi viene bloccata. Nel promotore dell'operone è presente una sequenza cis-regolatrice che è riconosciuta da un regolatore di trascrizione che quando si lega alla sequenza blocca l'accesso della RNA polimerasi al

### 8.3. I REGOLATORI DI TRASCRIZIONE ATTIVANO E DISATTIVANO I GENI

---

promotore impedendo la trascrizione. Questo regolatore è detto repressore del triptofano e la sequenza cis-regolatrice triptofano operatore. Il repressore può legarsi al DNA solo se ha anche legate molte molecole di triptofano. Il repressore è una proteina allosterica e il legame con il triptofano causa un cambio conformazionale in modo che possa legarsi alla sequenza operatore. Quando la concentrazione di triptofano diminuisce questo si disassocia dal repressore che non può più rimanere legato al DNA.

#### 8.3.2 I repressori disattivano i geni e gli attivatori li attivano

Le proteine repressori trascrizionali disattivano i geni o li reprimono. Alcuni regolatori attivano i geni. Questi attivatori trascrizionali lavorano con il promotore che è solo marginalmente capace di legarsi a sequenze cis-regolatorie e contatta la RNA polimerasi per aiutarla a iniziare la trascrizione. Queste proteine attivatrici legate al DNA possono aumentare il tasso anche di 1000 volte. Queste proteine devono interagire con una seconda molecola per riuscire a legarsi al DNA: l'attivatore batterico CAP deve legare AMP ciclici prima che possa legarsi al DNA. I geni attivati da esso vengono trascritti in risposta a un aumento della concentrazione di cAMP che aumenta quando il glucosio non è più disponibile, guidando la produzione di enzimi che permettono al batterio di digerire altri zuccheri.

#### 8.3.3 Un attivatore e un repressore controllano l'operone Lac

In molte istanze l'attività di un singolo promotore è controllata da diversi regolatori di trascrizione. L'operone Lac è controllato sia dai repressori Lac e dagli attivatori CAP. L'operone Lac codifica le proteine richieste per l'importazione e la digestione del lattosio. In assenza di glucosio il batterio crea cAMP che attiva CAP attivando i geni che permettono alla cellula di utilizzare alternative fonti di carbonio come il lattosio. Il repressore Lac disattiva l'operone in assenza di lattosio. Questo permette al controllo della regione dell'operone Lac di integrare due segnali diversi in modo che il gene è altamente espresso solo quando il glucosio è assente e il lattosio è presente. Tutti i regolatori di trascrizione devono legarsi al DNA per sortire un effetto. In questo modo ogni proteina regolatoria agisce selettivamente, controllando solo i geni che portano una sequenza cis-regolatoria riconosciuta da essa.

#### 8.3.4 L'inanellamento del DNA può avvenire durante la regolaizione genica dei batteri

Gli attivatori e i repressori sono molto simili in quanto devono riconoscere la sequenza cis-regolatorio attraverso lo stesso motivo strutturale. Alcune protein possono pertanto agire sia come repressori che attivatori in base al loro posizionamento esatto nella sequenza. La maggior parte dei batteri hanno un genoma piccolo e compatto e le sequenze cis-regolatorie si trovano molto vicino al punto di inizio della trascrizione, con delle eccezioni in cui è distante centinaia o migliaia di coppie. In questi casi il DNA è inanellato permettendo a un legame proteico ad un sito distante di contattare la RNA polimerasi e il DNA si comporta come un legame aumentando la probabilità che le proteine collidano.

#### 8.3.5 Interruttori complessi controllano la trascrizione genica negli eucarioti

I regolatori di trascrizione negli eucarioti coinvolgono molte proteine lunghe sequenze di DNA. Come nei batteri il tempo e luogo della trascrizione di un gene è regolato da sequenze cis-regolatorie che

sono lette dai regolatori di trascrizione che si legano a esse. Gli attivatori aiutano a legare la RNA polimerasi ad iniziare la trascrizione e i repressori bloccano il processo. Queste interazioni sono per la maggior parte indirette con molte proteine intermediarie come gli istoni. Negli organismi multicellulari dozzine di regolatori di trascrizione controllano un singolo gene con sequenze cis-regolatorie diffuse in decine di migliaia di paia di nucleotidi. L'inallentamento del DNA permette le proteine regolatorie di interagire tra di loro e con la polimerasi al promotore. Infine l'iniziazione della trascrizione deve superare il blocco imposto dai nucleosomi e strutture di livello superiore.

#### 8.3.6 Una regione di controllo di un gene eucariotico consiste di un promotore e molte sequenze cis-regolatorie

Negli eucarioti la RNA polimerasi II trascrive tutti i geni codificatori delle proteine e molti non codificatori e richiede 5 fattori di trascrizione generali il cui assemblaggio mette a disposizione multipli passi per il regolamento della trascrizione. Si usa il termine regione del controllo del gene per descrivere l'intera lunghezza di DNA coinvolta nella regolazione e iniziazione della trascrizione che include il promotore, dove i fattori generali di trascrizione e la polimerasi si assemblano e tutte le sequenze cis-regolatorie in cui i regolatori si legano per controllare il tasso di assemblaggio al promotore. Alcune parti delle regioni regolatorie sono trascritte come lncRNA e si possono considerare come sequenze spaziatrici che i regolatori di trascrizione non riconoscono direttamente. In contrasto con il piccolo numero di fattori generali di trascrizione ci sono migliaia di diversi regolatori di trascrizione che attivano o disattivano un singolo gene. Negli eucarioti gli operoni sono rari e ogni gene è regolato individualmente.

#### 8.3.7 I regolatori di trascrizione eucariotici lavorano in gruppi

I regolatori di trascrizione eucariotici si assemblano in gruppi alle loro sequenze cis-regolatorie con spesso interazioni cooperative. Oltre a questi sono presenti proteine coattivatrici o corepressori che si assemblano sul DNA con essi e non riconoscono specifiche sequenze di DNA e sono portati in queste posizioni dai regolatori di trascrizione. Queste interazioni proteina-proteina sono troppo deboli per avvenire in soluzione ma possono cristallizzarsi con l'appropriata combinazione di sequenze cis-regolatorie. I coattuatori e i corepressori influenzano la trascrizione dopo che sono stati localizzati sul genoma dai regolatori. Un regolatore può partecipare in più di un tipo di complesso regolatorio e funzionano pertanto come parti di un complesso che reprime la trascrizione. Ogni gene eucariotico è regolato da un insieme di proteine che devono essere tutte presenti per esprimere il gene al livello appropriato.

#### 8.3.8 Le proteine attivatrici promuovono l'assemblaggio di RNA polimerasi al punto di inizio della trascrizione

Le sequenze cis-regolatorie in cui le proteine attivatrici si legano aumentano il tasso di iniziazione della trascrizione attraendo e posizionando RNA polimerasi II al promotore e rilasciarla in modo che il processo inizi. Alcune proteine attivatrici si legano direttamente a fattori di trascrizione generali velocizzando il loro assemblaggio al promotore che è stato portato in prossimità grazie all'inallentamento del DNA. La maggior parte degli attivatori attraggono coattuatori che poi svolgono il compito biochimico per iniziare la trascrizione. Uno di questi è il complesso di proteine mediatore composto da più di 30 subunità e serve come ponte tra gli attivatori legati al DNA, la RNA polimerasi e i fattori di trascrizione generali facilitando il loro assemblaggio al promotore.



#### 8.3.9 Gli attivatori di trascrizione eucariotici direzionano la modifica di strutture cromatiniche locali

I fattori generali di trascrizione e l'RNA polimerasi sono incapaci di assemblarsi su un promotore che è condensato in un nucleosoma. Gli attivatori devono pertanto svolgere questo compito causando cambi alla struttura cromatinica rendendo il DNA più accessibile. Il modo più importante per compiere questo è attraverso cambi locali con modifiche covalenti degli istoni, rimodellamento dei nucleosomi, la loro rimozione e sostituzione. Gli attivatori usano tutti questi meccanismi e attraggono coattivatori che includono enzimi di modifica degli istoni, complessi di rimodellamento della cromatina dipendenti dall'ATP e accompagnatori di istoni, ognuno dei quali può alterare la struttura cromatinica del promotore. L'alterazione della struttura cromatinica può persistere per diverso tempo: in alcuni casi le modifiche sono eliminate appena i regolatori si disassociano dal DNA, cosa importante per geni che la cellula deve rapidamente attivare o disattivare in risposta a segnali esterni. In altri casi la struttura persiste in modo da estendere questa informazione alle generazioni successive e modificare la memoria dei pattern di espressione genica. Un tipo speciale di modifica cromatinica accade quando l'RNA polimerasi trascrive un gene: gli istoni dopo la polimerasi possono essere acetilati da enzimi, rimossi dagli accompagnatori degli istoni e depositati dietro la polimerasi. Questi istoni sono poi deacetilati e metilati rapidamente da complessi trasportati dalla polimerasi lasciando nucleosomi resistenti a trascrizione. Questo processo sembra prevenisca reiniziazioni spurie dietro a una polimerasi che deve crearsi una via attraverso la cromatina.

#### 8.3.10 Gli attivatori di trascrizione possono promuoverla rilasciando la RNA polimerasi dai promotori

In alcuni casi l'iniziazione della trascrizione richiede che un attivatore di trascrizione legato al DNA rilasci la RNA polimerasi dal promotore in modo da permettere l'inizio della trascrizione. In altri casi la polimerasi si blocca dopo aver trascritto 50 nucleotidi e ulteriore allungamento richiede un attivatore legato dietro a essa. Queste polimerasi bloccate sono comuni negli esseri umani. Il rilascio della RNA polimerasi può avvenire attraverso un complesso di rimodellamento della cromatina che rimuove un blocco nucleosomico o attraverso un attivatore che comunica con la RNA polimerasi tipicamente attraverso un coattivatore segnalando di procedere. Infine i fattori di allungamento permettono alla polimerasi di trascrivere attraverso la cromatina che in alcuni casi è il passo chiave per il controllo. Una volta che i fattori di allungamento sono caricati permettono alla polimerasi di muoversi attraverso i blocchi cromatinici. Avere una RNA polimerasi già posizionata ad un promotore permette di bypassare il passaggio di assemblaggio e di rispondere più velocemente a segnali esterni.

#### 8.3.11 Gli attivatori di trascrizione lavorano sinergicamente

I complessi di attivatori e coattivatori si assemblano cooperativamente sul DNA e possono promuovere diversi passi dell'iniziazione di trascrizione. L'aumento del tasso di reazione finale dovuto a un insieme di attivatori è il prodotto dei contributi dei singoli. Si nota pertanto come gli attivatori esibiscano sinergia trascrizionale.

### 8.3.12 Repressori di trascrizione eucariotici possono inibire la trascrizione in molti modi

I repressori di trascrizione possono diminuire il tasso di trascrizione e disattivare geni rapidamente. Grandi regioni di genoma possono essere disattivate dalla condensazione del DNA in forme cromatiniche specialmente resistenti, ma la maggior parte non sono organizzati per funzione e pertanto non è una strategia sempre disponibile. La maggior parte dei repressori lavorano gene per gene. Non competono direttamente con la RNA polimerasi per l'accesso al DNA ma usano molti meccanismi che bloccano la trascrizione. Tipicamente agiscono portando un corepressore al DNA. I repressori possono agire su più di un meccanismo ad un gene. La repressione è specialmente importante per organismi la cui crescita dipende da programmi di sviluppo elaborati e complessi.

### 8.3.13 Sequenze di DNA isolanti impediscono a regolatori di influenzare geni distanti

Per evitare che regolatori di geni diversi si influiscano tra di loro elementi del DNA compartimentalizzano il DNA in domini regolatori discreti. Sono presenti sequenze barriera che impediscono la diffusione di eterocromatina in geni che devono essere espressi. Un isolatore impedisce a sequenze cis-regolatorie di attivare geni inappropriati. Funzionano formando anelli di cromatina attraverso proteine specializzate che si legano a loro. Gli anelli tengono un gene e la sua regione di controllo in prossimità e aiutano a prevenire l'uscita di una regione di controllo in geni adiacenti. La distribuzione degli isolatori e sequenze barriera si pensa divida il genoma in domini indipendenti di regolazione genica e struttura cromatinica.

### 8.3.14 Meccanismi genetici molecolari che creano e mantengono tipi di cellula specializzati

Le cellule di organismi multicellulari mantengono la scelta della differenziazione in un tipo di cellula specifico, ricordando i cambi nell'espressione genica coinvolti nella scelta. Questa memoria è un prerequisito per la creazione di tessuti organizzati e per il mantenimento di tipi cellulari stabilmente differenziati.

### 8.3.15 Complessi interruttori genici che regolano lo sviluppo della *Drosophila* sono creati da molecole più piccole

Considerando il gene *Even-skipped* (*Eve*) della *Drosophila* la cui espressione è fondamentale per lo sviluppo dell'embrione si nota come se il gene è disattivato per mutazione, molte parti dell'embrione non si formano e questo muore. Quando *Eve* comincia ad essere espresso l'embrione è una singola cellula gigante contenente nuclei multipli in un citoplasma comune che contiene anche un insieme di regolatori di trascrizione distribuiti lungo la lunghezza dell'embrione fornendo informazioni posizionali che distinguono una parte dell'embrione dall'altra. I nuclei cominciano rapidamente ad esprimere geni diversi in quanto sono esposti a differenti regolatori. Le sequenze di DNA regolatorie che controllano il gene *Eve* si sono evolute per leggere le concentrazioni di regolatori di trascrizione ad ogni posizione e causano l'espressione del gene in sette fasce precisamente posizionate, ognuna larga tra 5 e 6 nuclei. La regione regolatoria del gene *Eve* è molto lunga ed è formata da una serie di moduli regolatori con una semplice sequenza cis-regolatoria responsabile per la specificazione di una particolare fascia di espressione lungo l'embrione.

#### 8.3.16 Il gene Eve della Drosophila è regolato da controlli combinatori

Il modello del modulo regolatorio della seconda faccia contiene sequenze di riconoscimento per due regolatori di trascrizione (Bicoid e Hunchback) che attivano la trascrizione di Eve e per altri due (Küppel e Giant) che la reprimono. La concentrazione relativa di queste quattro proteine determina se il complesso presente alla fascia attiva la trascrizione. Tale elemento, come quelli di tutte le altre fasce sono autonomi. L'intera regione di controllo lega più di 20 regolatori di trascrizione. 7 combinazioni di regolatori ne specificano l'espressione, mentre altre combinazioni lo mantengono silente. La regione di controllo è pertanto costituita da una serie di moduli più piccoli, ognuno dei quali consiste di un unico ordinamento di corte sequenze cis-regolatorie riconosciute da regolatori di trascrizione specifici. Il gene Eve codifica un regolatore di trascrizione che controlla l'espressione di altri geni. L'embrione viene diviso in regioni sempre più specifiche fino a che costituisce le parti del corpo della Drosophila.

#### 8.3.17 I regolatori di trascrizione sono messi ingioco da segnali extracellulari

Nella maggior parte degli embrioni e nelle cellule adulte i nuclei si trovano in cellule separate e si necessita un passaggio di informazioni extracellulari attraverso la membrana plasmatica per generare segnali nel citosol che causano l'attivazione di diversi regolatori.

#### 8.3.18 Il controllo dei geni combinatorio diversi tipi di cellule

Ogni regolatore di trascrizione in un organismo contribuisce al controllo di molti geni. A causa di questo controllo combinatorio un regolatore di trascrizione non ha una funzione definita come un comandante di uno specifico insieme di geni o di un tipo cellulare, ma è la sua combinazione relativa con altri a stabilire quest'informazione. Il controllo genico combinatorio causa l'addizione di un nuovo regolatore di trascrizione che dipende dalla storia precedente. Durante lo sviluppo una cellula accumula una serie di regolatori che ne alterano l'espressione genica solo quando l'ultimo membro presente viene aggiunto. Questo meccanismo permette ad alcune cellule di convertire da un tipo all'altro.

#### 8.3.19 Combinazioni di regolatori di trascrizione master specificano il tempo della cellula controllando l'espressione di molti geni

I pattern di espressione del tipo cellulare sono determinati da una combinazione di regolatori di trascrizione master che in molti casi si legano direttamente a sequenze cis-regolatorie del gene. MyiD si lega direttamente alla sequenza cis-regolatoria delle regioni di controllo sui geni specifici ai muscoli. In altri casi i regolatori master controllano l'espressione di regolatori a valle che si legano alle regioni di controllo di altri geni specifici e controllano la loro sintesi. La specifica di un particolare tipo di cellula coinvolge cambi nell'espressione di migliaia di geni: quelli richiesti dal particolare tipo sono prodotti, gli altri no.

#### 8.3.20 Cellule specializzate devono rapidamente attivare e disattivare insieme di geni

Le cellule specializzate devono rispondere ai cambi nell'ambiente come segnali da altre cellule che coordinano il comportamento dell'intero organismo. Molti di questi segnali inducono cambi temporanei alla trascrizione dei geni. L'effetto di un singolo regolatore può avere effetto, completando la

#### 8.4. MECCANISMI CHE RINFORZANO LA MEMORIA CELLULARE IN PIANTE E ANIMALI

---

combinazione necessaria per attivare o reprimere il gene. Un esempio è la proteina glucocorticoide-recettrice umana. Per elgarsi alla sua sequenza cis-regolatoria deve prima formare un complesso con un ormone glucocorticoide steriode come il cortisolo che viene rilasciato dal corpo durante periodi di fame e intensa attività ficia, stimolando le cellule del fegato ad aumentare la produzione di glucosio dagli amminoacidi. Tali cellule aumentano l'espressione di molti geni che codificano gli enzimi metabolici. Nonostante questi geni abbiano regioni di controllo diverse e complesse la loro espressione massimale dipende dal legame del complesso ormone-glucocorticoide alla sua sequenza cis-regolatoria, presente nella regione di controllo di ogni gene.

##### 8.3.21 Cellule differenziate mantengono la loro identità

Una volta che una cellula si è differenziata generalmente rimarrà differenziata e tutta la sua progenie rimarrà dello stesso tipo. Alcune cellule altamente specializzate non si dividono più, ovvero si differenziano terminalmente. Molte altre cellule differenziate si dividono molte volte nella vita di un individuo generando progenie dello stesso tipo. Affinchè il tipo venga mantenuto (memoria cellulare) i pattern di espressione genica responsabili devono essere mantenuti e passati alle figlie durante le divisioni. Questo viene ottenuto attraverso un loop a feedback positivo, dove un regolatore di trascrizione master attiva la trascrizione per il proprio gene e a tutti gli altri specifici al sito. Ogni volta che la cellula si divide questo viene passato alle figlie.

##### 8.3.22 Circuiti di trascrizione permettono alla cellula di compiere operazioni logiche

Semplici regolatori possono essere combinati per creare dispositivi di controllo che se ordinati in motivi di rete possono essere trovati continuamente in cellule di specie diverse. Sono comuni anelli di feedback positivi e negativi. Con più regolatori i comportamenti dei circuiti possono diventare più complessi come flip-flops, loop a feed-forward.

### 8.4 Meccanismi che rinforzano la memoria cellulare in piante e animali

#### 8.4.1 Pattern di metilazione del DNA possono essere ereditati quando le cellule dei vertebrati si dividono

Nelle cellule dei vertebrati la metilazione della citosina fornisce un meccanismo attraverso cui i pattern di espressione genica possono essere passati alla progenie della cellula. La 5-metil citosina ha la stessa relazione con la citosina che la timina ha con l'uracile e la modifica non ha effetto sull'accoppiamento delle basi. La metilazione del DNA avviene principalmente in sequenze CG accoppiate alla stessa sequenza opposta all'altro filamento. Un semplice meccanismo permette a questo pattern di essere ereditato. L'enzima metil trasferasi di manutenzione agisce sulle sequenze già metilate e il pattern di metilazione serve come stampo per la metilazione del filamento figlio. Tali pattern sono dinamici durante lo sviluppo. Dopo la fertilizzazione c'è un'onda di metilazione dove la maggior parte dei gruppi metili sono persi dal DNA attraverso soppressione dell'attività della metil trasferasi o grazie a enzimi di demetilazione. Successivamente nello sviluppo nuovi pattern sono stabiliti da nuove DNA metil trasferasi direzionate al DNA da proteine specifiche alla sequenza. Una volta che i nuovi pattern sono stabiliti possono essere propagati attraverso la replicazione. La metilazione de DNA lavora in congiunzione con altri meccanismi di controllo genici

#### 8.4. MECCANISMI CHE RINFORZANO LA MEMORIA CELLULARE IN PIANTE E ANIMALI

---

per stabilire una forma efficiente di repressione genica. In questo modo i geni eucariotici possono essere altamente repressi. I gruppi metili sulle citosine si trovano nella scanalatura maggiore del DNA e interferiscono direttamente con le proteine di legame richieste per l'iniziazione della trascrizione e la cellula contiene un insieme di proteine che si legano a tali citosine come enzimi di modifica degli istoni che portano a stati di cromatina repressivi dove la metilazione e la struttura cromatinica agiscono sinergicamente.

##### 8.4.2 Isole ricche di CG sono associate con molti geni nei mammiferi

La citosina metilata tende ad essere eliminata nei genomi dei vertebrati in quanto deaminazione accidentale di un non metilato C genera U che viene riconosciuto dalla DNA glicosilasi, esportato e sostituito con un C. Deaminazione accidentale di una citosina metilata non può essere riparata in quanto il prodotto è una T. Esiste pertanto uno speciale sistema di riparazione per rimuovere queste T mutanti che non sono precisi. Le sequenze CG rimanenti sono distribuite non uniformemente nel genoma e maggiormente nelle isole CG e tipicamente includono promotori dei geni. Sono maggiormente presenti nei geni che codificano proteine necessarie alla vita della cellula. Isole CG rimangono non metilate nella maggior parte delle cellule somatiche e tale stato è mantenuto da proteine specifiche alla sequenza. Legandosi a queste sequenze proteggono il DNA dalla metiltransferasi e reclutano DNA demetilasi che converte la 5-metil C in idrossi-metil C che è successivamente sostituita da C attraverso riparazione del DNA o passivamente. Tali isole permettono ad alcune proteine che si legano alle isole di CG e le proteggono dall'enzima modificatore di istoni che recluta la metilazione e rendono l'isola amichevole ai promotori. La RNA polimerasi è pertanto spesso legata a promotori nelle isole di CG. La polimerasi si trova pertanto maggiormente rispetto ai nucleosomi. Altri passi sono necessari per iniziare la trascrizione e sono diretti da regolatori che si legano alle sequenze cis-regolatorie del DNA che rilasciano la polimerasi con i fattori di allungamento.

##### 8.4.3 L'imprinting genomico si basa sulla metilazione del DNA

Le cellule dei mammiferi sono diplodi e l'espressione di una minoranza dei geni dipende dal fatto se sono stati ereditati dalla madre o dal padre: quando la copia del gene paterno è attiva quello materno è disattivo o viceversa. Questo fenomeno è detto imprinting genomico. L'imprinting può smascherare mutazioni coperte dall'altra coppia funzionale. Nel primo embrione i geni soggetti a imprinting sono marcati da metilazione secondo il fatto che siano derivati da un cromosoma dello sperma o dell'uovo. In questo modo la metilazione del DNA viene usata per distinguere due geni altrimenti identici. In qualche modo sono protetti dall'onda di demetilazione le cellule somatiche si ricordano l'origine parentale di ogni gene e regolano la loro espressione secondo questo. Nella maggior parte dei casi vengono silenziati geni vicini, ma in alcuni casi possono essere attivati. In altri casi l'imprinting coinvolge long noncoding RNA definiti come molecole di RNA più lunghe di 200 nucleotidi che non codificano le proteine. Nel caso del gene *Kcnq1* che codifica per un canale di calcio legato al voltaggio necessario per la funzione del cuore l'ncRNA è fatto dall'allele paterno non metilato ma non è rilasciato dalla RNA polimerasi e rimane nel sito della sintesi e recluta enzimi modificatori degli istoni e per la metilazione del DNA che direzionano la formazione di cromatina repressiva che silenzia il gene associato sul cromosoma derivato dal padre. Il gene materno è immune a questi effetti.

### 8.4.4 Alterazioni globali al cromosoma nella struttura cromatinica possono essere ereditate

Nei mammiferi alterazioni nella struttura cromatinica di un cromosoma possono modulare il livello di espressione della maggior parte dei geni su quel cromosoma. Maschi e femmine differiscono nei loro cromosomi sessuali: le femmine possiedono due cromosomi X mentre i maschi un X e un Y. La cellula femmina contiene pertanto il doppio delle coppie dei geni del cromosoma X, diverso dall'Y: il primo è grande e contiene più di mille geni, mentre il secondo è piccolo e ne contiene meno di 100. Si è evoluto un meccanismo di compensazione del dosaggio per equalizzare il dosaggio di prodotti dei geni del cromosoma X tra maschi e femmine. Il tasso corretto dei geni del cromosoma X a autosoma è controllato e mutazioni che lo coinvolgono sono generalmente letali. I mammiferi ottengono questa compensazione disattivando uno dei due cromosomi X nelle cellule femminili somatiche nella disattivazione-X: due cromosomi X possono coesistere nello stesso nucleo e differire completamente nella loro espressione. In un primo momento durante lo sviluppo dell'embrione femminile uno dei due cromosomi X in ogni cellula diventa altamente condensato in un tipo di eterocromatina. La scelta iniziale è casuale e una volta che viene disattivato rimane silente attraverso tutte le cellule sequenti. A causa della casualità ogni femmina è un mosaico di gruppi clonali di cellule in cui uno dei due è silenziato. Questi gruppi si trovano in piccoli cluster. La disattivazione del cromosoma X inizia e si diffonde da un sito vicino al centro del cromosoma, il centro di inattivazione X (XIC) dove è trascritto un lncRNA (Xist) che è espresso unicamente dal cromosoma inattivo. L'RNA Xist si diffonde dal XIC all'intero cromosoma e guida il silenziamento. Lo Xist che si diffonde è prima passato attraverso la base degli anelli che creano i cromosomi. Imprinting e disattivazione del cromosoma X sono esempi di espressione genetica monoallelica.

### 8.4.5 Meccanismi epigenetici assicurano che pattern stabili di espressione genica possono essere trasmessi a cellule figlie

Il modo più semplice per una cellula di ricordare la sua identità è attraverso loop a feedback positivo in cui un regolatore di trascrizione chiave attiva la trascrizione del proprio gene. Tali loop concatenati forniscono grande stabilità bufferando il circuito contro fluttuazioni nel livello di qualcuno dei regolatori di trascrizione. Siccome i regolatori di trascrizione sono sintetizzati nel citosol e si diffondono attraverso il nucleo i loop hanno effetto su entrambe le copie del gene nella cellula diploide, ma anche le differenze di livello di espressione tra lo stesso gene viene ereditata. L'abilità della cellula di mantenere la memoria di pattern di espressione genica è un esempio di ereditarietà epigenetica: un'alterazione nel fenotipo ereditabile che non risulta da cambi nella sequenza del DNA.

## 8.5 Controlli post trascrizionali

I controlli post trascrizionali operano dopo che la RNA polimerasi si è legata al promotore del gene e ha iniziato la sintesi sono cruciali per la sintesi di molti geni.

### 8.5.1 L'attenuazione della trascrizione causa la terminazione prematura di alcune molecole di RNA

L'espressione di alcuni geni è inibita da terminazione prematura della trascrizione, fenomeno detto attenuazione della trascrizione. In alcuni di questi casi la catena di RNA adotta una struttura che causa la sua interazione con la polimerasi in modo da abortire la sua trascrizione. Quando il gene è richiesto proteine regolatorie si legano alla catena e rimuovono l'attenuazione. Un esempio di

questo avviene durante il ciclo dell'HIV in cui il DNA virale è trascritto dalla polimerasi II che di solito termina prima di averne sintetizzato l'intero genoma. Quando le condizioni per la crescita virale sono ottime una proteina codificata dal virus Tat si lega a una struttura specifica dell'RNA che contiene una base rigonfia impedendo la terminazione prematura.

### 8.5.2 Riboswitches rappresentano probabilmente antiche forme di controllo genico

I riboswitches dimostrano come anche l'RNA può formare dispositivi di controllo: piccole sequenze di RNA che cambiano la loro conformazione quando si legano a piccole molecole come i metaboliti. Ogni riboswitch riconosce una piccola molecola e il cambio conformazionale viene utilizzato per regolare l'espressione dei geni. Sono spesso localizzati vicino alla terminazione 5' dell'mRNA e si piegano mentre viene sintetizzato bloccando o permettendo il progresso della RNA polimerasi. Sono comuni nei batteri, altamente specifici. Non necessitano proteine regolatorie. Controllano l'allungamento della trascrizione.

### 8.5.3 RNA splicing alternativo può produrre diverse forme di proteine dallo stesso gene

L'RNA splicing accorcia il trascritto di molti geni rimuovendo sequenze di introni dal precursore. Una cellula può fare splicing sullo stesso precursore in maniera diversa creando diverse catene polipeptidiche durenate RNA splicing alternativo. Quando esistono diverse possibilità di splicing un singolo gene può produrre dozzine di diverse proteine. In alcuni casi accade a causa di ambiguità nella sequenza dell'introne: lo spliceosoma standard è incapace di distinguere tra due o più coppie alternative di accoppiamenti 5' e 3' di siti di splicing e possono essere fatte scelte diverse. In molti casi viene regolato per passare dalla produzione di una proteina non funzionale a una funzionale o viceversa. Oltre a questo può produrre diverse versioni di una proteina secondo la necessità della cellula. Lo splicing può essere regolato negativamente da una molecola regolatoria che impedisce al macchinario di accedere a un sito di splice particolare all'RNA o positivamente da una che aiuta a direzionare lo splicing. Si può pensare allo splicing di una molecola di pre-mRNA come a un equilibrio tra siti di splicing in competizione.

### 8.5.4 La definizione di un gene è stata modificata

Fino alla scoperta che i geni eucariotici contengono introni e che le loro sequenze codificanti possono essere assemblate in vari modi il gene era definito operazionalmente come una regione del genoma che segrega una singola unità durante la meiosi e dà origine a un tratto fenotipico definibile. Successivamente si è passati a una definizione secondo cui la maggior parte dei geni corrispondono a una regione del genoma che direziona la sintesi di un singolo enzima. La giorno d'oggi un gene viene definito come una sequenza di DNA che è trascritta come una singola unità e codifica un insieme di catene polipeptidiche strettamente imparentate (proteine isoformiche).

### 8.5.5 Un cambio nel sito della rottura del trascritto a RNA e dell'addizione Poli-A può cambiare la terminazione C di una proteina

La terminazione di una molecola di mRNA non è formata dalla sintesi dell'RNA da parte della polimerasi, ma risulta da una reazione di rottura catalizzata da altre proteine mentre il trascritto si allunga. Una cellula può controllare il sito di questa rottura cambiando la terminazione C della

proteina risultante. In alcuni casi si produce una versione troncata, in altri rottura alternativa e siti di poliadenilazione si trovano tra le sequenze di introni e viene alterato il pattern di splicing. Questo processo può produrre due proteine che divergono solo nella sequenza alla terminazione C. Un esempio di questo è il passaggio dalla forma legata a membrana a quella di antibiotico secreto durante lo sviluppo dei linfociti B. Presto nella vita del linfocita B l'anticorpo che produce è ancorato alla membrana plasmatica dove è un recettore per antigeni la cui presenza causa una sua moltiplicazione e l'inizio della secrezione di anticorpi. La forma secreta è identica a quella legata a membrana ad eccezione della terminazione C. In questa parte si trova una lunga stringa di amminoacidi idrofobici che attraversano il bistrato lipidico della membrana, mentre la forma secreta è corta e contiene amminoacidi idrofilici. Questo passaggio avviene a causa del cambio del sito di rottura e poliadenilazione causato da un aumento di concentrazione della proteina CstF che promuove la rottura dell'RNA: il primo sito di rottura e addizione poliA viene ignorato in quanto subottimo solitamente, ma quando viene attivata per produrre anticorpi viene considerato a causa di un aumento di concentrazione di proteina CstF.

### 8.5.6 Modifiche dell'RNA possono cambiare il significato del messaggio RNA

Il processo di RNA editing modifica la sequenza nucleotidica dei trascritti dopo che sono stati sintetizzati. Negli animali avvengono principalmente la deamminazione dell'adenina per produrre inosina (A-to-I editing) e la deamminazione della citosina per produrre uracile (C-to-U editing). Essendo che queste modifiche cambiano le proprietà di accoppiamento delle basi hanno profondi effetti sul significato dell'RNA. Se questo avviene in una regione codificante può cambiare la sequenza di amminoacidi o produrre una proteina troncata, nella sequenza intronica modificare i pattern di splicing, il trasporto dell'mRNA, l'efficienza della traduzione o l'accoppiamento di basi tra microRNA (miRNA) e i loro obiettivi. Il processo A-to-I è prevalente negli umani e enzimi ADAR (adenina deaminasi agente sull'RNA) lo svolge riconoscendo una struttura di RNA a doppio filamento che è formata dall'accoppiamento tra le basi e una sequenza complementare in altro luogo, tipicamente un introne. La struttura a doppiofilamento specifica se l'mRNA debba essere editato e dove debba essere fatto.

### 8.5.7 Il trasporto di RNA dal nucleo può essere regolato

Nei mammiferi solo circa un ventesimo della massa totale dell'RNA lascia il nucleo in quanto i residui del processamento e RNA danneggiati sono degradati nel nucleo. L'esportazione dal nucleo è ritardata fino a che il processamento è terminato e meccanismi che possono annullare questo passaggio possono essere usati per regolare l'espressione genica. Questa strategia forma le basi del trasporto nucleare controllato dell'mRNA.

### 8.5.8 Alcuni mRNA sono localizzati in regioni specifiche del citosol

Una volta che una molecola di mRNA esce nel citosol incontra un ribosoma che la traduce in una catena polipeptidica. Una volta che passa il test di correttezza viene rapidamente tradotta. Se la proteina è destinata alla secrezione o espressa nella superficie cellulare una sequenza di segnali alla terminazione N la direziona al reticolo endoplasmatico ER in cui componenti dell'apparato di ordinamento delle proteine riconoscono la sequenza di segnale appena emerge dal ribosoma e direziona l'intero complesso all'ER dove viene sintetizzata la catena rimanente. In altri casi ribosomi liberi nel citosol sintetizzano la proteina e segnali possono poi direzionarla in altri siti della cellula. Molti



mRNA sono direzionati a specifiche locazioni intracellulari prima che inizi la traduzione permettendo alla cellula di posizionarli vicino al sito dove la proteina è necessitata. Altri vantaggi includono stabilimenti di asimmetrie nel citosol e mRNA localizzati con controlli di traduzione permettono alla cellula di regolare l'espressione genica indipendentemente in regioni diverse. Altri meccanismi di localizzazione dell'mRNA richiedono specifici segnali sull'mRNA, concentrati nella regione non tradotta 3' o UTR, che si estende dal codone di stop. La localizzazione è tipicamente accoppiata con controlli trascrizionali per garantire che l'mRNA rimane quiescente fino a che è stato mosso nel luogo corretto. Le molecole di mRNA portano numerose marcature quando escono dal nucleo che segnalano il completamento del processamento e l'UTR 3' può essere pensato come un codice postale che direziona l'mRNA in diversi luoghi.

### 8.5.9 Le regioni non tradotte 5' e 3' controllano la traduzione dei propri mRNA

Una volta che l'mRNA viene sintetizzato si controlla il livello di prodotto proteico controllando l'iniziazione della traduzione. Nei mRNA batterici una lunghezza di nucleotidi conservata (sequenza Shine-Dalgarno) si trova pochi nucleotidi a monte del codone AUG iniziante. I controlli traduzionali sono svolti da proteine o molecole di RNA e coinvolgono esposizione o blocco della sequenza Shine-Dalgarno. Gli mRNA eucariotici non contengono tale sequenza ma la selezione di un codone AUG dipende dalla sua prossimità al cappuccio alla terminazione 5', il sito in cui la piccola subunità ribosomica si lega all'mRNA e inizia lo scano. I repressori si possono legare a tale terminazione e inibire l'iniziazione della traduzione. Altri repressori riconoscono sequenze nucleotidiche nella 3' UTR e interferiscono con la comunicazione tra il cappuccio 5' e la coda poli-A 3' o da miRNA che si legano all'mRNA e riducono l'output di proteine.

### 8.5.10 La fosforilazione di un fattore di iniziazione regola globalmente la sintesi proteica