

# Biologia molecolare della cellula

Giacomo Fantoni

Telegram: @GiacomoFantoni

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

**23 marzo 2020**

# Indice

<b>1</b>	<b>Introduzione</b>	<b>2</b>
1.1	Microscopia . . . . .	2
1.1.1	Risoluzione . . . . .	2
1.1.2	Tecniche di microscopia ottica . . . . .	3
1.2	Storia . . . . .	3
<b>2</b>	<b>Cellule e genomi</b>	<b>5</b>
2.1	Caratteristiche universali della vita sulla Terra . . . . .	5
2.1.1	Tutte le cellule conservano la loro informazione ereditaria nello stesso codice chimico lineare (DNA) . . . . .	5
2.1.2	Tutte le cellule replicano la loro informazione ereditaria mediante polimerizzazione su uno stampo . . . . .	6
2.1.3	Tutte le cellule trascivono porzioni della loro informazione ereditaria nella stessa forma intermedia (RNA) . . . . .	6
2.1.4	Tutte le cellule usano proteine come catalizzatori . . . . .	7
2.1.5	Tutte le cellule traducono RNA nello stesso modo . . . . .	7
2.1.6	Il frammento di informazione genica che corrisponde ad una proteina è un gene . . . . .	7
2.1.7	La vita richiede energia libera . . . . .	8
2.1.8	Tutte le cellule funzionano come fabbriche biochimiche che impiegano le stesse strutture molecolari base . . . . .	8
2.1.9	Tutte le cellule sono chiuse in una membrana plasmatica attraverso la quale nutrienti e materiali di scarto devono passare . . . . .	8
2.2	La diversità dei genomi e l'albero della vita . . . . .	8
2.2.1	Le cellule possono ottenere energia da una varietà di fonti di energia libera . . . . .	8
2.2.2	Alcune cellule fissano azoto e anidride carbonica per altre . . . . .	9
2.2.3	La più grande diversità biochimica esiste tra le cellule procariote . . . . .	9
2.2.4	L'albero della vita possiede tre rami principali: Batteri, Archei ed Eucarioti . . . . .	9
2.2.5	Alcuni geni evolvono rapidamente, altri sono altamente conservati . . . . .	9
2.2.6	Nuovi geni sono generati da geni preesistenti . . . . .	10
2.2.7	La duplicazione genica permette la creazione di famiglie di geni imparentati in una stessa cellula . . . . .	10
2.2.8	I geni possono essere trasferiti tra organismi . . . . .	10
2.2.9	Molti geni sono comuni tra tutti i rami primari dell'albero della vita . . . . .	11
2.2.10	Le mutazioni rivelano la funzione dei geni . . . . .	11
2.3	Informazione genetica negli eucarioti . . . . .	11
2.3.1	Le cellule eucariote potrebbero essersi originate come predatori . . . . .	11

2.3.2	Le moderne cellule eucariote si sono evolute da una simbiosi . . . . .	11
2.3.3	Gli eucarioti possiedono genomi ibridi . . . . .	12
2.3.4	I genomi eucarioti sono grandi . . . . .	12
2.3.5	Il genoma definisce il programma dello sviluppo multicellulare . . . . .	12
2.3.6	Molti eucarioti vivono come cellule solitarie . . . . .	13
<b>3</b>	<b>Chimica e bioenergetica della cellula</b>	<b>14</b>
3.1	Le componenti chimiche della cellula . . . . .	14
3.1.1	L'acqua è mantenuta da legami a idrogeno . . . . .	14
3.1.2	Quattro tipi di attrazione non covalente aiuta a unire le molecole nelle cellule	14
3.1.3	Alcune molecole polari formano acidi e basi in acqua . . . . .	15
3.1.4	La cellula è formata da composti di carbonio . . . . .	15
3.1.5	Le cellule contengono quattro principali famiglie di piccole molecole organiche	15
3.1.6	La chimica della cellula è dominata da macromolecole con proprietà notevoli	16
3.1.7	I legami non covalenti specificano la forma di una macromolecola e i suoi legami con le altre molecole . . . . .	16
3.2	Catalisi e utilizzo dell'energia da parte delle cellule . . . . .	16
3.2.1	Il metabolismo della cellula è organizzato dagli enzimi . . . . .	16
3.2.2	L'ordine biologico è reso possibile dal rilascio di calore dalla cellula . . . . .	17
3.2.3	Le cellule ottengono energia ossidando molecole organiche . . . . .	17
3.2.4	Ossidazione e riduzione coinvolgono il trasferimento di elettroni . . . . .	17
3.2.5	Gli enzimi abbassano la barriera di energia di attivazione che blocca la reazione chimica . . . . .	17
3.2.6	Gli enzimi possono guidare i substrati lungo specifici cammini di reazioni . .	18
3.2.7	Come gli enzimi trovano i loro substrati, la rapidità dei movimenti molecolari	18
3.2.8	Il cambio di energia libera di una reazione determina se può avvenire spontaneamente . . . . .	18
3.2.9	Molecole vettore attivate sono essenziali per la biosintesi . . . . .	19
3.2.10	L'ATP è il vettore attivo più utilizzato . . . . .	19
3.2.11	NADH e NADPH sono vettori di elettroni . . . . .	20
3.2.12	Esistono molte altre molecole vettori nella cellula . . . . .	20
3.2.13	La sintesi di polimeri biologici è guidata dall'idrolisi dell'ATP . . . . .	20
3.3	Come le cellule ottengono energia dai nutrienti . . . . .	21
3.3.1	La glicolisi è un cammino centrale per produrre ATP . . . . .	21
3.3.2	La fermentazione produce ATP in assenza di ossigeno . . . . .	21
3.3.3	La glicolisi mostra come gli enzimi accoppiano l'ossidazione alla conservazione di energia . . . . .	21
3.3.4	Gli organismi conservano le molecole nutrienti in serbatoi speciali . . . . .	22
3.3.5	Zuccheri e acidi grassi sono degradati in acetile CoA nei mitocondri . . . . .	22
3.3.6	Il ciclo di acido citrico genera NADH ossidando i gruppi acetili in CO <sub>2</sub> . . . .	22
3.3.7	Il trasporto di elettroni guida la sintesi della maggior parte dell'ATP nelle cellule	23
3.3.8	Amminoacidi e nucleotidi sono parte del ciclo azotato . . . . .	23
<b>4</b>	<b>Proteine</b>	<b>24</b>
4.1	Forma e struttura delle proteine . . . . .	24
4.1.1	La forma di una proteina è specificata dalla sua sequenza degli amminoacidi .	24
4.1.2	Le proteine si ripiegano nella conformazione a minore energia . . . . .	25
4.1.3	L' $\alpha$ -elica e il $\beta$ -foglietto sono pattern di piegamento comuni . . . . .	25

4.1.4	I domini proteici sono unità modulari da cui le proteine più grandi sono costruite	25
4.1.5	Poche delle catene polipeptidiche possibili saranno utili alla cellula . . . . .	25
4.1.6	Le proteine possono essere classificate in molte famiglie . . . . .	26
4.1.7	Alcuni domini proteici si trovano in molte proteine diverse . . . . .	26
4.1.8	Il genoma umano codifica un complesso insieme di proteine, rivelando che molto rimane sconosciuto . . . . .	26
4.1.9	Proteine più grandi contengono più di una catena polipeptidica . . . . .	27
4.1.10	Alcune proteine globulari formano lunghi filamenti elicoidali . . . . .	27
4.1.11	Molte proteine hanno forme allungate e fibrose . . . . .	27
4.1.12	Le proteine contengono una grande quantità di catene polipeptidiche intrinsecamente disordinate . . . . .	27
4.1.13	Legami incrociati covalenti stabilizzano proteine extracellulari . . . . .	27
4.1.14	Le molecole proteiche spesso svolgono la funzione di subunità per l'assemblaggio di grandi strutture . . . . .	28
4.1.15	I fattori di assemblaggio aiutano la formazione di complesse strutture biologiche	28
4.1.16	Fibrille amiloide possono formarsi da molte proteine . . . . .	28
4.1.17	Molte proteine contengono domini a bassa complessità che possono formare amiloidi reversibili . . . . .	29
4.2	Funzione delle proteine . . . . .	29
4.2.1	Tutte le proteine si legano ad altre molecole . . . . .	29
4.2.2	La conformazione superficiale di una proteina determina la sua chimica . . .	29
4.2.3	La comparazione di sottosequenze tra membri di una famiglia proteica evidenzia i siti di legame cruciali . . . . .	30
4.2.4	Le proteine si legano con altre proteine attraverso molti tipi di interfacce . .	30
4.2.5	I siti di legame degli anticorpi sono estremamente versatili . . . . .	30
4.2.6	La costante di equilibrio misura la forza di legame . . . . .	30
4.2.7	Gli enzimi sono catalizzatori potenti e molto specifici . . . . .	31
4.2.8	Il legame con il substrato è il primo passo nella catalisi enzimatica . . . . .	31
4.2.9	Gli enzimi velocizzano la reazione stabilizzando selettivamente stati transitivi	31
4.2.10	Gli enzimi possono usare simultaneamente catalizzatori acidi e basici . . . . .	31
4.2.11	Il lisozima illustra come un enzima funziona . . . . .	32
4.2.12	Molecole strettamente legate aggiungono altre funzioni alle proteine . . . . .	32
4.2.13	Complessi multienzima aiutano ad aumentare il tasso del metabolismo cellulare	32
4.2.14	La cellula regola le attività catalitiche dei propri enzimi . . . . .	33
4.2.15	Gli enzimi allosterici possiedono due o più siti di legame che interagiscono . .	33
4.2.16	Due ligandi i cui siti di legame sono accoppiati devono avere effetto reciprocamente sul legame dell'altro . . . . .	33
4.2.17	Insiemi di proteine simmetrici producono transizioni allosteriche cooperative	34
4.2.18	Molti cambi nelle proteine sono guidati dalla fosforilazione proteica . . . . .	34
4.2.19	Una cellula eucariotica contiene un gran numero di proteina chinasi e proteina fosfatasi . . . . .	34
4.2.20	La regolazione della proteina chinasi Src rivela come una proteina può funzionare come un microprocessore . . . . .	34
4.2.21	Le proteine che si legano e regolano il GTP sono onnipresenti regolatori . . .	35
4.2.22	Le proteine regolatorie GAP e GEF controllano l'attività delle proteine GTP-leganti determinando se è legato GTP o GDP . . . . .	35
4.2.23	Le proteine possono essere regolate dall'aggiunta covalente di altre proteine .	35

4.2.24	Un elaborato sistema di coniugazione di ubiquitina è utilizzato per marcare le proteine . . . . .	35
4.2.25	I complessi proteici con parti interscambiabili fanno un efficiente uso dell'informazione genica . . . . .	36
4.2.26	Una proteina GTP-legante mostra come grandi movimenti proteici siano possibili	36
4.2.27	Le proteine motrici producono grandi movimenti nelle cellule . . . . .	36
4.2.28	I trasportatori di membrana imbrigliano energia per pompare le molecole attraverso le membrane . . . . .	37
4.2.29	Le proteine formano grandi complessi che funzionano come macchinari proteici	37
4.2.30	Impalcature concentrano insieme di proteine che interagiscono . . . . .	37
4.2.31	Molte proteine sono controllate da modifiche covalenti che le dirigono verso siti specifici all'interno della cellula . . . . .	37
4.2.32	Una complessa rete di interazioni proteiche sottosta alle funzioni della cellula	37
<b>5</b>	<b>DNA, cromosomi e genomi</b>	<b>39</b>
5.1	La struttura e la funzione del DNA . . . . .	39
5.1.1	Una molecola di DNA consiste di due catene di nucleotidi complementari . .	39
5.1.2	La struttura del DNA mette a disposizione un meccanismo per l'ereditarietà	40
5.1.3	Negli eucarioti il DNA è racchiuso nel nucleo cellulare . . . . .	40
5.2	DNA cromosomico e il suo confezionamento nella fibra cromatina . . . . .	40
5.2.1	Il DNA eucariotico è confezionato in un insieme di cromosomi . . . . .	40
5.2.2	I cromosomi contengono lunghe stringhe di geni . . . . .	41
5.2.3	La sequenza nucleotidica del genoma umano mostra come i geni sono ordinati	41
5.2.4	Ogni molecola di DNA che forma un cromosoma lineare deve contenere un centromero, due telomeri e un'origine di replicazione . . . . .	41
5.2.5	Le molecole di DNA sono altamente conservate nei cromosomi . . . . .	42
5.2.6	I nucleosomi sono unità base della struttura del cromosoma eucariote . . . .	42
5.2.7	La struttura del nucleo del nucleosoma rivela come il DNA è condensato . . .	42
5.2.8	I nucleosomi hanno strutture dinamiche e sono soggetti a cambi catalizzati da complessi rimodellanti a cromatina ATP-dipendenti . . . . .	43
5.2.9	I nucleosomi sono tipicamente condensati in una fibra cromatinica . . . . .	43
5.3	La struttura e la funzione della cromatina . . . . .	43
5.3.1	L'eterocromatina è altamente organizzata e limita l'espressione genica . . . .	43
5.3.2	Lo stato eterocromatico è autopropagante . . . . .	44
5.3.3	I nuclei istonici sono modificati covalentemente a molti siti diversi . . . . .	44
5.3.4	La cromatina acquisisce varietà addizionale attraverso inserzioni specifiche al sito di un insieme di varianti di istoni . . . . .	44
5.3.5	Le modifiche covalenti e le varianti istoniche agiscono insieme per controllare le funzioni dei cromosomi . . . . .	44
5.3.6	Un complesso di proteine lettrici e scrittrici possono espandere modifiche alla cromatina lungo un cromosoma . . . . .	45
5.3.7	Barriere nella sequenza di DNA bloccano l'espansione dei complessi di lettura-scrittura separando domini di cromatina vicini . . . . .	45
5.3.8	La cromatina nei centromeri rivela come le varianti degli istoni possono creare strutture speciali . . . . .	45
5.3.9	Alcune strutture cromatiniche possono essere ereditate direttamente . . . . .	46
5.3.10	L'attivazione e la repressione di strutture di cromatina può essere ereditato epigeneticamente . . . . .	46

5.3.11	Le strutture di cromatina sono importanti per le funzioni del cromosoma eucariotico . . . . .	46
5.4	La struttura globale dei cromosomi . . . . .	46
5.4.1	I cromosomi sono piegati in grandi anelli di cromatina . . . . .	46
5.4.2	Cromosomi politene sono utili per visualizzare le strutture di cromatina . . .	46
5.4.3	Esistono multiple forme di cromatina . . . . .	47
5.4.4	Gli anelli di cromatina decondensano quando i geni al loro interno sono espressi	47
5.4.5	La cromatina può muoversi a siti specifici nel nucleo per alterare l'espressione dei geni . . . . .	47
5.4.6	Reti di macromolecole formano un insieme di ambienti biochimici distinti nel nucleo . . . . .	48
5.4.7	I cromosomi mitotici sono altamente condensati . . . . .	48
5.5	Come si evolvono i genomi . . . . .	48
5.5.1	Il confronto tra i genomi rivela sequenze di DNA attraverso la loro evoluzione del throughput di evoluzione . . . . .	49
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 . . .	49
5.5.3	La sequenza genomica di due specie differisce in proporzione al tempo in cui si sono evolute separatamente . . . . .	49
5.5.4	L'albero filogenetico costruito dal confronto delle sequenze di DNA traccia le relazioni di tutti gli organismi . . . . .	49
5.5.5	Un confronto tra cromosomi umani e dei topi mostra come la struttura dei genomi diverge . . . . .	50
5.5.6	La dimensione del genoma dei vertebrati riflette i tassi relativi di addizione e perdita del DNA in un lignaggio . . . . .	50
5.5.7	Si può inferire la sequenza di alcuni genomi antichi . . . . .	50
5.5.8	Il confronto tra sequenze multispecie identifica sequenze di DNA conservato di funzione sconosciuta . . . . .	51
5.5.9	Cambi in sequenze precedentemente conservate possono aiutare a decifrare passi critici dell'evoluzione . . . . .	51
5.5.10	Mutazioni nella sequenza di DNA che controlla l'espressione genica ha guidato molti dei cambi evolutivi nei vertebrati . . . . .	51
5.5.11	La duplicazione genica fornisce una fonte di novità genica durante l'evoluzione	51
5.5.12	Geni duplicati divergono . . . . .	52
5.5.13	L'evoluzione del gene globina mostra come la duplicazione del DNA contribuisce all'evoluzione degli organismi . . . . .	52
5.5.14	I geni che codificano nuove proteine possono essere creati ricombinando gli esoni	52
5.5.15	Mutazioni neutrali si diffondono diventando fissate in una popolazione, con probabilità dipendente dalla dimensione della popolazione . . . . .	53
5.5.16	Molto può essere imparato dall'analisi della variazione tra gli umani . . . . .	53
<b>6</b>	<b>Replicazione, riparazione e ricombinazione del DNA</b>	<b>55</b>

# 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, escludendo 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 ortologhi 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 alghe 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

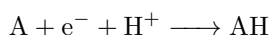
Le cellule devono ridurre il proprio livello di entropia e pertanto deve recuperare energia dall'ambiente sotto forma di cibo o fotoni, che viene poi utilizzata per generare l'ordine necessario. Nel corso di queste reazioni una parte dell'energia utilizzata viene trasformata in calore. L'energia, nel 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:

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

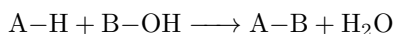
Dove  $R$  è la costante dei gas reali e  $T$  la temperatura assoluta. Si noti come al procedere della reazione il rapporto tra i reagenti cambia e avvicina  $\Delta G$  a zero, dove si raggiunge l'equilibrio chimico e non esiste un cambio di energia libera per guidare la reazione in nessuna direzione 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

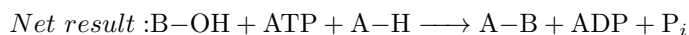
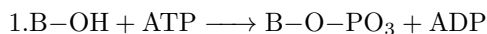
L'energia rilasciata dall'ossidazione delle molecole deve essere temporaneamente conservata prima che possa essere canalizzata nella sintesi. Nella maggior parte dei casi questo avviene come energia chimica di legame in un sottoinsieme di molecole dette molecole vettore con dei legami covalenti ricchi 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 accoppiati 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, permando 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 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 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

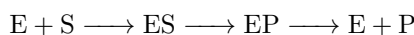
Le molecole che collidono con superfici che si accoppiano male formano pochi legami deboli e si disassociano rapidamente mentre se ne formano troppi l'associazione può persistere nel tempo. Associazioni forti accadono quando una funzione biologica richiede che le molecole rimangano associate per molto 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 segregare 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 ligande 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'aggiunta 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 aggiunta 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.

## 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 dettato da 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 se 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 sintenici 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 e 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 prole 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 eterozigotità nella popolazione. Un sottoinsieme di questi mutazioni sono responsabili per gli aspetti ereditabili degli individui.

## Capitolo 6

# Replicazione, riparazione e ricombinazione del DNA