



UNIVERSITÀ DI PISA

Corso di Laurea Triennale in Informatica (L-31)

TESI DI LAUREA

**Protein Folding: dai metodi classici alla
rivoluzione di AlphaFold**

Relatore

Prof. Paolo Milazzo

Candidato

Ludovico Venturi

ANNO ACCADEMICO 2020/2021

Indice

1	Introduzione	2
2	Background	3
2.1	Background biologico	4
2.1.1	Organizzazione della vita: dagli atomi alle cellule	4
2.1.2	Dogma centrale della biologia	6
2.1.3	Proteine: le macromolecole più importanti della vita	6
2.2	Background informatico	7
3	Protein Folding	8
4	Predizione della struttura di proteine	9
5	AlphaFold	10
6	Uso di AlphaFold e visualizzazione	11
7	Scenari aperti e conclusioni	12

Capitolo 1

Introduzione

Capitolo 2

Background

Cos'è la vita? Da dove viene? - Fino al 18° secolo per rispondere a tale quesito si faceva riferimento alla fede nel vitalismo: l'esistenza di una forza vitale non subordinata a leggi della chimica e della fisica. Il cambiamento avvenne nel 19° secolo. Un'importante svolta fu il lavoro di Louis Pasteur che stabilì un collegamento fra processi vitali e reazioni chimiche: la conversione di zucchero in alcool (fermentazione) era un risultato della crescita di microorganismi.

Successivamente vi sono i lavori di Berthelot e Buchner (premio Nobel in Chimica 1907) il quale dimostrò che era possibile ottenere la fermentazione in assenza di microorganismi, usando solamente sostanze estratte da essi. Queste sostanze furono chiamate *enzimi* (dal ted. Enzym, letteralmente «dentro il lievito»[1]). Non si conosceva la loro natura chimica, si scoprì successivamente che tutti gli enzimi sono *proteine* (dal greco «primario», «che occupa la prima posizione»). Queste proteine agivano da catalizzatori: acceleravano le reazioni chimiche all'interno delle cellule e nei tessuti senza cambiare la loro natura, quindi senza consumarsi, e senza entrare nei prodotti finali della reazione.

La scoperta degli enzimi portò ad un cambio di paradigma nel pensiero scientifico riguardo le origini della vita: veniva ora considerata come la conseguenza di numerosi processi chimici resi possibili dalle proteine[2]. I fondamenti del pensiero biologico si spostarono dal vitalismo al meccanicismo secondo il quale tutti i fenomeni naturali, vita compresa, sono governati dalle stesse leggi, sia per sostanze organiche che inorganiche.

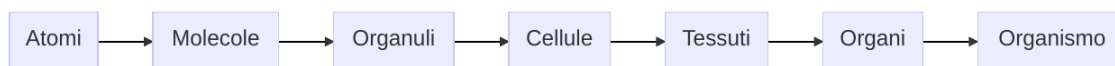
L'inconorazione delle proteine a *macromolecole più importanti della vita* si può legare ad un'altra svolta nel pensiero scientifico avvenuta nella seconda metà del 20° secolo: la rivoluzione genetica. Le proteine sono ben più che "macchine molecolari": sono i prodotti primari dei geni, responsabili, fra altri, dell'espressione dell'informazione genetica. È sullo sfondo di questa rivoluzione che l'informatica si è inserita all'interno del mondo della biologia.

2.1 Background biologico

2.1.1 Organizzazione della vita: dagli atomi alle cellule

Nonostante le grandi differenze in dimensione, dieta, riproduzione, morfologia, comportamento, vi è un tratto comune a tutti gli organismi viventi: sono composti di cellule. Tutte le cellule sono caratterizzate da una stupefacente somiglianza chimica poiché utilizzano molecole simili e hanno ereditato tutte le stesse intuizioni genetiche. Si pensa quindi vi sia un antenato comune a tutti i viventi: una cellula vissuta circa 3,5 miliardi di anni fa che conteneva un prototipo del macchinario universale della vita sulla Terra oggi[3].

Prima di parlare di cellule è opportuno richiamare l'attenzione sulle strutture biologiche. L'organizzazione biologica si basa su una gerarchia di livelli strutturali, ognuno dei quali poggia su un gradino sottostante:



Tutta la materia è costituita da 94 elementi chimici in natura (tralasciando quelli non stabili). La materia organica è composta per il 96% da atomi di C, O, N, H (carbonio, ossigeno, azoto, idrogeno). Un atomo ha un nucleo composto da neutroni e protoni circondato da una nube di elettroni in rapido movimento. Il Dalton (Da) è l'unità della massa atomica, corrisponde al peso di un protone o neutrone: $1Da = 1.7 \times 10^{-24}g$. Un elettrone pesa $0.0005Da$. Gli elettroni più esterni sono chiamati *elettroni di valenza* e determinano il comportamento chimico di un atomo.

Lo scheletro dei composti organici è formato da catene carboniose, lunghe catene di atomi di carbonio legati fra loro da legami covalenti (il tipo di legame chimico più forte). Salendo di un livello nella gerarchia strutturale si arriva alle macromolecole biologiche, fondamentali per le cellule: carboidrati, lipidi, acidi nucleici e proteine. I carboidrati sono combustibili cellulari e materiale da costruzione, i lipidi sono sia depositi di energia che gusci protettivi, gli acidi nucleici permettono di codificare l'informazione genica e le proteine sono alla base delle funzioni vitali.

La cellula è la più piccola unità in grado di vivere. Per *vivente* si intende un essere dotato di: organizzazione interna, metabolismo, omeostasi, interazione con l'ambiente, adattamento, crescita e riproduzione.

Le cellule hanno dimensioni che variano dai $2\mu m$ ai cm delle uova di rana, gallina o struzzo ai m di neuroni con lunghi assoni:

È possibile dividere gli esseri viventi in due domini: *procarioti* ed *eucarioti*. Il primo include i due regni Bacteria e Archaea. Sono caratterizzati da cellule piccole, circa

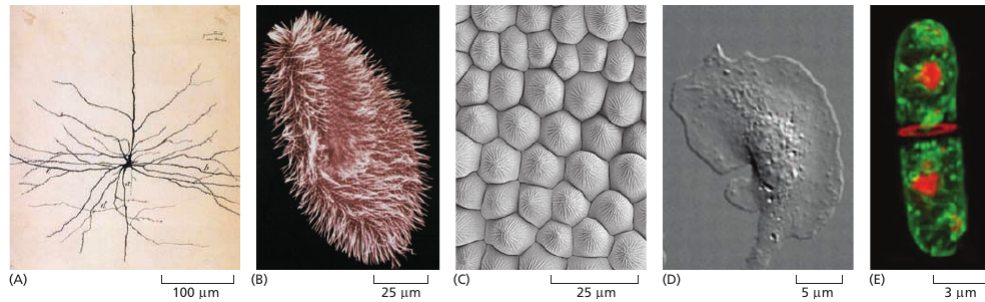
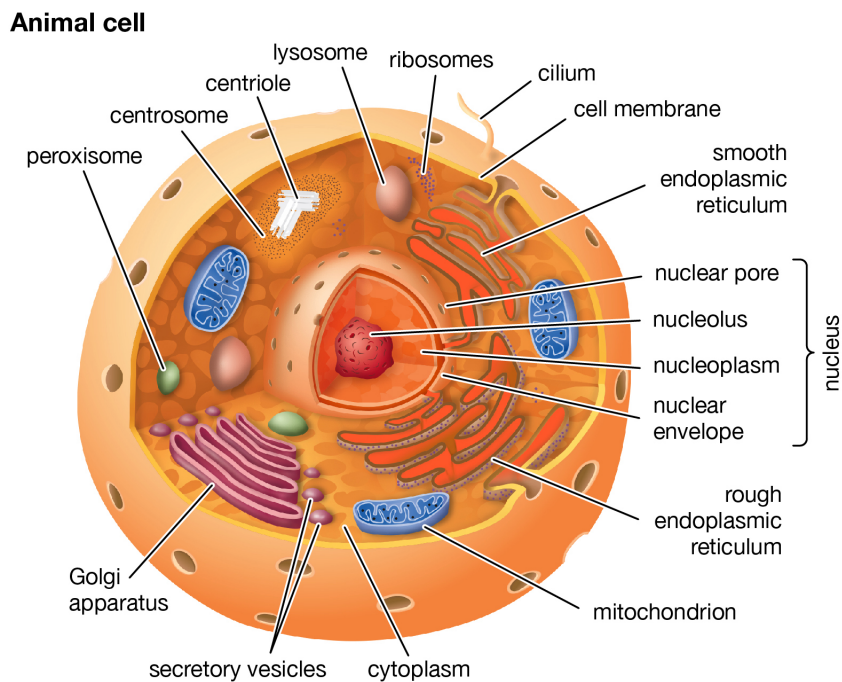


Figura 2.1: (A) disegno di un neurone. (B) Paramecium. (C) superficie di un petalo di fiore di bocca di leone. (D) Macrofago. (E) Un lievito di fissione viene catturato nell'atto di divisione cellulare

$1\mu m$. Il secondo dominio include cinque regni: animali, piante, funghi, protisti e cromisti. Dispongono di cellule più grandi (circa $10-100\mu m$) dotate di compartimenti interni che dividono i processi cellulari.

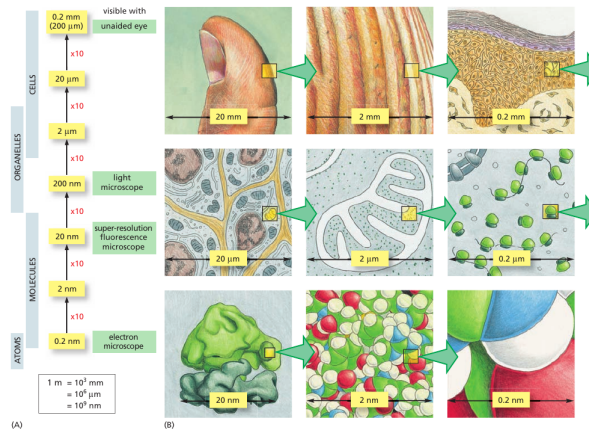
La struttura tipica di una cellula animale è mostrata nella seguente figura:



© Encyclopædia Britannica, Inc.

L'interno delle cellule è composto da ambiente acquoso per il 70-95%.

È importante notare che passando da un livello al superiore si palesano nuove proprietà non riconducibili ai livelli più semplici: le proprietà *emergenti*. Una singola molecola d'acqua non è né solida né liquida.



Un altro concetto fondamentale in biologia è la relazione forma-funzione, valida a tutti i livelli biologici. Se le ali degli uccelli non fossero così come sono essi non potrebbero volare, se i mitocondri non avessero striature non potrebbero svolgere la respirazione cellulare, se i neuroni non avessero lunghi assoni non riuscirebbero a comunicare oppure si pensi al *paramecium* che si muove come un sommergibile grazie alle sue ciglia (vedi figura 2.1B).

L'evoluzione rappresenta il tema centrale ed unificante della biologia, come si è già accennato sopra. Gli organismi sono sistemi aperti che interagiscono continuamente con l'ambiente, dotati di variabilità individuale e finalizzati alla competizione per la sopravvivenza.

Vi sono da 5 a 30 milioni di specie differenti eppure scendendo sempre di più nella struttura degli organismi si osserva una similitudine quasi sconcertante. Un esempio che ci riguarda è la somiglianza fra le ciglia di *paramecium* e le ciglia di una cellula epiteliale delle vie aeree degli esseri umani: presentano la stessa sezione trasversale. Diversità e unità della vita sulla Terra sono due facce della stessa medaglia.

2.1.2 Dogma centrale della biologia

- DNA, fenotipo e genotipo • RNA
- codoni, amminoacidi

2.1.3 Proteine: le macromolecole più importanti della vita

Oltre agli enzimi ci sono altre proteine importanti, uno degli esempi più noti è l'emoglobina, proteina animale adibita a trasportare ossigeno dai polmoni agli organi e ai tessuti del corpo così come a riportare CO₂ ai polmoni.

Importante funzione degli enzimi è correlata alla digestione negli animali. Enzimi come le amilasi e le proteasi sono in grado di ridurre le macromolecole (nella fattispecie amido e proteine) in unità semplici (maltosio e amminoacidi), assorbibili dall'intestino

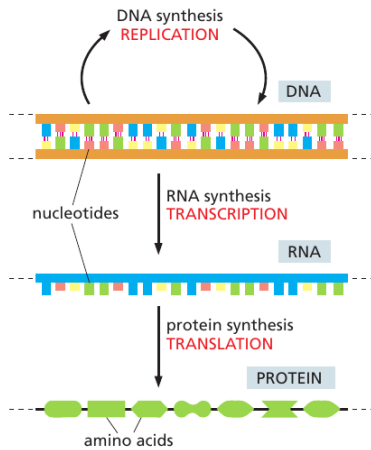


Figura 2.2: Dogma centrale in biologia

Tutti gli enzimi sono proteine, ma non tutti i catalizzatori biologici sono enzimi, dal momento che esistono anche catalizzatori costituiti di RNA, chiamati ribozimi

In un organismo, nonostante tutte le cellule condividano gli stessi geni, cellule afferenti a organi o tessuti diversi esprimono geni differenti (*espressione genica*).

2.2 Background informatico

Background informatico • bioinformatica • database bioinformatici • machine learning
• reti neurali, deep learning

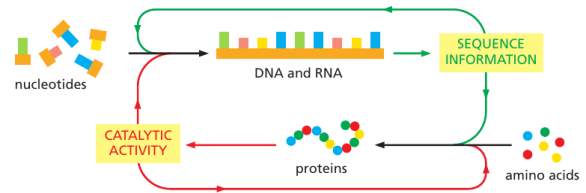


Figura 2.3: La vita è un processo autocatalitico

Capitolo 3

Protein Folding

Capitolo 4

Predizione della struttura di proteine

Capitolo 5

AlphaFold

Capitolo 6

Uso di AlphaFold e visualizzazione

Capitolo 7

Scenari aperti e conclusioni

Bibliografia

- [1] “enzima nell’Enciclopedia Treccani.” [Online; accessed 21. Jan. 2022]. (), indirizzo: <https://www.treccani.it/enciclopedia/enzima>.
- [2] A. Kessel e N. Ben-Tal, *Introduction to proteins: Structure, function and motion*, 2^a ed. Chapman e Hall/CRC, 2018.
- [3] B. Alberts, D. Bray, K. Hopkin et al., *Essential cell biology*, 5^a ed. W. W. Norton e Company, 2019.