

2) BACKGROUND

Background biologico

- organizzazione molecolare dei viventi
- come funziona una cellula
- DNA, fenotipo e genotipo
- RNA
- Dogma centrale della biologia
- codoni, amminoacidi
- Proteine, le macromolecole più importanti della vita

Background informatico

- bioinformatica
- database bioinformatici
- machine learning
- reti neurali, deep learning

3) IL PROBLEMA DEL PROTEIN FOLDING

- cos'è il problema del protein folding: non solo la struttura finale
- 3 sottoproblemi: folding code, protein structure prediction e folding process
- le domande di principio del protein folding
- struttura delle proteine (4 livelli)
- Domini, Residui, Motivi, Giri
- postulato di Anfinsen, paradosso di Levinthal
- interazioni (covalenti, polari, van der Waals, ..)
- una forza dominante o tante piccole forze?
- Cambio di paradigma da metà anni '80
- chaperonine, sintesi
- misfolding (denaturazione)
- struttura e funzione, malattie e prioni
- Limiti al ripiegamento: angoli di torsione e piano di Ramachandran
- processo spontaneo: energia di Gibbs, entalpia, entropia

4) PREDIZIONE DELLA STRUTTURA DI PROTEINE

COME VENGONO DETERMINATE SPERIMENTALMENTE LE STRUTTURE

- cristallografia a raggi-x
- NMR, risonanza magnetica nucleare
- Cryo-EM (electron microscopy)

STRUMENTI E METODI COMPUTAZIONALI

- utilità: farmaci e funzione dalla sequenza genica
- come sono descritte le strutture delle proteine (es. PDB format)
- rappresentazione DNA e sequenze amminoacidiche
- PDB, Protein Databank
- CASP

- Workflow oggi (2015)
- proprietà dati input
- output: modelli e informazioni
- metriche di valutazione e verifica
 - IDDT, GDT...

- **Paradigmi (approssimazione biologica)**
 - ab initio
 - spazio configurazionale
 - MD: molecular dynamics
 - complessità e utilità
 - template-based (homology-based)
 - 5 step principali
 - threading
- metodi integrativi

- **Paradigmi (soft computing)**
 - ANN
 - EC
 - SVM
 - altri approcci: statistici, matematici, correlated mutation, ...

- **ML e task biologici**
 - MSA

- attention, self-attention, transformer
- equivarianza
- un po' di storia:
 - primi passi: predizione struttura secondaria
 - anni '90: database, omologia e sequence alignment

5) ALPHAFOLD

- rivoluzione
- cosa consente di fare
- CASP 13, AlphaFold1
- CASP 14, AlphaFold2 e rivoluzione
- AlphaFold DB
- limiti, intrinsically disordered protein regions

PRINCIPI DI FUNZIONAMENTO

- distogramma (contact map)
- dataset bioinformatici usati
- architettura di AlphaFold
 - input
 - database usati
 - rappresentazioni intermedie
 - evoformer
 - cosa è cambiato da AlphaFold 1
 - Invariant Point Attention (IPA) modulo
 - output: struttura, stime di confidenza (pLDDT, PAE)
- DeepMind, Alphabet Inc, Isomorphic Labs

6) USO DI ALPHAFOLD E VISUALIZZAZIONE

- github
- colab
- esempio d'uso nel server unipi
- visualizzazione
- differenti rappresentazioni:
 - space-fill, wire-frame, ribbon, surface
 - colorazione: evoluzione, potenziale elettrostatico
 - combinata

7) SCENARI APERTI E CONCLUSIONI

- impatto di AlphaFold e AlphaFold DB sulle scienze della vita
 - box etico-morale: si sta andando in una direzione troppo data-centric?
- AlphaFold e covid-19
- drug engineering
- il problema del protein folding non è risolto