

# Esercitazione di Bioinformatica





**Expasy**  
Swiss Bioinformatics Resource Portal

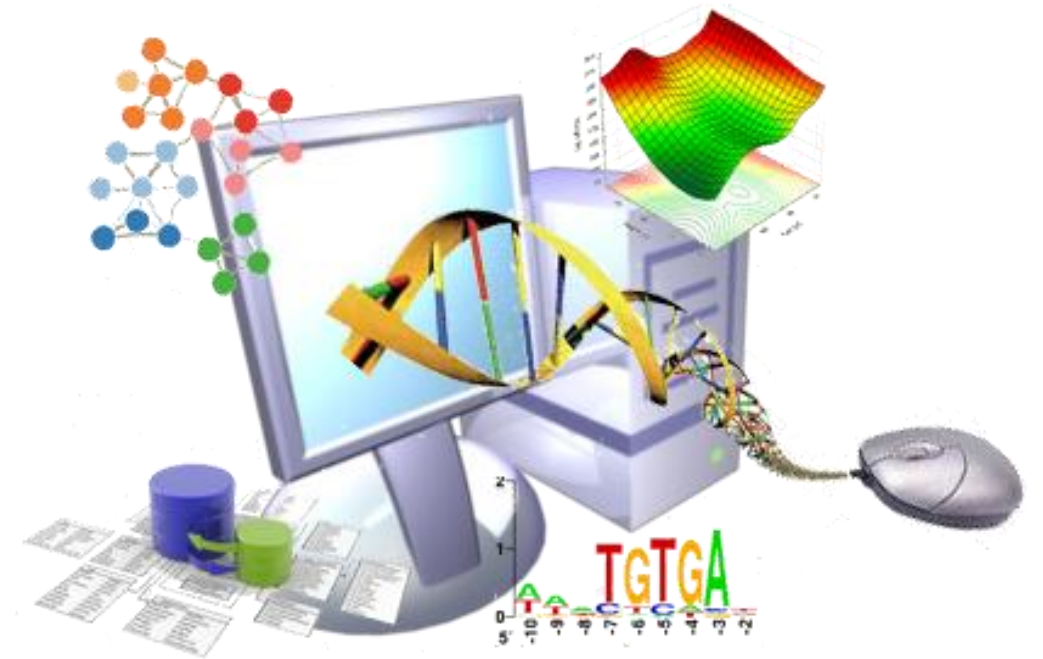


**iCn3D** *Explore 3D structures*

---

La **bioinformatica** è una disciplina interdisciplinare che utilizza informatica, statistica e matematica per analizzare e interpretare grandi quantità di dati biologici (genomica, proteomica, trascrittomica).

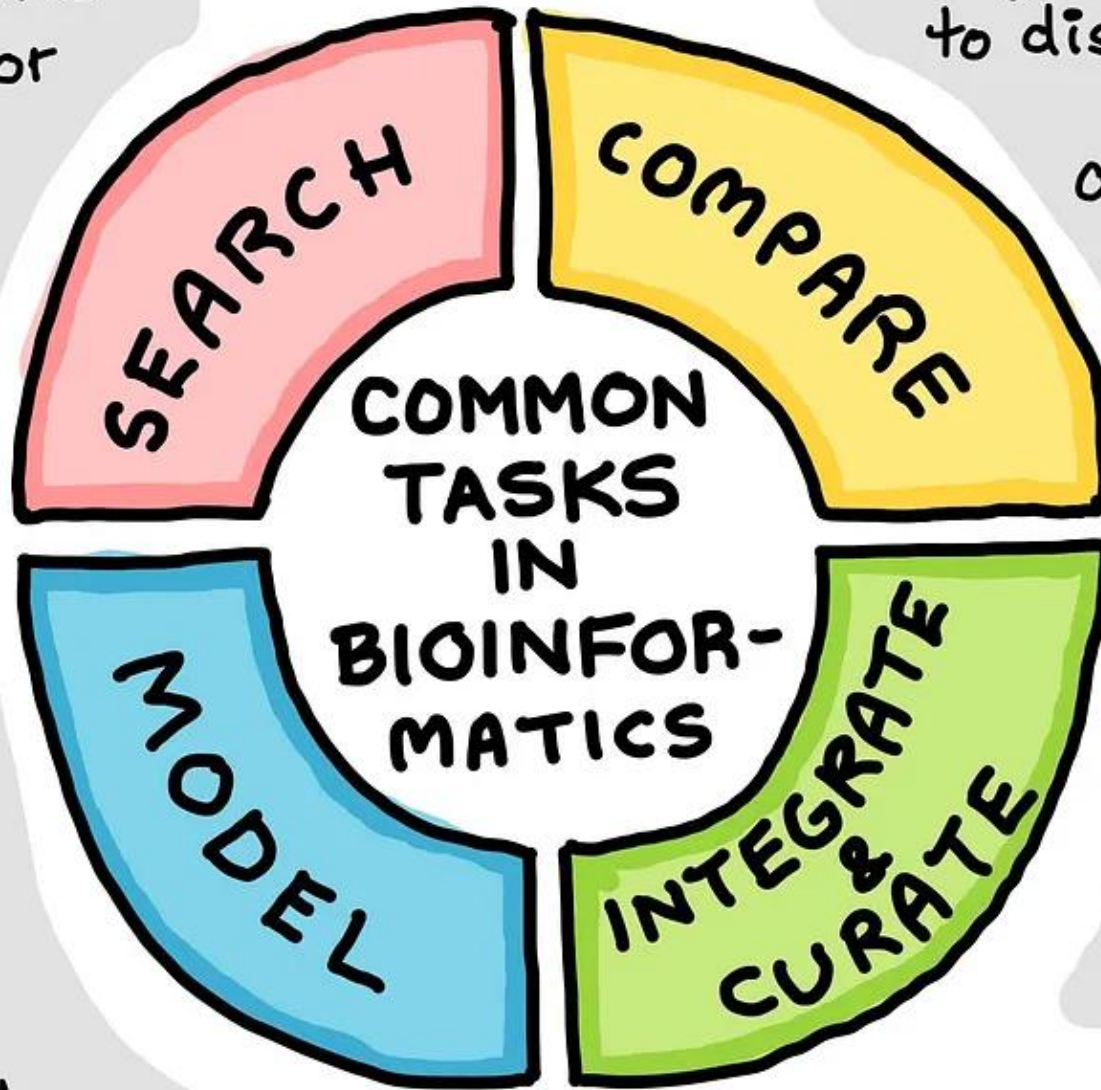
È **fondamentale** per sviluppare algoritmi, software e database che permettono di studiare sequenze di DNA, strutture proteiche e processi molecolari, accelerando la ricerca biomedica e farmaceutica.



**Search** public databases for information on genes / proteins / RNA / pathways

Build **structural model** of protein structure

Build **predictive model** using retrospective data



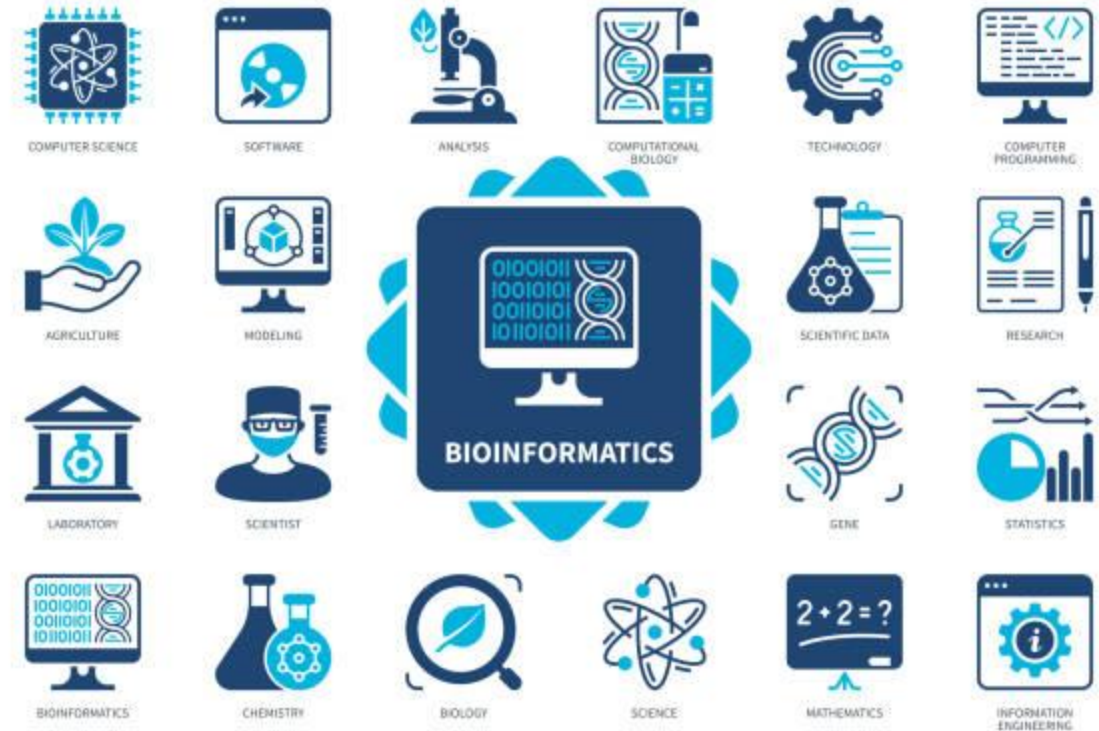
Sequence alignment to discern **similarity / differences** amongst various genes / proteins / RNA

**Combine** heterogeneous data sources & **Remove** redundancy

La bioinformatica è il ponte perfetto per far capire ai ragazzi che la biologia moderna non si fa solo in "**wet lab**" (con le provette), ma anche in "**dry lab**" (al computer).

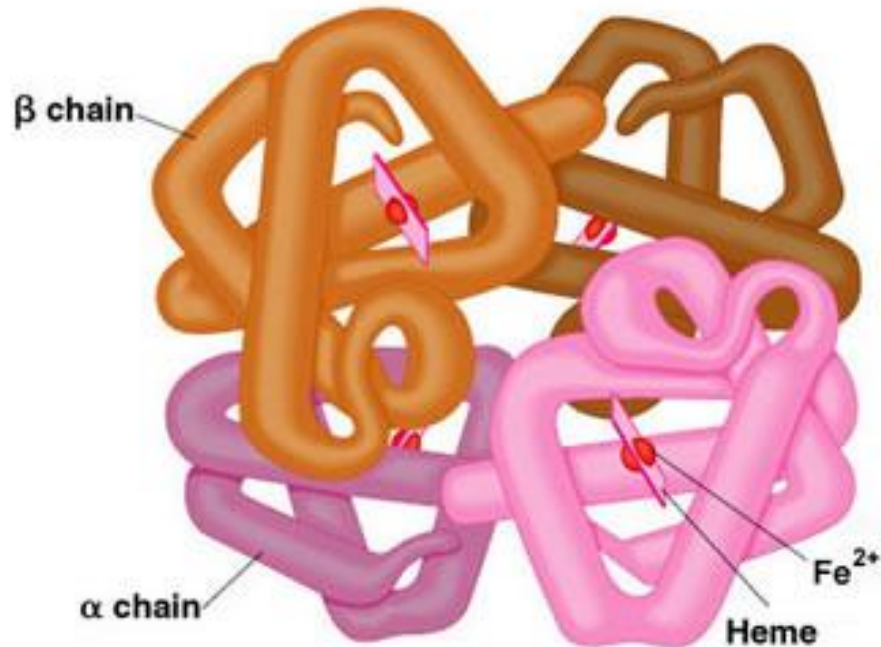


*Prof. Marco Rinaldo Oggioni  
is an expert in bacterial  
genetics and pathogenesis  
of infection.*

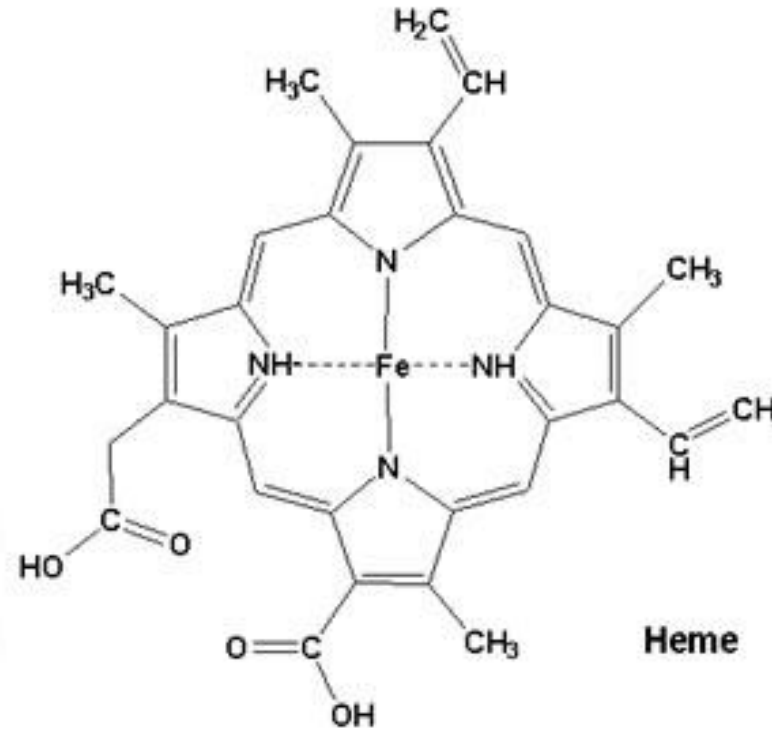


# Studio di caso: l'anemia falciforme

L'**emoglobina** (Hb o Hgb) è una proteina globulare ricca di ferro contenuta nei globuli rossi, responsabile del trasporto dell'ossigeno dai polmoni ai tessuti e del colore rosso del sangue.

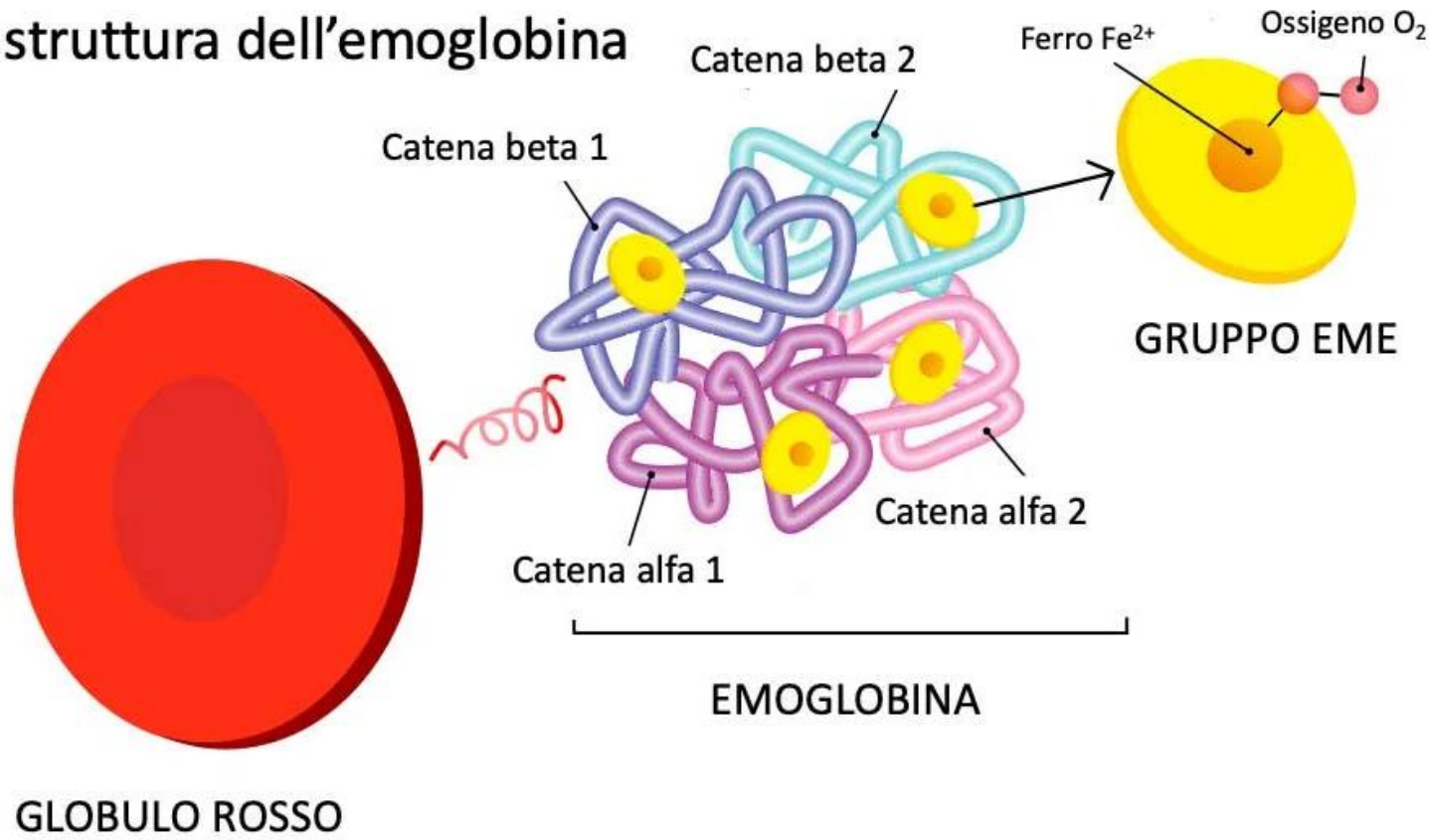


Struttura molecolare emoglobina



Gruppo EME

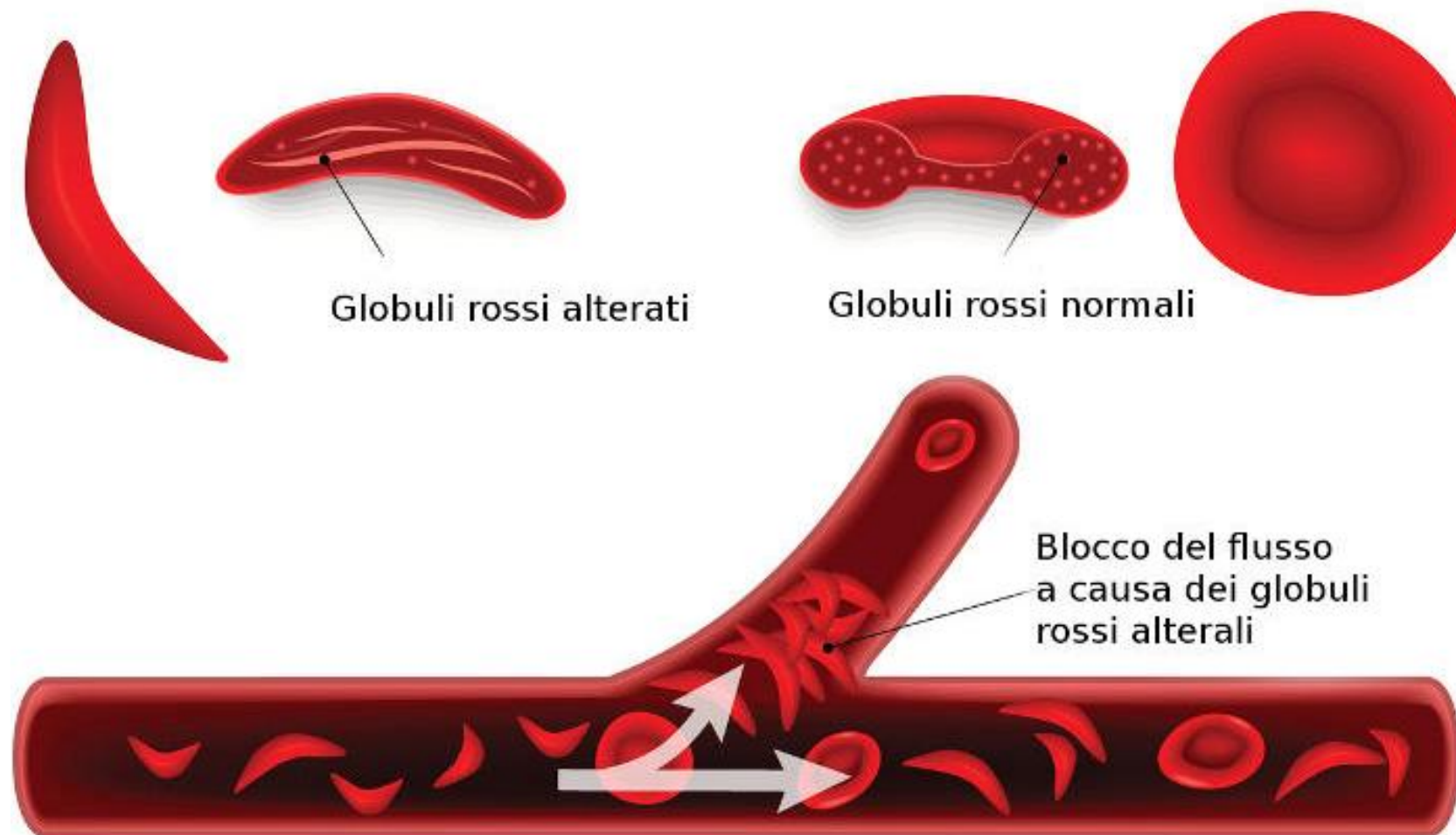
# La struttura dell'emoglobina



**Catena  $\alpha$** : è un polipeptide composto da **141 amminoacidi**. I geni che codificano per la catena alfa si trovano sul **cromosoma 16**.

**Catena  $\beta$** : è un polipeptide costituito da **146 amminoacidi**. I geni per la catena beta sono situati sul **cromosoma 11**.

L'**anemia falciforme**, o **drepanocitosi**, è una grave malattia genetica ereditaria del sangue, caratterizzata dalla produzione di emoglobina anomala (**HbS**) che deforma i globuli rossi a forma di falce. Queste cellule rigide causano ostruzioni vascolari, provocando crisi dolorose, anemia cronica e danni d'organo.



DNA dell'emoglobina normale



mRNA



Emoglobina normale



DNA dell'emoglobina mutante



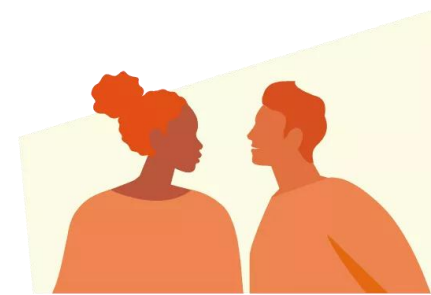
mRNA



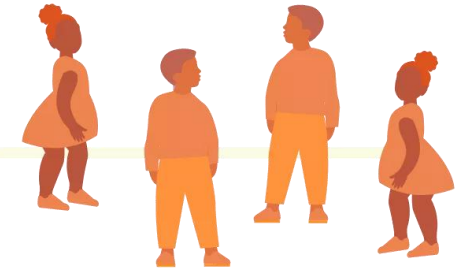
Emoglobina falciforme



L'anemia falciforme è una malattia genetica ed ereditaria, trasmessa con **modalità autosomica recessiva**.



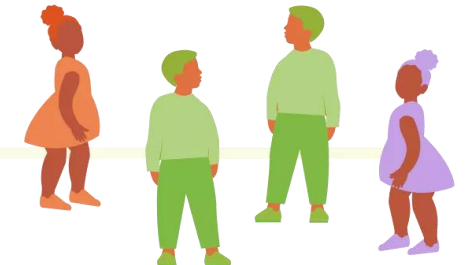
Entrambi i genitori hanno l'anemia falciforme



Tutti i bambini avranno l'anemia falciforme



Entrambi i genitori hanno il tratto falciforme



1 bambino su 4 avrà l'anemia falciforme, e 2 su 4 avranno il tratto falciforme

 Anemia falciforme    Tratto falciforme    Emoglobina normale

# 1. Ricerca della sequenza su NCBI



L'obiettivo è estrarre il "codice sorgente" della vita.

1. **Accesso**: vai su [NCBI](#).
2. **Ricerca**: nel menu a tendina seleziona **Gene** e digita **HBB human** (HBB è il simbolo del gene della beta-emoglobina).
3. **Selezione**: clicca sul primo risultato (HBB hemoglobin subunit beta [Homo sapiens]).
4. **Download**: scorri in basso fino alla sezione "Genomic regions, transcripts, and products".
5. **Clicca** su FASTA.

La sequenza inizia con >. Il formato FASTA è un semplice file di testo che i computer leggono velocemente.

6. Copiate la sequenza e incollatela in un file Word o Blocco Note.



## 2. Confronto con BLAST (allineamento)



Detective molecolari all'opera... per trovare le differenze tra specie o mutazioni.

**1. Accesso:** vai su BLASTn (Nucleotide BLAST).

**2. Input:** incolla la sequenza FASTA ottenuta precedentemente nel box "Enter Query Sequence".

**3. Confronto:** in "Choose Search Set", seleziona "Standard databases".

4. In "**Program Selection**" scegli "Highly similar sequences (megablast)".

**5. Analisi dei risultati:** apparirà una lista di geni simili. Notate la *Percent Identity*.



Ora dovete ottenere la sequenza genica della catena  $\beta$  del Pan troglodytes.

Potete cercarla di nuovo su **NCBI Gene** digitando **HBB Pan troglodytes**, oppure (per velocizzare) con il codice di accesso: **NM\_001009111**.

An official website of the United States government [Here's how you know](#) ▼

**NIH** National Library of Medicine  
National Center for Biotechnology Information

Gene  ▼ HBB Pan troglodytes  
[Create RSS](#) [Save search](#) [Advanced](#)

Gene sources [Tabular](#) ▼ [20 per page](#) ▼ [Sort by Relevance](#) ▼ [Send to:](#) ▼  
Genomic

Categories  
[Alternatively spliced](#)  
[Annotated genes](#)  
[Protein-coding](#)

Sequence content  
[Ensembl](#)  
[RefSeq](#)

Status [clear](#)  
 **Current**

GENE [Was this helpful?](#)

[HBB – hemoglobin subunit beta](#)  
[Pan troglodytes \(chimpanzee\)](#)  
Also known as: GLNA1  
Gene ID: 450978

## 3. Configurazione di BLAST

1. Vai sulla pagina principale di [BLASTn](#).
2. **Sequenza Query**: incolla la sequenza umana nel box grande in alto.
3. Il comando chiave: spunta la casella "**Align two or more sequences**" (si trova subito sotto il primo box). Apparirà un secondo box di inserimento.
4. **Sequenza Subject**: incolla nel secondo box la sequenza dello scimpanzé.
5. Esecuzione: scorri in basso e clicca sul tasto blu BLAST.

Il sistema impiegherà qualche secondo per "**sovrapporre**" le due stringhe di lettere e cercare le differenze.

# Come leggere i risultati (analisi dei dati)

Una volta terminata l'elaborazione, gli studenti devono concentrarsi sulla scheda "**Alignments**": Identità (*Identity*).

Vedrete una percentuale altissima (solitamente superiore al 99%). Questo significa che su centinaia di basi azotate, solo pochissime sono diverse.

Il "Match" visivo: BLAST mostra due righe di lettere. Tra le due righe ci sono dei trattini verticali **|**.

- Se c'è il trattino: le basi sono identiche (es. A con A).
- Se manca il trattino: c'è stata una mutazione in quel punto.
- Gap: Se vedete un trattino orizzontale **-** in una delle due sequenze, significa che lì si è verificata una delezione o un'inserzione (Indel).

Score	Expect	Identities	Gaps	Strand
2839 bits(1537)	0.0	1588/1612(99%)	6/1612(0%)	Plus/Plus
Query 1		ACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATC		60
Sbjct 12		ACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACC		71
Query 61		TGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAA		120
Sbjct 72		TGACTCCTGAGGAGAAGTCTGCCGTTACTG-CCCTGTGGGGCAAGGTGAACGTGGATGAA		130
Query 121		GTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACC		180
Sbjct 131		GTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGCTTAAGGAGACC		190
Query 181		AATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTC		240
Sbjct 191		AGTAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTC		250



# Sfide

1. Contate quante basi sono diverse tra noi e lo scimpanzé.
2. Queste differenze cambiano la proteina?
3. Perché la percentuale non è il 100%?
4. E se provassimo a fare lo stesso esperimento confrontando l'uomo con un topo (*Mus musculus*), quale percentuale di identità mi aspetto di trovare? Maggiore? Minore? Uguale?

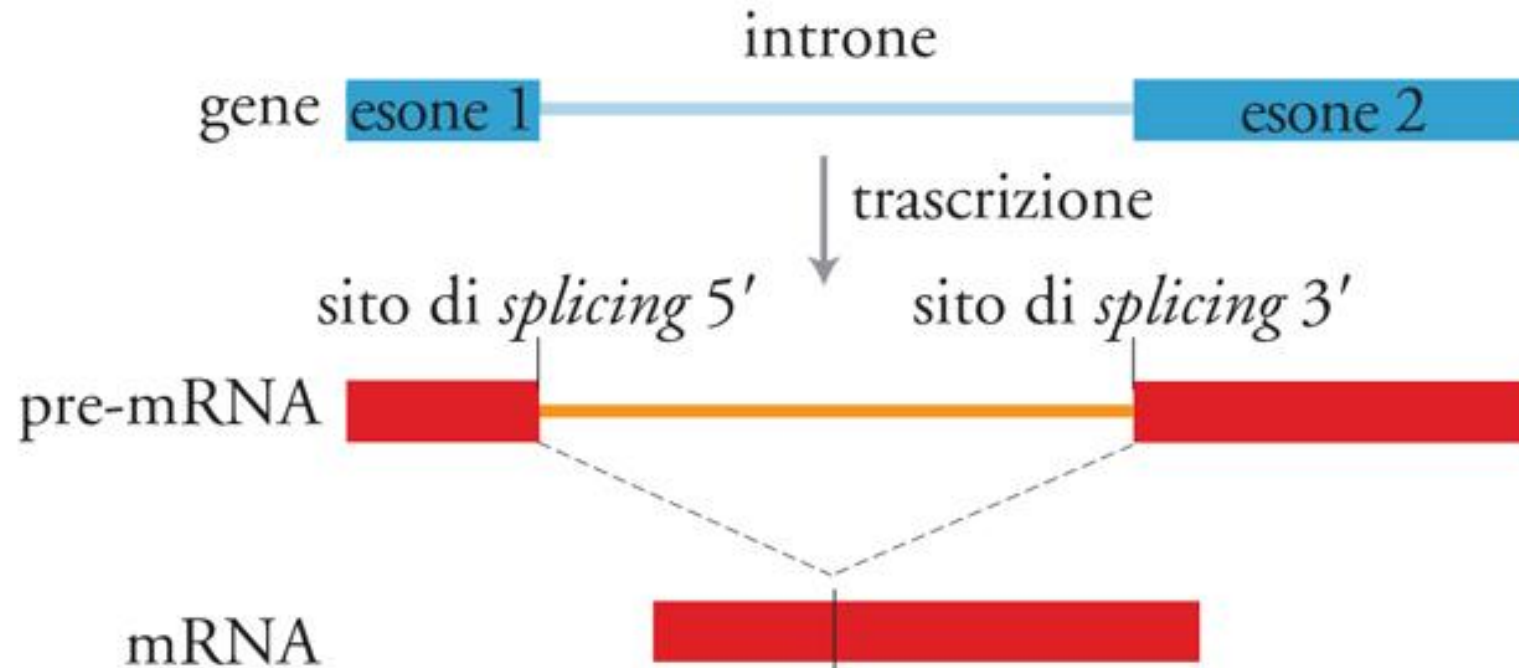


## 4. Traduzione con ExPASy Translate Tool

In biologia molecolare, questo passaggio rappresenta la **traduzione** (o sintesi proteica) che avviene nei ribosomi.

ExPASy

Swiss Bioinformatics Resource Portal



part\_of hemoglobin complex  
located\_in tertiary granule lumen

TAS PubMed  
TAS

### General protein information

Preferred Names  
hemoglobin subunit beta

Names  
Hb Monza protein  
beta globin chain  
hemoglobin beta subunit  
hemoglobin, beta

### NCBI Reference Sequences (RefSeq)

**NEW** Try the new [Transcript table](#)

☐ [RefSeqs maintained independently of Annotated Genomes](#)

These reference sequences exist independently of genome builds. [Explain](#)

#### Genomic

1. NG_059281.1 RefSeqGene	
Range	5001..6608
Download	<a href="#">GenBank</a> , <a href="#">FASTA</a> , <a href="#">Sequence Viewer (Graphics)</a> , <a href="#">LRG_1232</a>

Questo è il passaggio **più critico** perché spesso la pagina del Gene di NCBI è molto densa di informazioni e ci si può perdere!



Ecco la procedura guidata per trovare la **sequenza "pulita" di mRNA** (identificata dalla sigla **NM\_**), che è quella che contiene solo gli **esoni** già montati e pronti per la traduzione:

1. Partire dalla pagina "**Gene**": una volta cercato HBB human su NCBI Gene e selezionato il primo risultato, vi troverete in una pagina molto lunga.
2. Individuare la sezione "NCBI Reference Sequences (RefSeq)": **non fermarti alla prima sequenza che vedi in alto**. Scorri la pagina verso il basso (**circa a metà**) finché non trovi una sezione chiamata: "**NCBI Reference Sequences (RefSeq)**"

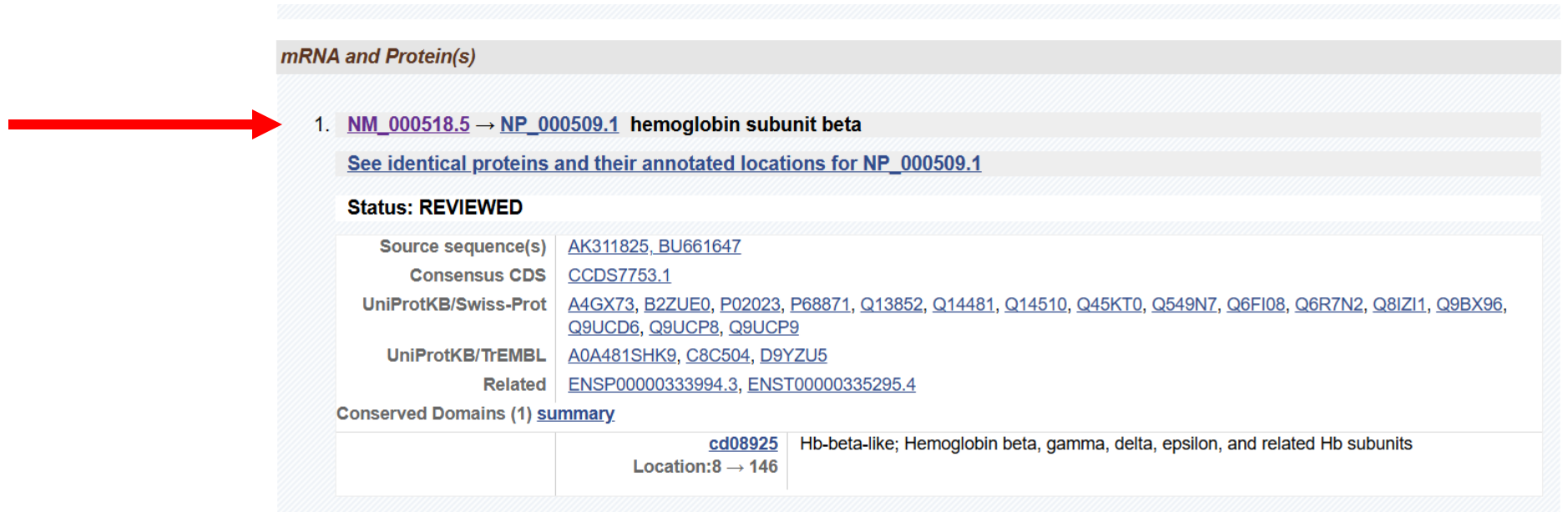
## Cercare la riga "mRNA and Protein(s)"

Sotto questa intestazione vedrai una lista di codici.  
Quello che serve a noi è nella colonna di sinistra.

Cerca un codice che inizi con **NM\_**. Per l'emoglobina beta umana, il codice standard è **NM\_000518.5**.

**Nota bene:** Accanto a "NM\_..." vedrai spesso un link che inizia con **NP\_**....

*Il codice NP si riferisce alla proteina già tradotta; a noi serve il codice NM (mRNA) perché vogliamo che siate voi studenti a fare il lavoro di traduzione.*



**mRNA and Protein(s)**

1. [NM\\_000518.5](#) → [NP\\_000509.1](#) hemoglobin subunit beta

[See identical proteins and their annotated locations for NP\\_000509.1](#)

Status: REVIEWED

Source sequence(s)	<a href="#">AK311825</a> , <a href="#">BU661647</a>
Consensus CDS	<a href="#">CCDS7753.1</a>
UniProtKB/Swiss-Prot	<a href="#">A4GX73</a> , <a href="#">B2ZUE0</a> , <a href="#">P02023</a> , <a href="#">P68871</a> , <a href="#">Q13852</a> , <a href="#">Q14481</a> , <a href="#">Q14510</a> , <a href="#">Q45KT0</a> , <a href="#">Q549N7</a> , <a href="#">Q6FI08</a> , <a href="#">Q6R7N2</a> , <a href="#">Q8IZI1</a> , <a href="#">Q9BX96</a> , <a href="#">Q9UCD6</a> , <a href="#">Q9UCP8</a> , <a href="#">Q9UCP9</a>
UniProtKB/TrEMBL	<a href="#">A0A481SHK9</a> , <a href="#">C8C504</a> , <a href="#">D9YZU5</a>
Related	<a href="#">ENSP00000333994.3</a> , <a href="#">ENST00000335295.4</a>

Conserved Domains (1) [summary](#)

<a href="#">cd08925</a>	Hb-beta-like; Hemoglobin beta, gamma, delta, epsilon, and related Hb subunits
Location:8 → 146	

- Una volta cliccato sul link NM\_000518.5 si aprirà una nuova pagina (chiamata "Nucleotide").
- In alto a sinistra, subito sotto il titolo della sequenza, vedrai un piccolo link con scritto **FASTA**.
- **Cliccaci sopra.**
- La pagina si ricaricherà mostrando solo il "testo" genetico: una riga di intestazione che inizia con > seguita dai blocchi di lettere A, C, G, T.

FASTA ▾

## Homo sapiens hemoglobin subunit beta (HBB), mRNA

NCBI Reference Sequence: NM\_000518.5

[GenBank](#) [Graphics](#)

```
>NM_000518.5 Homo sapiens hemoglobin subunit beta (HBB), mRNA
ACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGA
GGAGAAGTCTGCCGTTACTGCCCTGTGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGC
AGGCTGCTGGTGGTCTACCCTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGCCACTCCTGATG
CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGC
TCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT
CCTGAGAAGTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCATCACTTTGGCAAAGAATTCA
CCCCACAGTGACAGGCTGCCTATCAGAAAGTGGTGGCTGGTGGCTAATGCCCTGGCCACAAAGTATCA
CTAAGCTCGCTTTCTTGTGTCCAATTTCTATTAAGGTTCCCTTTGTTCCCTAAGTCCAACACTAAACT
GGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTGCAA
```

An official website of the United States government [Here's how you know](#) ▾

**NIH** National Library of Medicine  
National Center for Biotechnology Information

Nucleotide

GenBank ▾ Send to: ▾

## Homo sapiens hemoglobin subunit beta (HBB), mRNA

NCBI Reference Sequence: NM\_000518.5

[FASTA](#) [Graphics](#)

---

Go to:

LOCUS	NM_000518	628 bp	mRNA	linear	PRI 01-FEB-2026
DEFINITION	Homo sapiens hemoglobin subunit beta (HBB), mRNA.				
ACCESSION	NM_000518				
VERSION	NM_000518.5				
KEYWORDS	RefSeq; MANE Select.				
SOURCE	Homo sapiens (human)				
ORGANISM	<a href="#">Homo sapiens</a>				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 628)				
AUTHORS	Civettini,I., Zappaterra,A., Corti,P., Messina,A., Aroldi,A., Biondi,A., Cavalca,F., Crippa,V., Crosti,F., Ferrari,G.M., Malighetti,F., Mologni,L., Piperno,A., Ramazzotti,D., Scollo,C., Zambon,A., Rossi,F. and Gambacorti-Passerini,C.				

# Passaggi operativi:

**Expasy Translate**  
Translate  
DNA/RNA to protein sequence

Expasy

- **Accesso:** Vai su [ExpASy Translate Tool](#).
- **Input:** incolla la sequenza di mRNA (quella che inizia con NM\_) ottenuta precedentemente.
- **Output:** clicca sul tasto Translate.
- Selezione del Reading Frame: il software mostrerà diversi "telai di lettura" (Frame).
- **Perché sono 6?**

5' AGGTGACACCGCAAGCCTTATATTAGC 3'

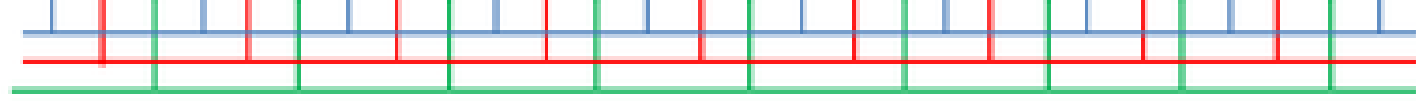


Illustration of possible reading frames:

**AGG·TGA·CAC·CGC·AAG·CCT·TAT·ATT·AGC**  
**A·GGT·GAC·ACC·GCA·AGC·CTT·ATA·TTA·GC**  
**AG·GTG·ACA·CCG·CAA·GCC·TTA·TAT·TAG·C**

Reading frame #1

5'-AGUCUUACCGCAUUGUGG-3'  
| | | | | |  
Ser--Leu--Thr--Ala--Leu-Ser

Reading frame #2

5'-AGUCUUACCGCAUUGUGG-3'  
| | | | | |  
Val--Leu--Pro--His--Cys

Reading frame #3

5'-AGUCUUACCGCAUUGUGG-3'  
| | | | | |  
Ser--Tyr--Arg--Ile--Val

Perché il DNA ha due filamenti (senso e antisenso) e il codice è a triplette (si può iniziare a leggere dalla base 1, 2 o 3).

### Quale scegliere?

Dovete cercare il blocco di testo che inizia con una M (Metionina, il codone di inizio START) e che prosegue per molto tempo **senza trattini** (i trattini - o la parola Stop indicano la fine della proteina). *Di solito è il 5'3' Frame 1.*



- Esercizio pratico: contate gli aminoacidi partendo dalla prima M (Metionina).
- Identificazione: il 6° aminoacido dopo la Metionina è una E (Acido Glutammico).
- Confronto "Malato": se i ragazzi avessero usato la sequenza del paziente con Anemia Falciforme, al 6° posto troverebbero una V (Valina).

## 5. Visualizzazione 3D con iCn3D

Ora trasformiamo quelle lettere in una struttura reale che occupa spazio.



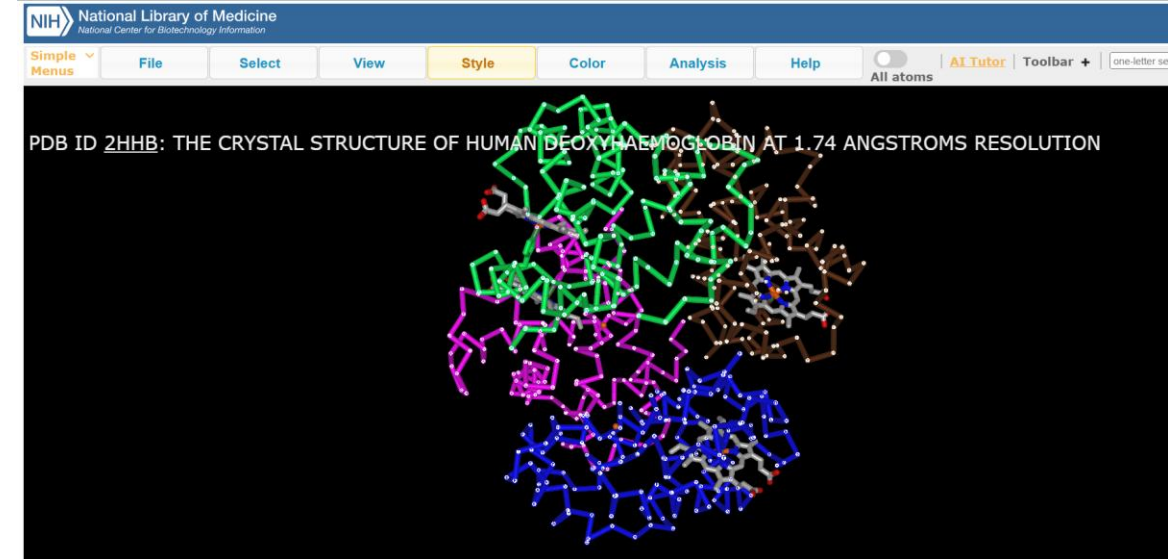
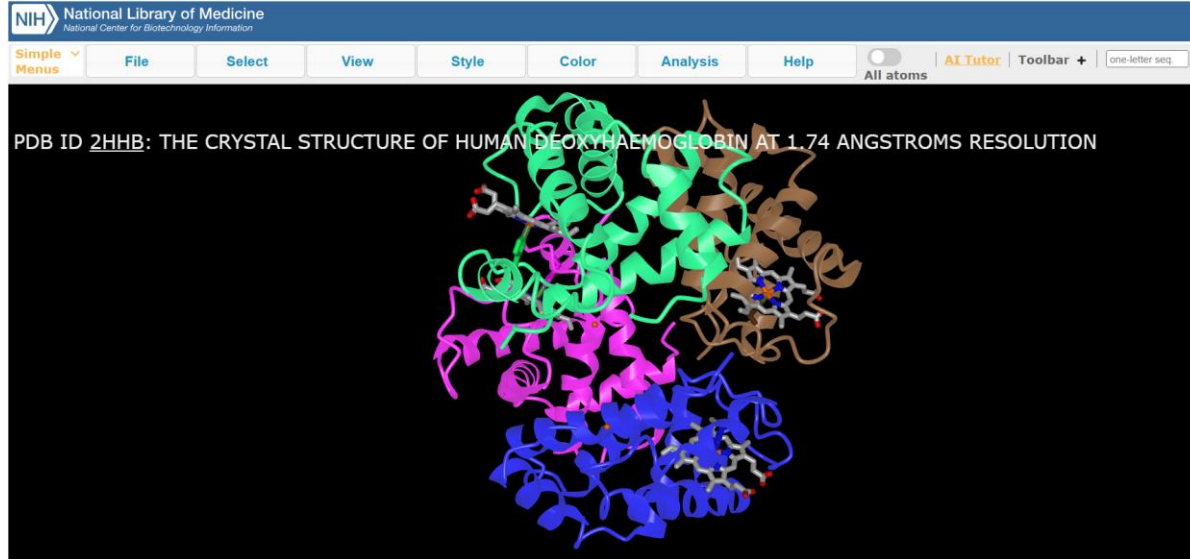
**iCn3D** *Explore 3D structures*

- 1. Apri il sito:** [NCBI iCn3D](https://www.ncbi.nlm.nih.gov/Structure/iCn3D/).
- 2. Vai nel menu in alto:** File > Retrieve by ID > PDB ID.
- 3. Inserisci il codice:** 2HHB e clicca su Load.

Vedrai apparire una struttura complessa composta da 4 catene (2 alfa e 2 beta).

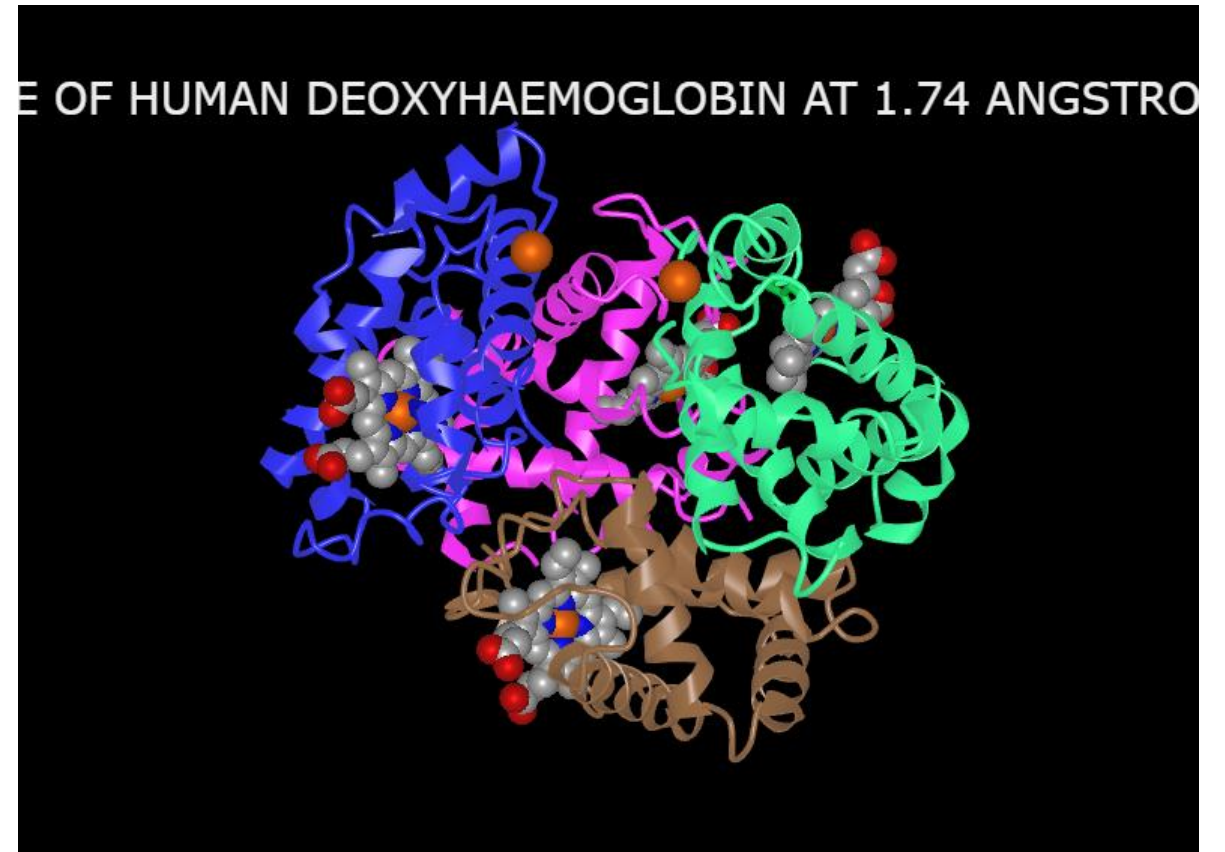
# Orientamento e Visualizzazione

- Ruotare: clicca con il tasto sinistro del mouse e trascina per ruotare la molecola.
- Zoom: usa la rotella del mouse.
- Semplificare la vista: vai su Style > Proteins > Schematic. Questo trasformerà la visualizzazione a "fili" in una struttura a "nastri" dove si vedono chiaramente le alfa-eliche.



# Evidenziare il gruppo eme

- Vai su Select > Nucleotides/Ligands.
- Vai su Style > Chemicals > Sphere.
- I gruppi Eme (dove si lega l'ossigeno) appariranno come grandi sfere al centro delle eliche.



# Trovare l'aminoacido della mutazione (punto cruciale)

Per l'anemia falciforme, la mutazione avviene sulla catena  $\beta$  al residuo 6.

- Vai su **Analysis** > Sequence & Annotations.
- Si aprirà una finestra a destra o in basso.
- Clicca sulla scheda **Details**.
- **Cerca** la riga relativa a una delle catene Beta (solitamente indicate come 2HHB\_B o 2HHB\_D).
- Vedrai la sequenza di lettere: V H L T P E E ...
- Clicca sulla **sesta lettera**, la E (Acido Glutammico).
- Noterai che sulla struttura 3D una piccola parte si illumina (**diventa gialla**). Quello è il punto esatto dove, nel paziente malato, la E viene sostituita dalla V (Valina).

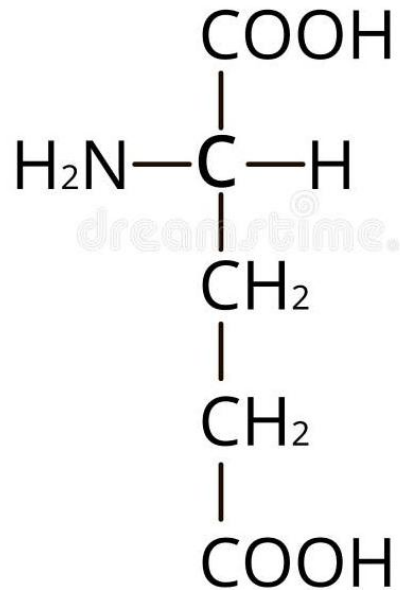
Sigla a tre lettere	Sigla a una lettera	Nome
Gly	G	Glicina
Ala	A	Alanina
Val	V	Valina
Leu	L	Leucina
Ile	I	Isoleucina
Met	M	Metionina
Cys	C	Cisteina
Pro	P	Prolina
Phe	F	Fenilalanina
Trp	W	Triptofano
Tyr	Y	Tirosina
Thr	T	Treonina
Ser	S	Serina
Asn	N	Asparagina
Gln	Q	Glutamina
Asp	D	Acido aspartico o Aspartato
Glu	E	Acido Glutammico o Glutammato
His	H	Istidina
Lys	K	Lisina
Arg	R	Arginina

## Dove si trova quella E (Acido Glutammico)?

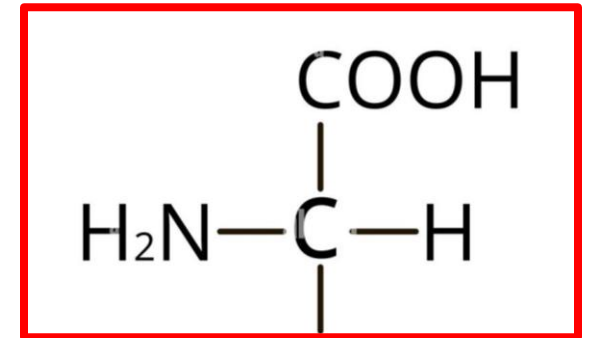
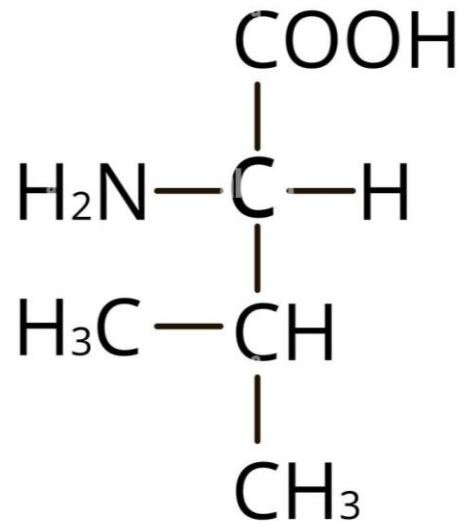
Noterete che si trova sulla superficie esterna della proteina.

Essendo l'**acido glutammico** idrofilico (ama l'acqua), sta bene all'esterno. Se viene sostituito dalla **valina** (che è idrofobica e "unta"), la proteina cercherà di "attaccarsi" ad altre emoglobine per nascondere la macchia di valina dall'acqua, formando i famosi polimeri che deformano il globulo rosso.

### Glutamic Acid



### Valine



PDB ID 2HHB: THE CRYSTAL STRUCTURE OF HUMAN

