Corso di studi in Biologia Molecolare

Informatica e Bioinformatica

Insegnamento di Bioinformatica

a.a. 2012/2013

Docente: **Prof.ssa Elisabetta Bergantino**

Dipartimento di Biologia – 6° piano nord

elisabetta.bergantino@unipd.it

Didattica di supporto: **Dott.ssa Chiara Gardin**

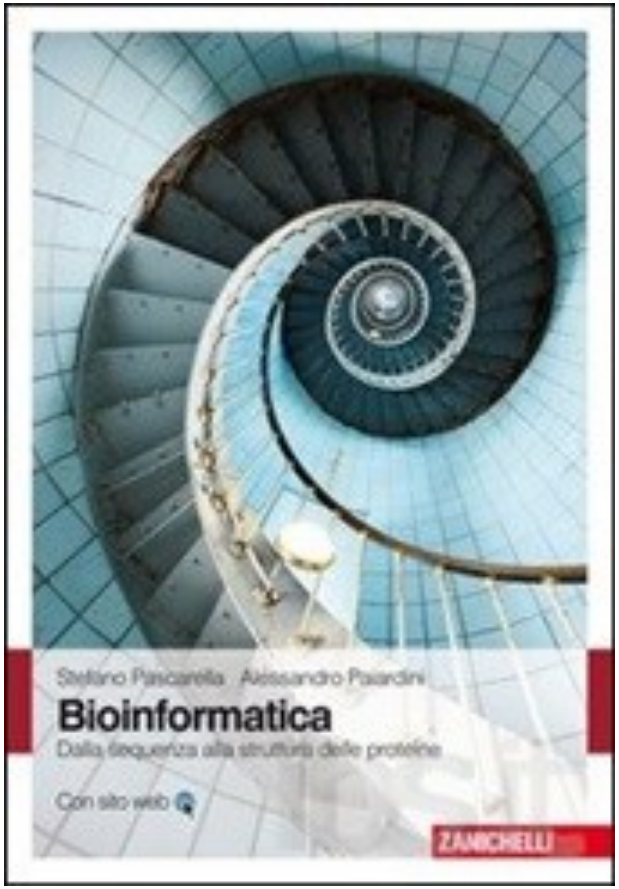
Dott. Andrea Telatin

Ringraziamenti: Proff.ri Ivano Zara e Alessandro Vezzi

Stefano Pascarella – Alessandro Paiardini

BIOINFORMATICA
Dalla sequenza alla struttura delle proteine

Zanichelli



Appunti di lezione

Presentazioni delle lezioni svolte nel sito E-learning

Presentazioni delle esercitazioni alla pagina:

http://didattica.cribi.unipd.it/bioinfouno/2011_2012/esercitazioni/index.html

Due siti interessanti per reperire informazioni e strumenti utili per la bioinformatica:

- ✓ **'2can Bioinformatic Support Portal'**, the bioinformatics educational resource presente all'EBI (European Bioinformatics Institute)

  http://www.ebi.ac.uk/2can/home.html

- ✓ **Basic Introduction to the Science Underlying NCBI Resources**, presente all'NCBI (National Centerfor Biotechnology Information)

  http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html

Prerequisiti: Internet, Inglese, ... Interesse

E' necessario una conoscenza (di base) di alcuni argomenti biologici (DNA, proteine, gene, genoma, procariote, eucariote).

COS'É LA BIOINFORMATICA?

"Applicazione dell'informatica alla gestione e all'analisi dei dati biologici"

Bioinformatics is an interdisciplinary research area that is the interface between the biological and computational sciences.

The ultimate goal of bioinformatics is to uncover the wealth of biological information hidden in the mass of data and obtain a clearer insight into the fundamental biology of organisms.

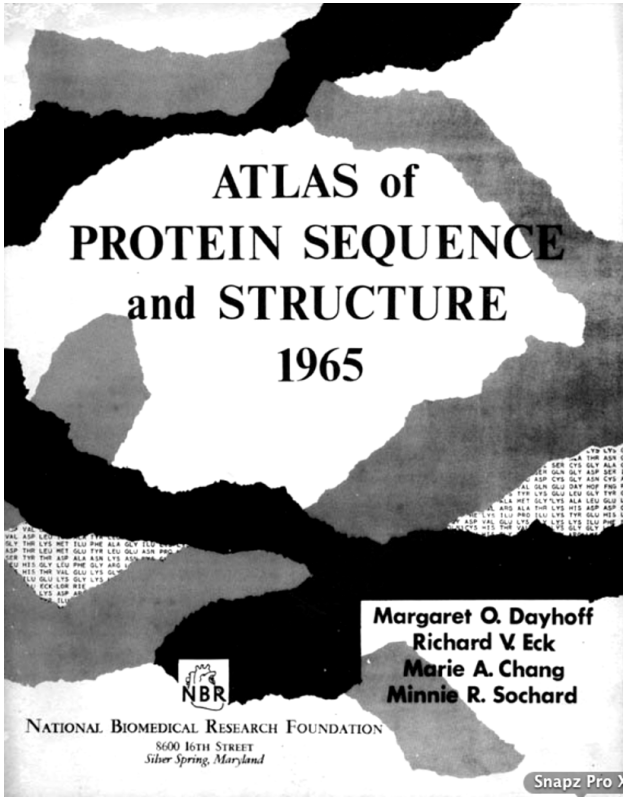
This new knowledge could have profound impacts on fields as varied as human health, agriculture, the environment, energy and biotechnology.

PERCHÉ STUDIARE LA BIOINFORMATICA?

Perché imparare a conoscere e usare le banche dati di Biologia Molecolare?

- ✓ Le banche dati raccolgono, organizzano e consentono l'accesso all'informazione.
- ✓ La portata d'informazione necessaria per la ricerca in biologia e medicina è cresciuta. ben al di là della capacità delle collezioni bibliografiche.
- ✓ Le conoscenze, i dati biologici è cresciuto in maniera esponenziale.
- ✓ Numero e varietà delle risorse disponibili sono aumentate (e aumentano) nel tempo.
- ✓ Aumenta via via l'uso e il numero di utilizzatori.
- ✓ Il livello di familiarità con le risorse (bio)informatiche è molto variabile.

I "dati" biologici

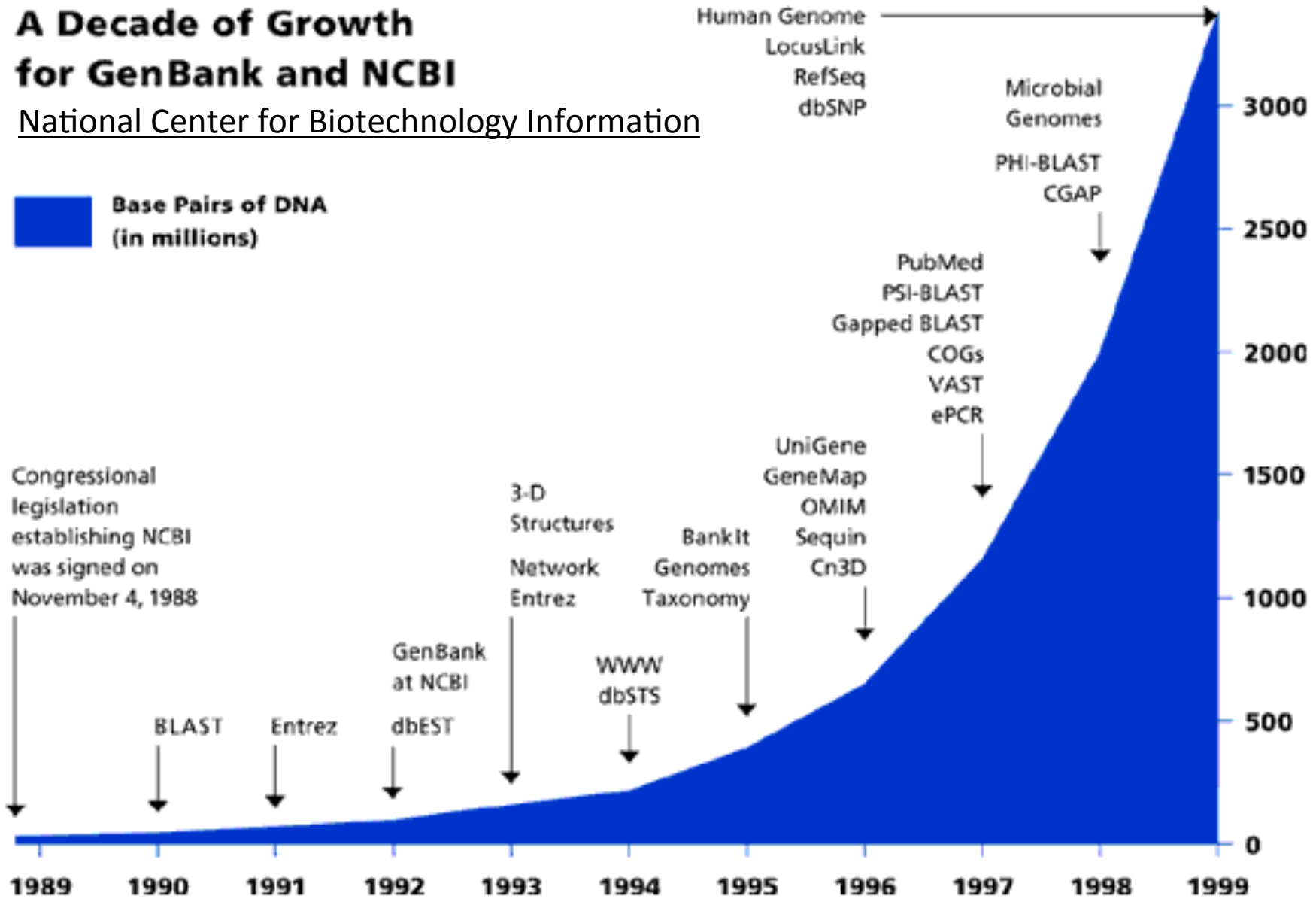


COSA HA DETERMINATO LO SVILUPPO DELLA BIOINFORMATICA:

- Sviluppo di biotecnologie innovative
- Sviluppo delle potenzialità informatiche (hardware)
- Sviluppo di nuovi programmi informatici (software)
- Diffusione delle conoscenze informatiche tra i biologi

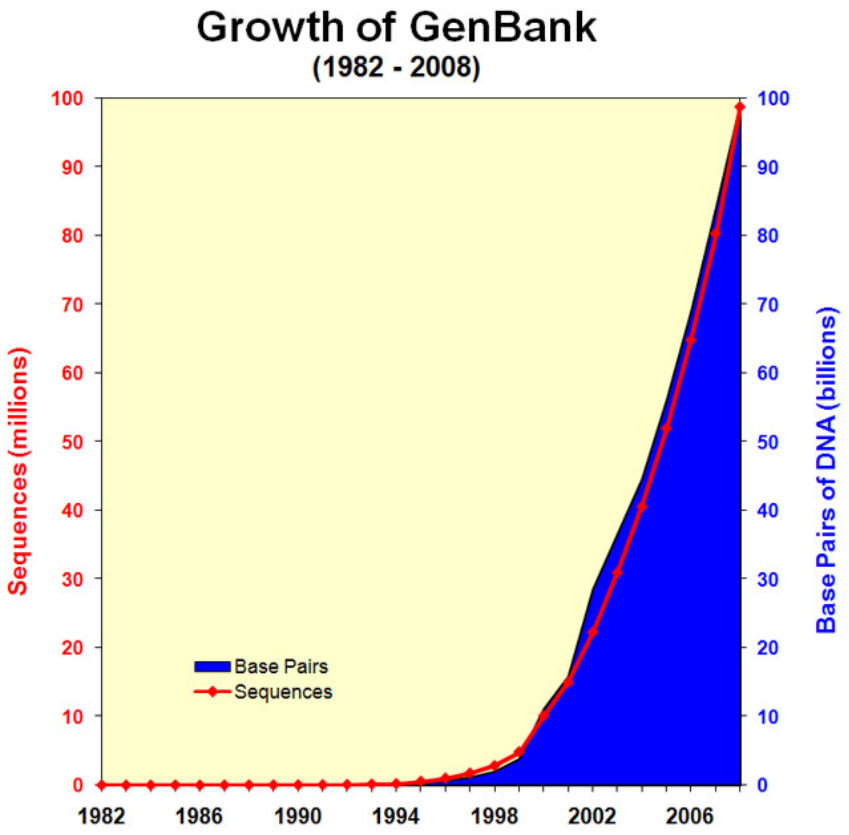
A Decade of Growth for GenBank and NCBI

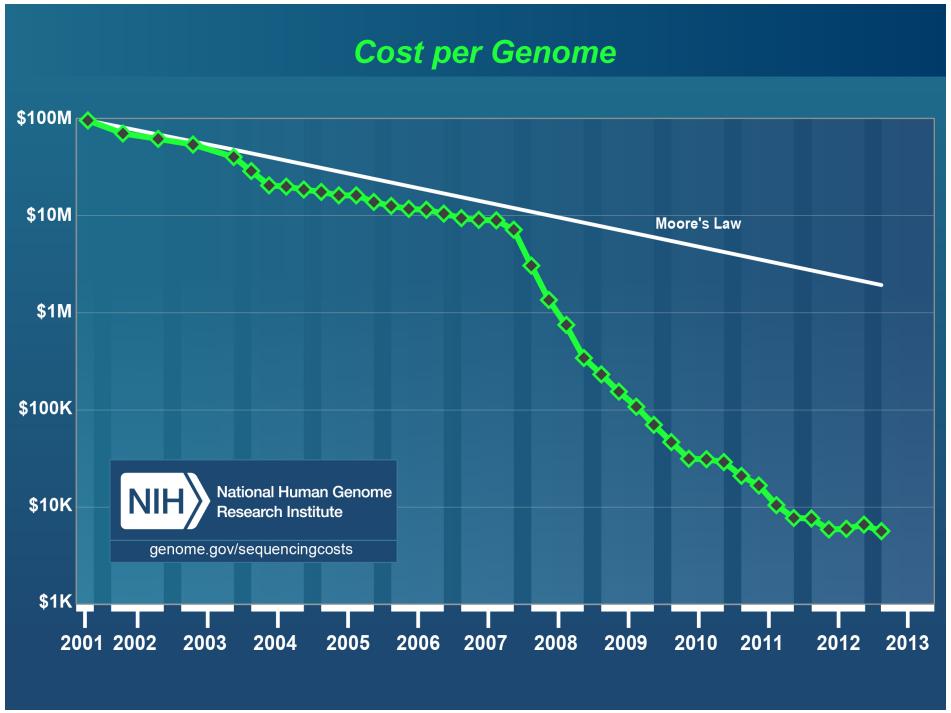
National Center for Biotechnology Information



Negli ultimi 20-30 anni abbiamo assistito ad una vera e propria esplosione nella produzione di dati biologici (aumento esponenziale)

Costo sequenze / costo hardware





Per una bioinformatica 'di base' è necessario:

- -Avere conoscenze biologiche
- -Sapere cosa sono e come sono strutturati i database
- -Conoscere dove sono archiviati i dati biologici
- -Conoscere come sono archiviati questi dati
- -Essere in grado di effettuare ricerche (anche complesse)
- -Saper utilizzare i numerosi strumenti ('tool') che sono pubblicamente disponibili

Questo corso ha lo scopo di introdurre sommariamente alcuni dei principali argomenti della biologia e di fornire gli strumenti ed i metodi per accedere all'informazione biologica in modo razionale ed efficiente, utilizzando le risorse disponibili in rete.

In tre parti:

Archiviazione dati: DATABASE
- come vengono memorizzati i dati
- come strutturare gli archivi

Database Biologici
- database di sequenze di DNA e proteine
- database articoli scientifici
- database "tematici"

Analisi computazionale dei dati.
- allineamento di sequenze e ricerca di similarità
- Uso di strumenti informatici (tools) per analizzare dati

Per una bioinformatica 'di base' è necessario:

- -Avere conoscenze biologiche → gli oggetti della bioinformatica

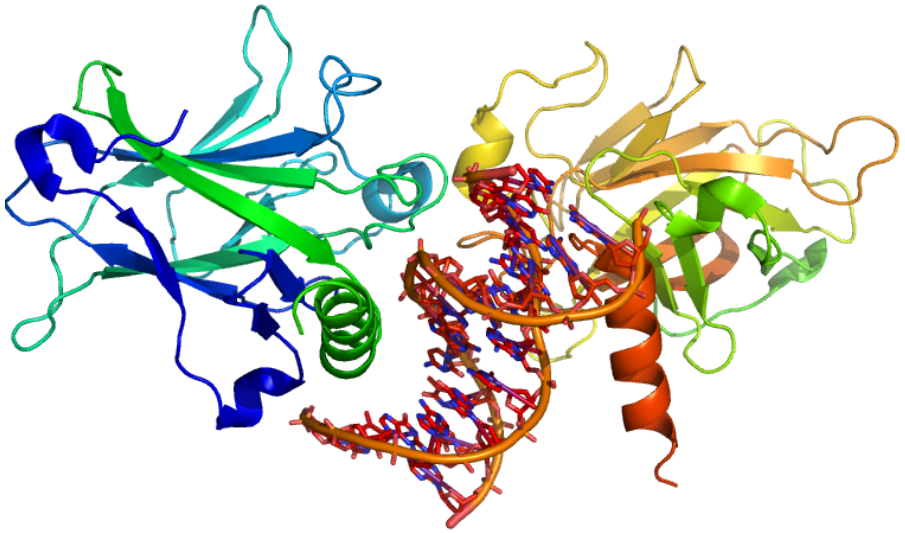
Sequenze di acidi nucleici

>gi|8886401|gb|AF162269.1|
CCCACTCCTCCATCTCACAAACACTTCTCTATACCCAACAATCCCTTTTACAAT
CCCTGCTCATTTAGTCAAAATGGTCAAGATTGCTGCTATCATCCTCCTCATGG
GCATTCTCGCCAATGCTGCCGCCATCCCTGTCATTTCAACACCCAAATTACAG
AGCCAACCGGCGAGGGCGACCGTGGGGACGTGGCCGAC

Sequenze proteiche

>P25032
MASSSSATSGDDRPPAAGGGTPAQAHAEWAASMHAYYAAAASAAGHPYAW
PLPPQAQQHGLVAAGAGAAYGAGAVPHVPPPPAGTRHAHASMAAGVPYMA

Strutture di macromolecole



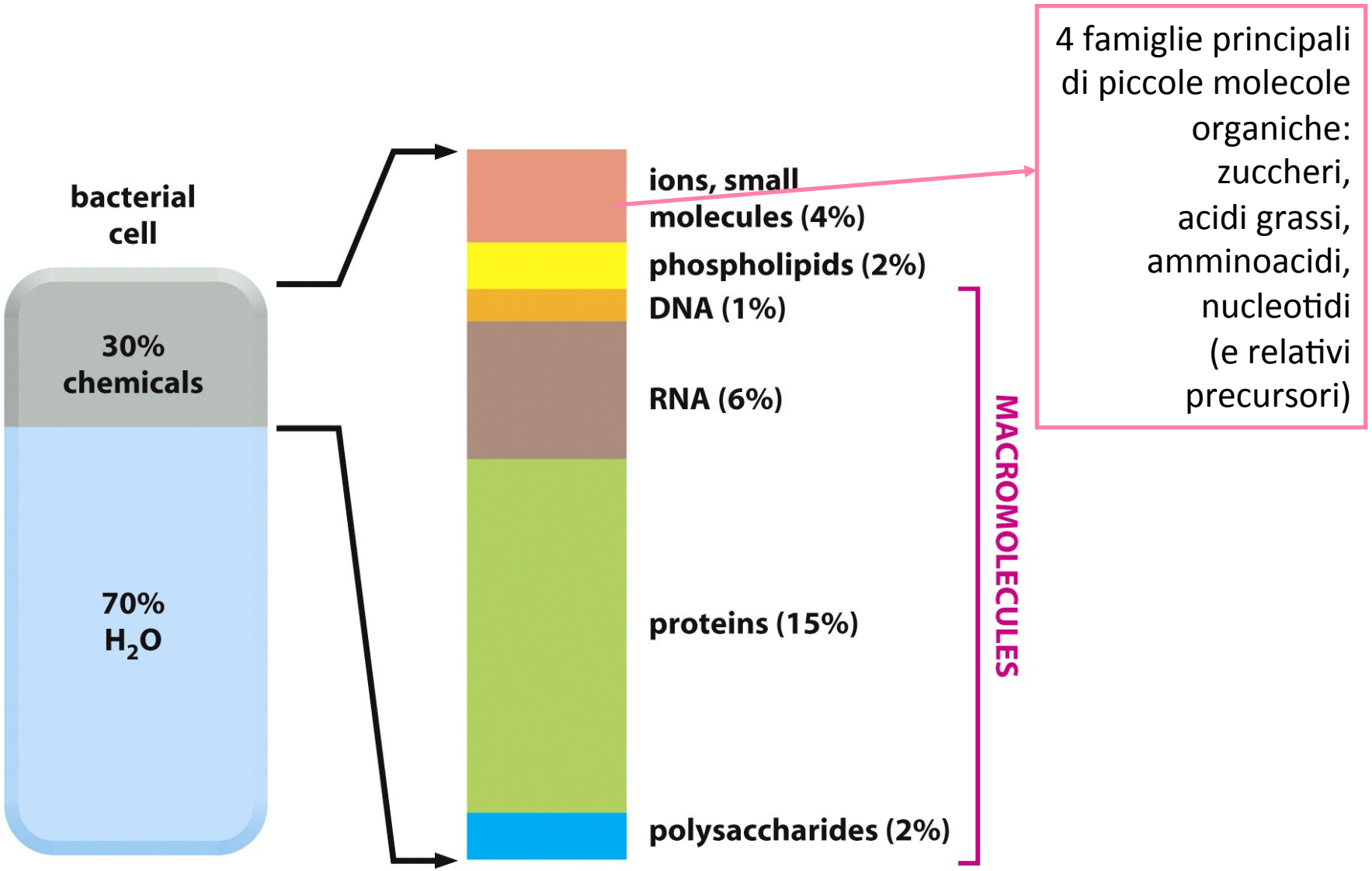


Figure 2-29 *Molecular Biology of the Cell* (© Garland Science 2008)

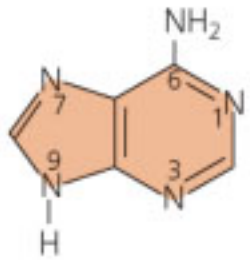
Per una bioinformatica 'di base' è necessario:

- **-Avere conoscenze biologiche** → gli elementi costituenti delle macromolecole biologiche

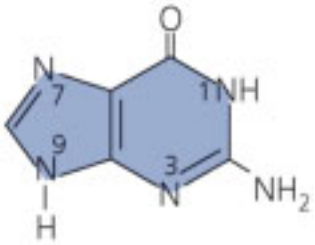
acidi nucleici

Basi azotate

Purine

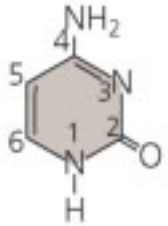


Adenina (A)

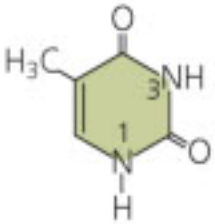


Guanina (G)

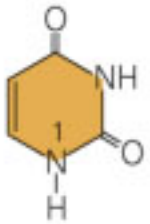
Pirimidine



Citosina (C)

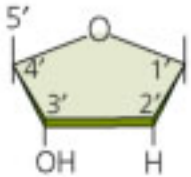


Timina (T)
(DNA)

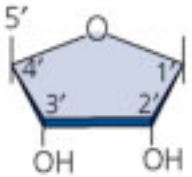


Uracile (U)
(RNA)

Zuccheri

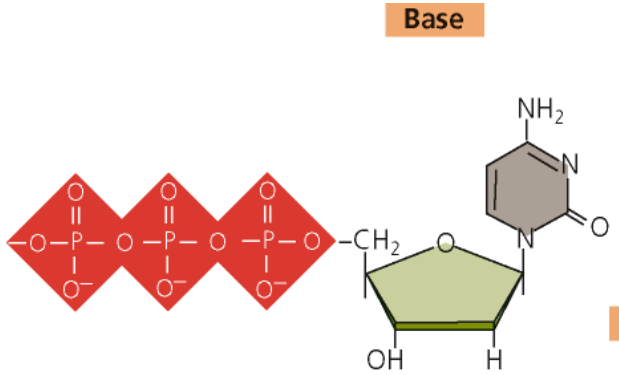


2'-deossiribosio
(DNA)



Ribosio
(RNA)

Gruppo fosfato



① Nucleoside

Nucleoside monofosfato

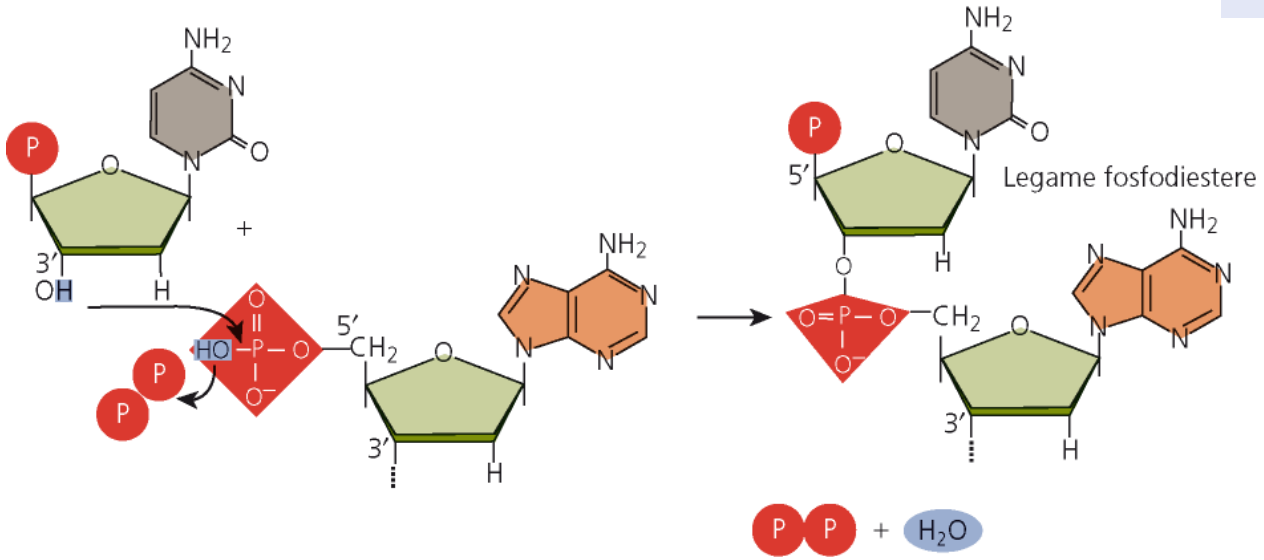
Nucleoside difosfato

Nucleoside trifosfato

② Nucleotidi

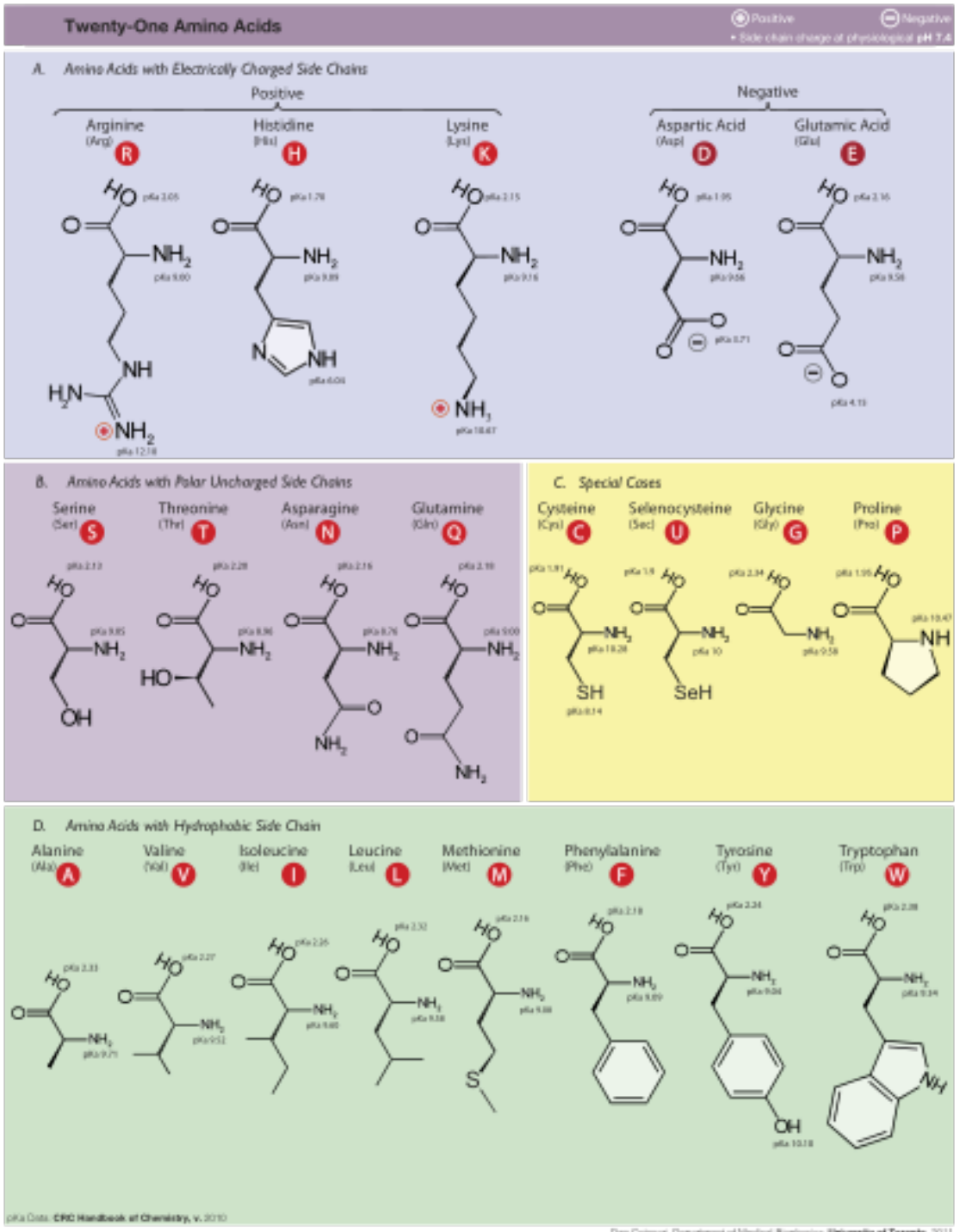
③ Catena di DNA – nucleotide + nucleotide + nucleotide + …

Direzione 5’ – 3’

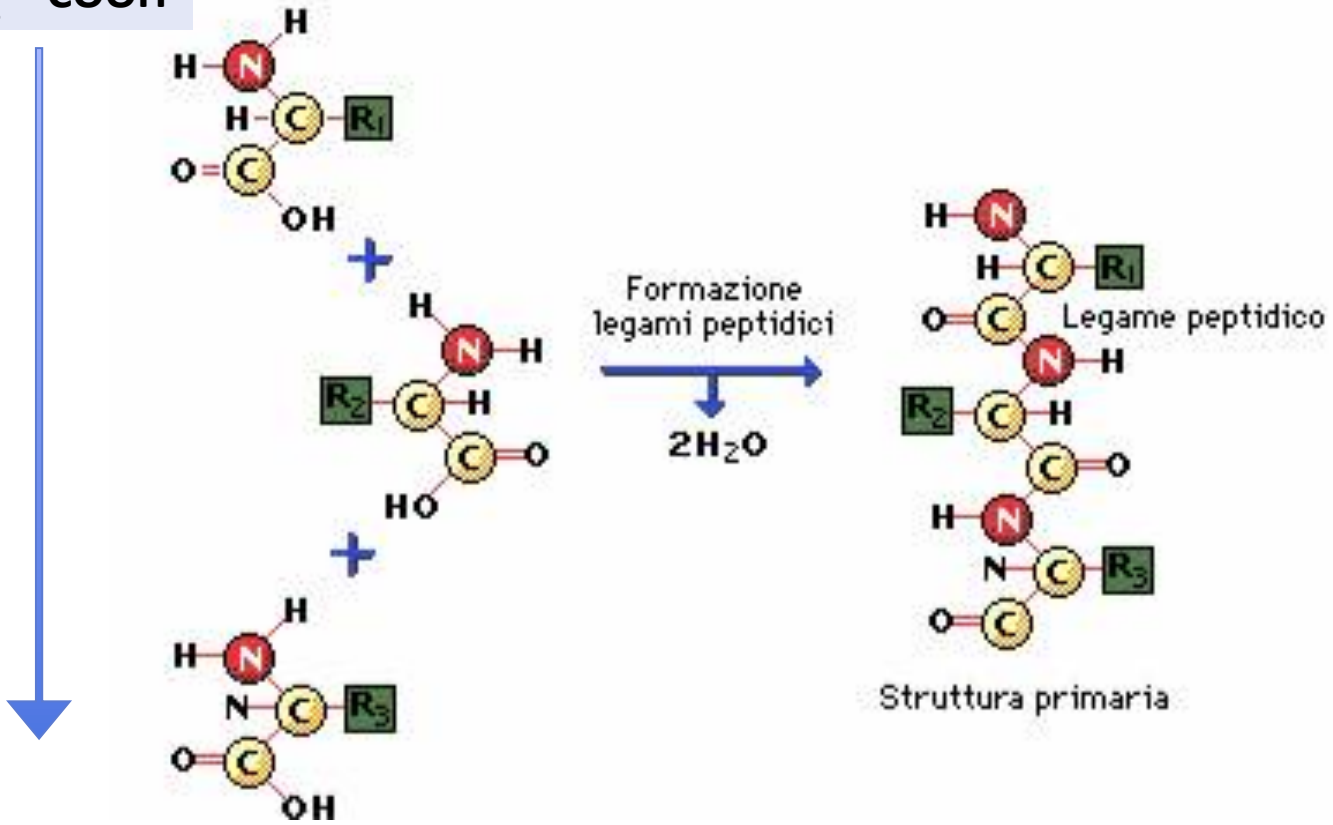


proteine

Amminoacidi



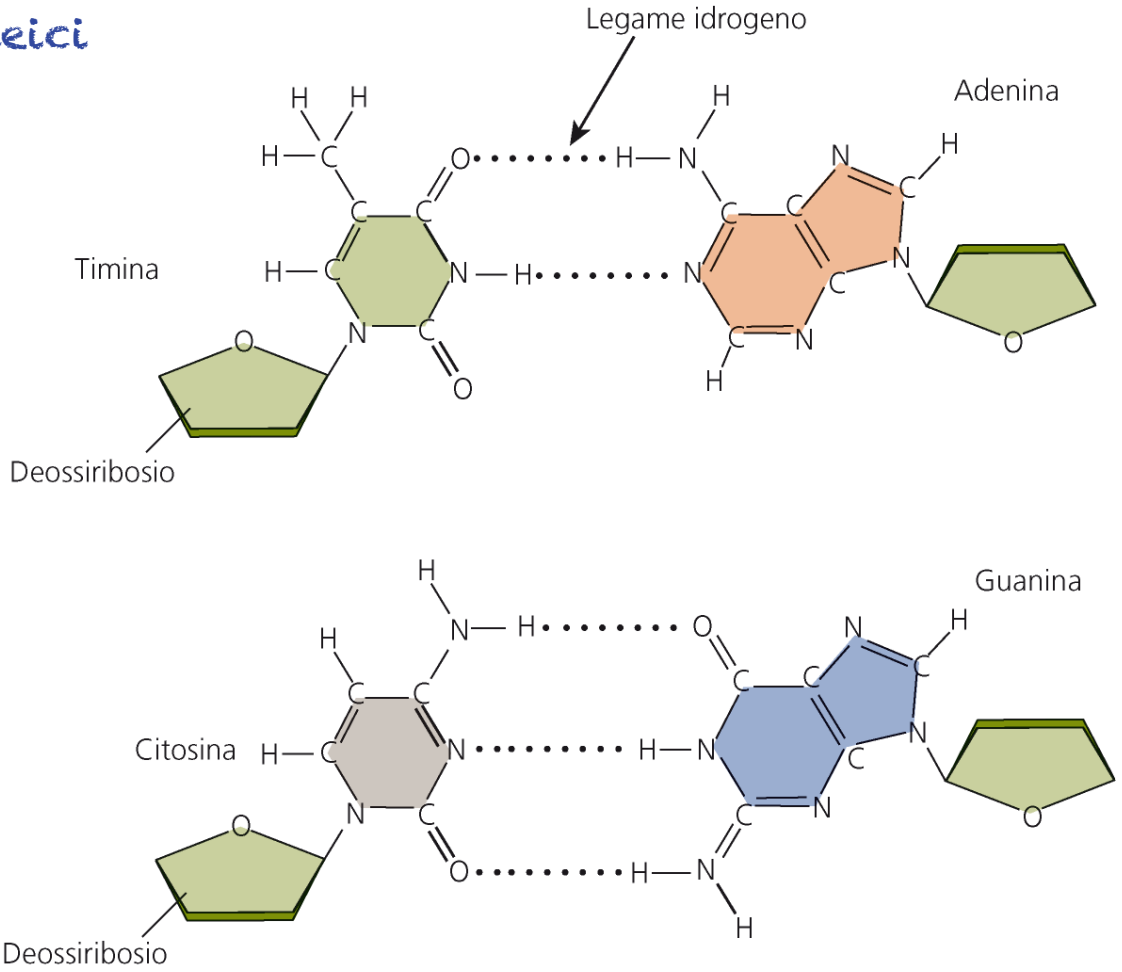
Direzione
NH_2 – COOH



Per una bioinformatica 'di base' è necessario:

- -Avere conoscenze biologiche ➔ L'organizzazione strutturale delle macromolecole biologiche

acidi nucleici



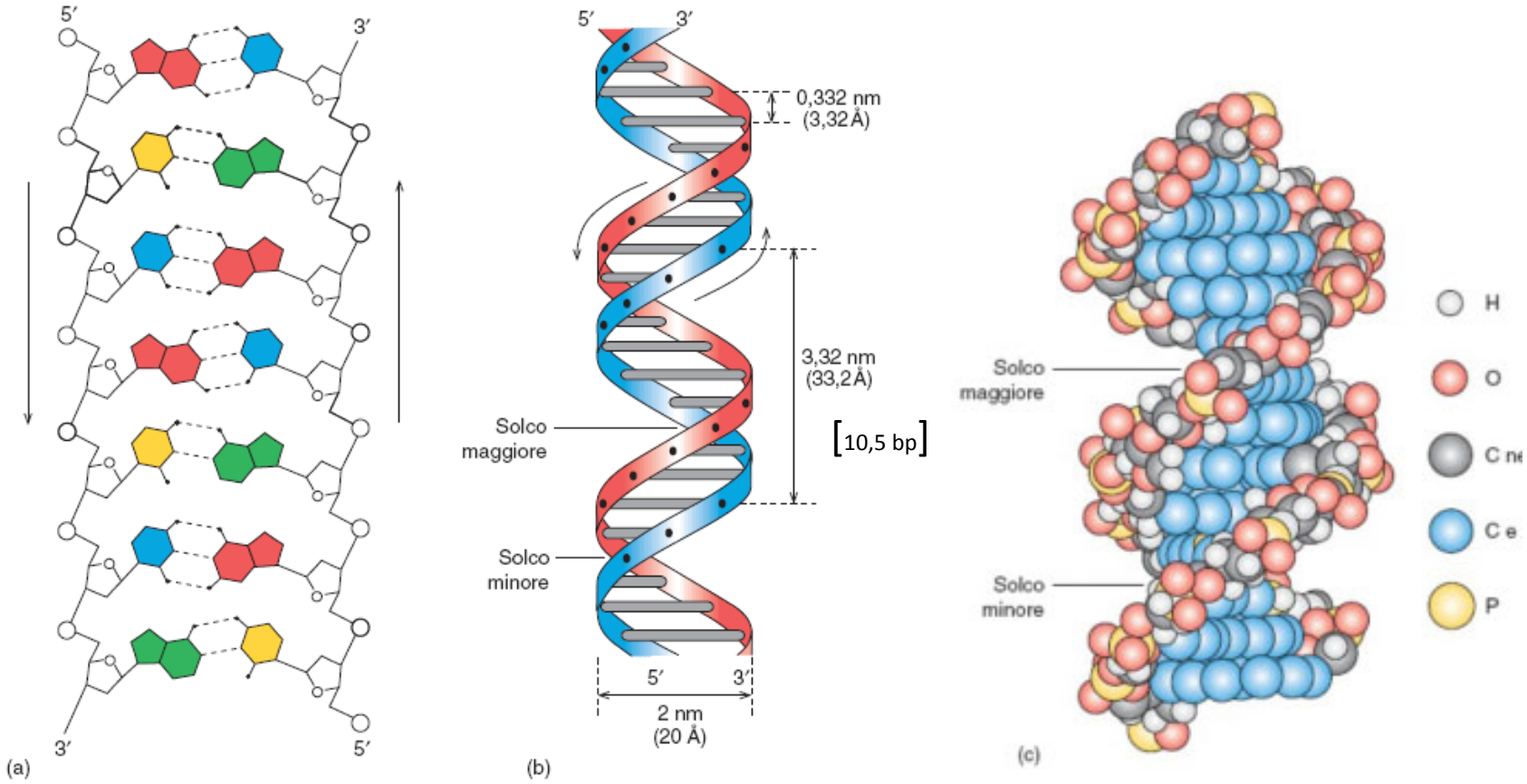
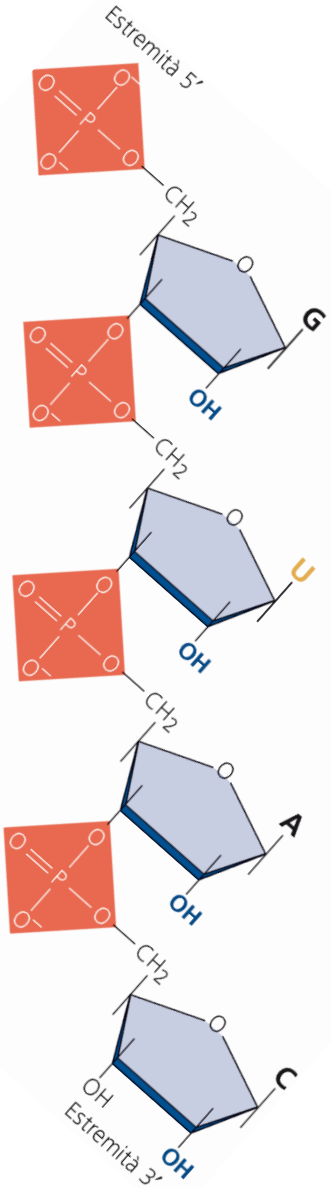
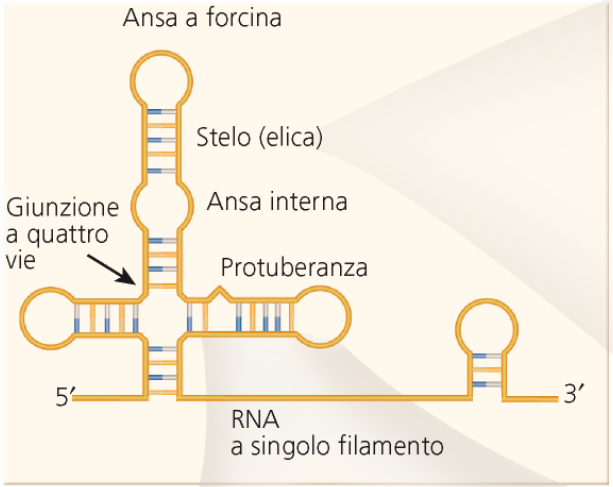
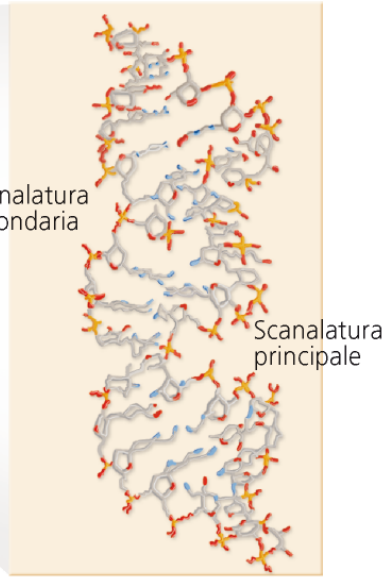


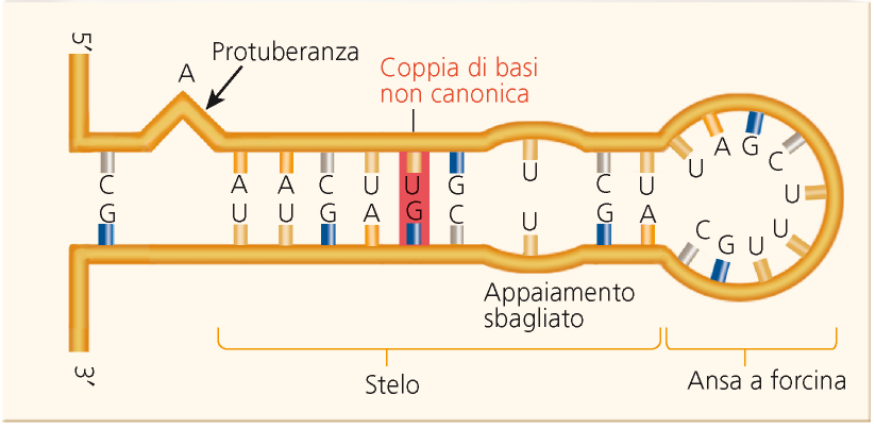
Figura 2.14 Tre modelli di struttura del DNA. **(a)** L'elica è rappresentata appiattita per mostrare gli appaiamenti tra le basi. Ciascuna base viene rappresentata con un colore diverso mentre lo scheletro costituito dai gruppi zucchero-fosfato viene mostrato in nero. Si notino i tre legami idrogeno presenti nell'accoppiamento tra G e C e i due legami idrogeno presenti tra A e T. Le frecce verticali presenti ai lati dei due filamenti indicano la direzione 5′→ 3′, e mostrano come i due filamenti di DNA siano antiparalleli l'uno rispetto all'altro. Il filamento rappresentato a sinistra corre in direzione 5′→ 3′, dall'alto verso il basso, mentre il filamento rappresentato a destra corre in direzione 5′→ 3′ dal basso verso l'alto. Anche gli anelli che rappresentano il deossiribosio (pentagoni bianchi, con la O che rappresenta l'ossigeno) mostrano come i due filamenti corrano in direzioni opposte: gli anelli presenti sul filamento di destra sono invertiti rispetto a quelli presenti sull'anello di sinistra. **(b)** La doppia elica di DNA viene rappresentata come una scala a chiocciola, in cui i lati a spirale rappresentano i gruppi zucchero-fosfato e i pioli rappresentano le coppie di basi. Le frecce ricurve presenti ai lati dei due filamenti indicano la direzione 5′→ 3′, mettendo in evidenza l'andamento antiparallelo dei due filamenti. **(c)** Questa figura rappresenta un modello molecolare a spazi pieni. Lo scheletro costituito dai gruppi zucchero-

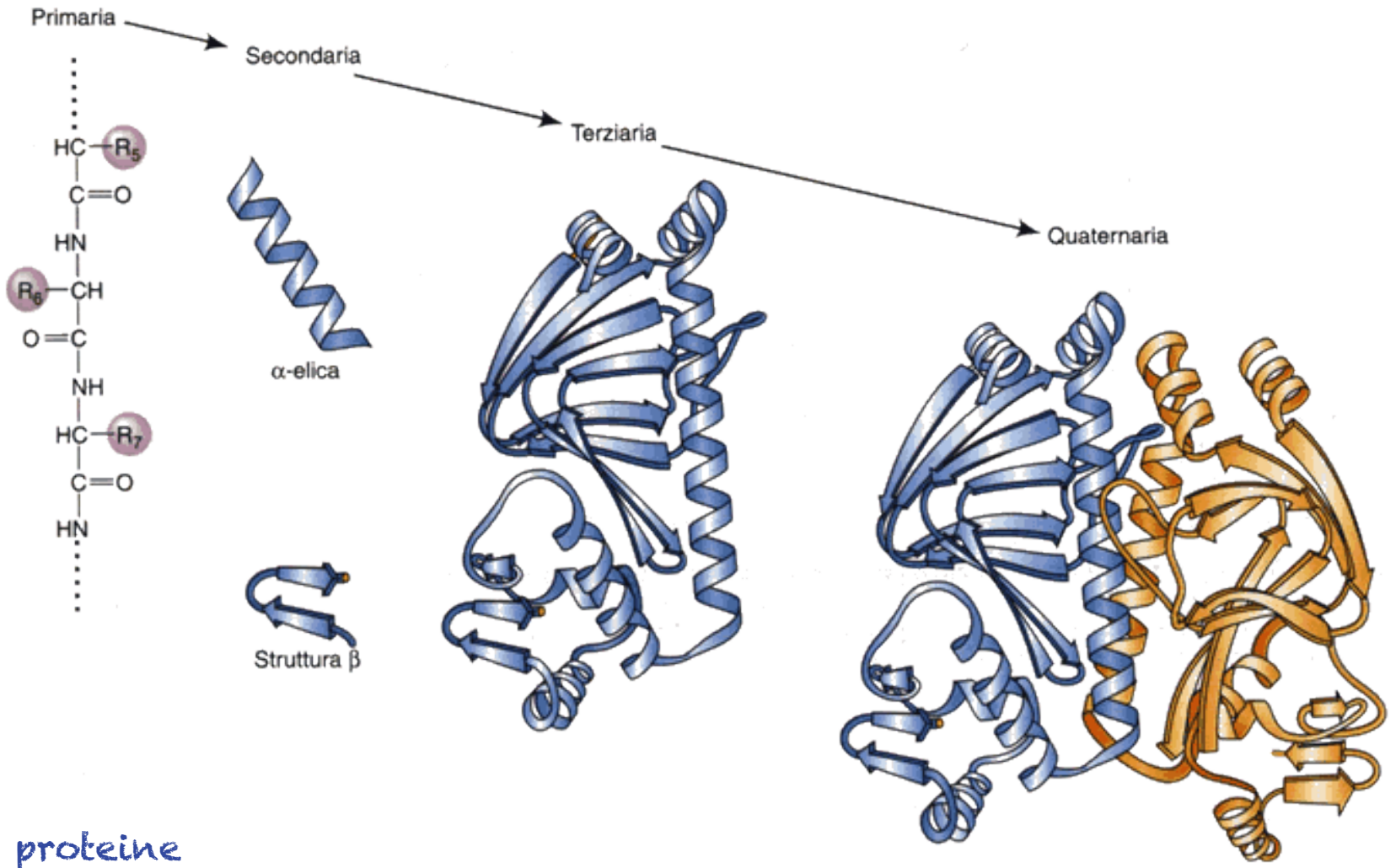






Una doppia elica di RNA

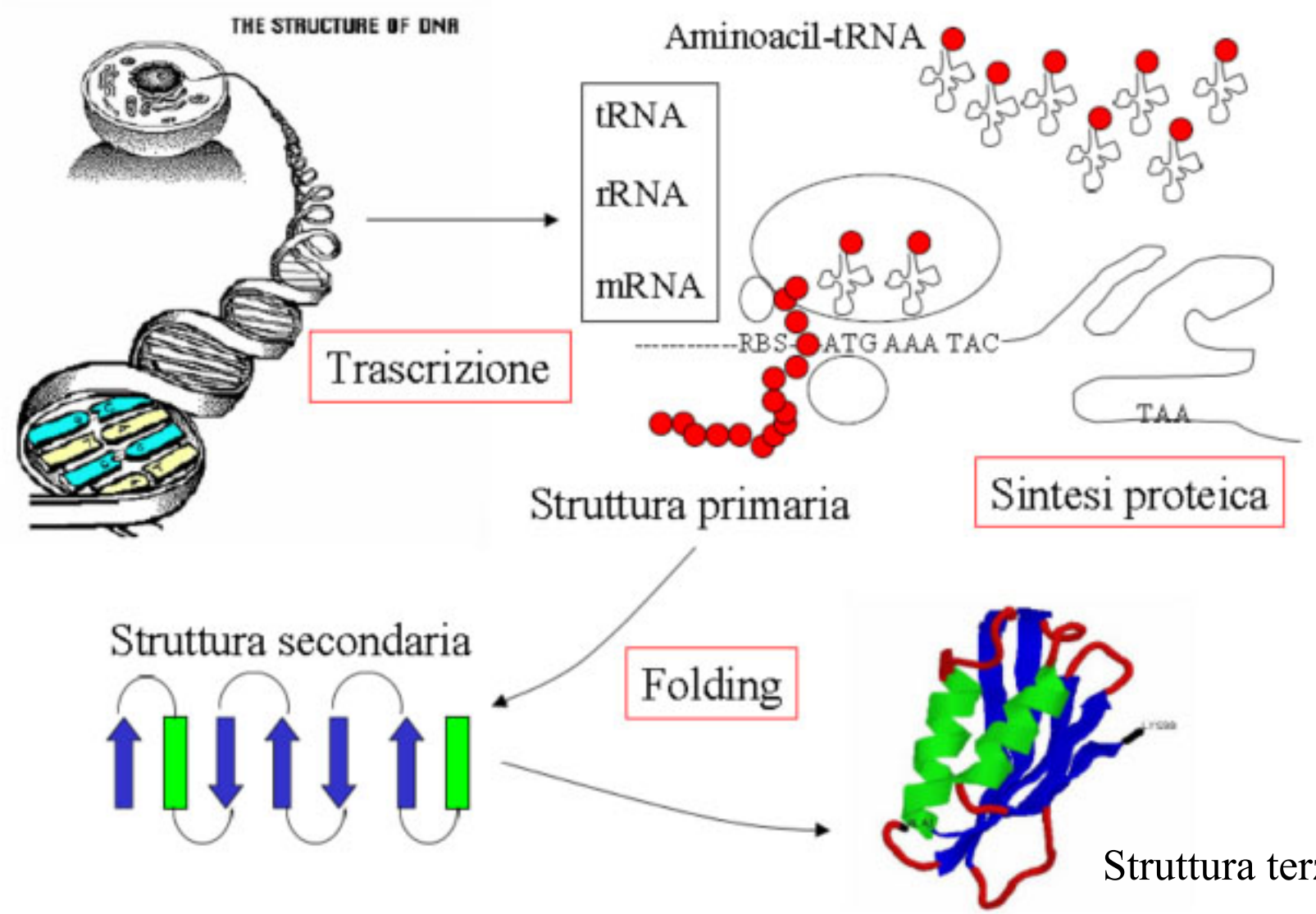


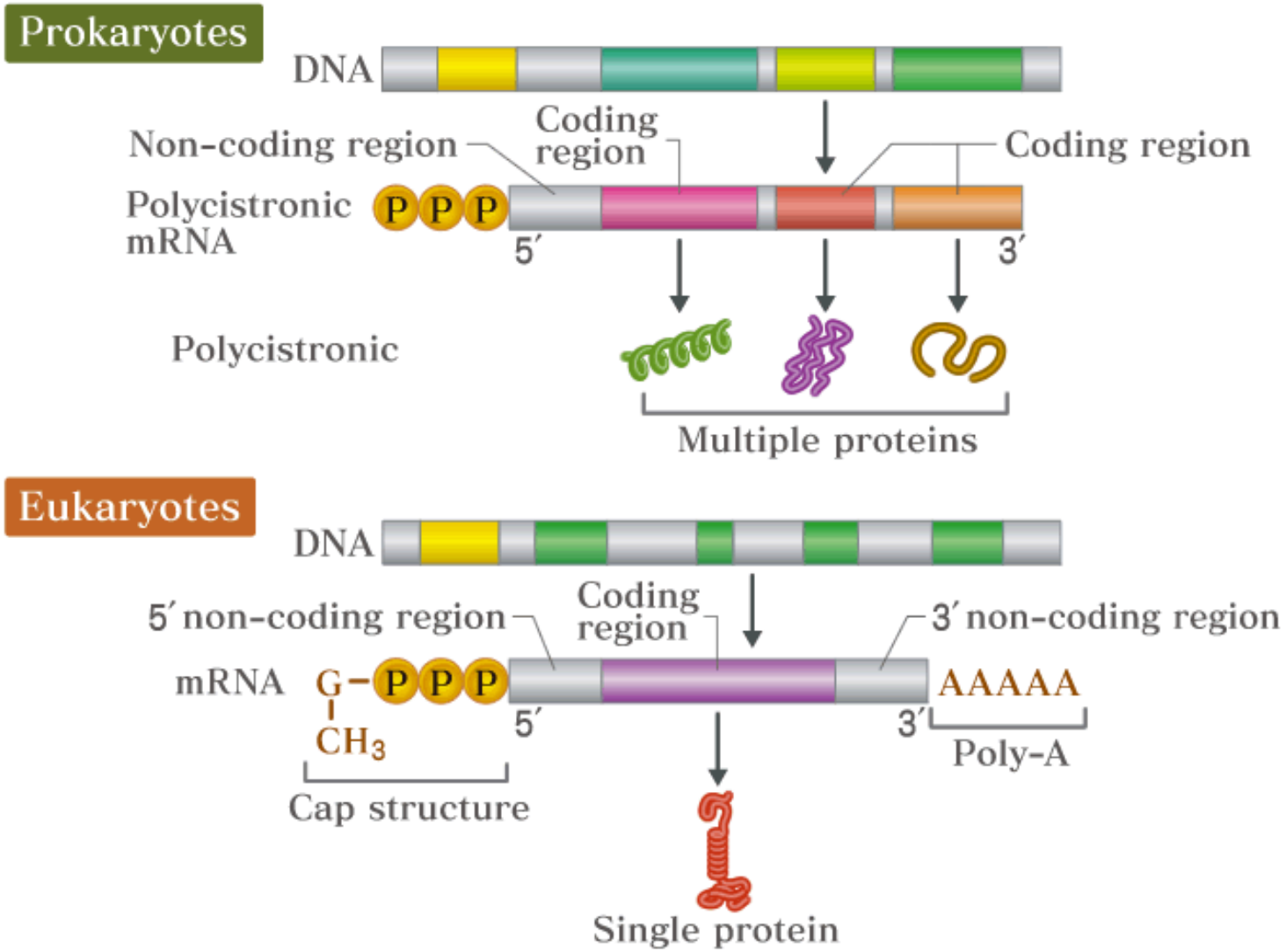

Primaria
Secondaria
Terziaria
Quaternaria
HC—R5
C=O
HN
R6—CH
O=C
NH
HC—R7
C=O
HN
α-elica
Struttura β

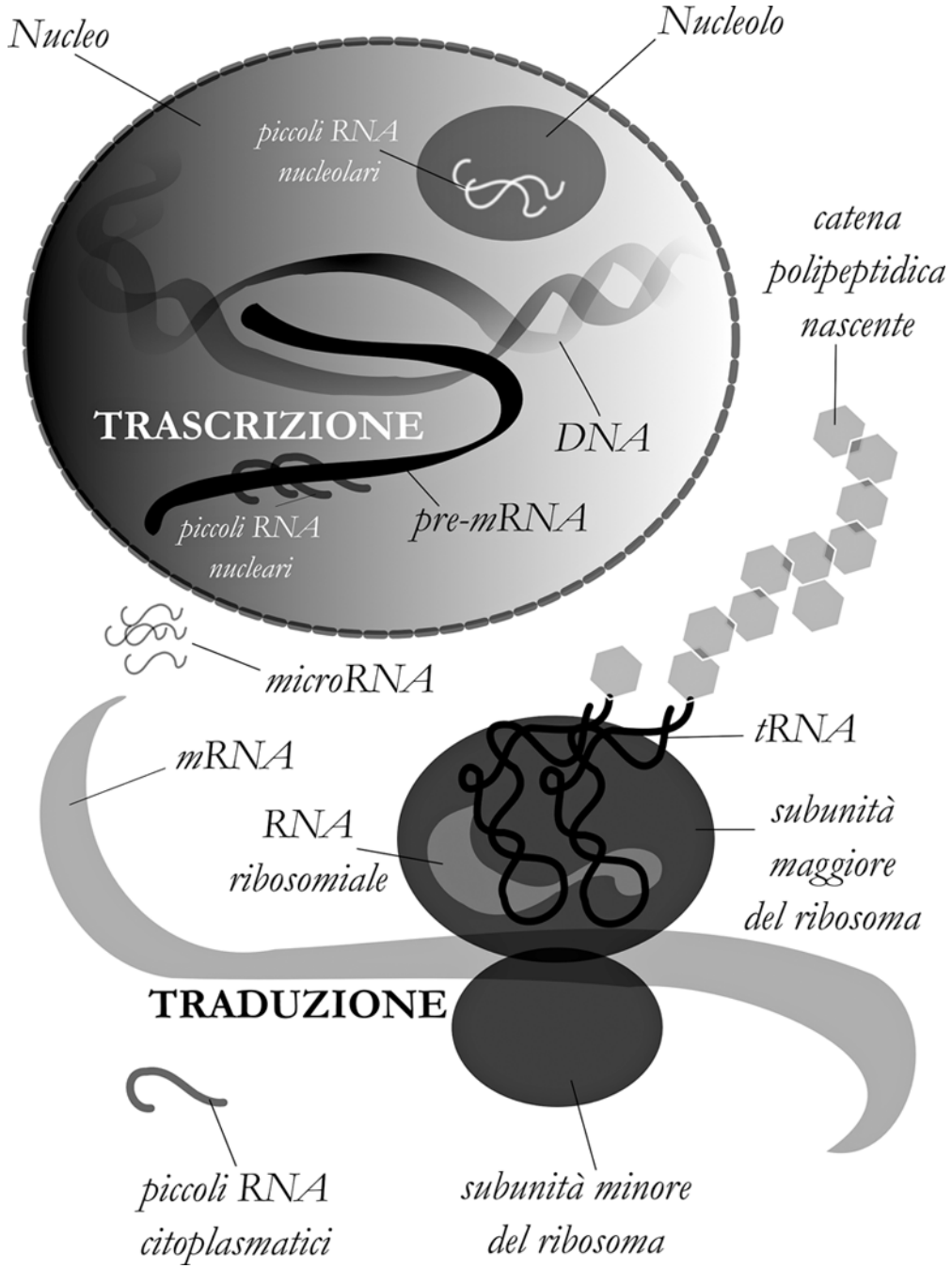
proteine

Per una bioinformatica 'di base' è necessario:

- -Avere conoscenze biologiche → i processi biologici in esame






Nucleo
Nucleolo
piccoli RNA nucleolari
catena polipeptidica nascente
TRASCRIZIONE
DNA
piccoli RNA nucleari
pre-mRNA
microRNA
tRNA
mRNA
RNA ribosomiale
subunità maggiore del ribosoma
TRADUZIONE
piccoli RNA citoplasmatici
subunità minore del ribosoma

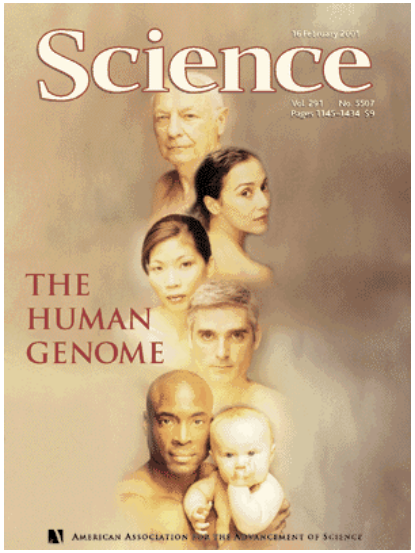
La bioinformatica 'avanzata' tratterà, in seguito:

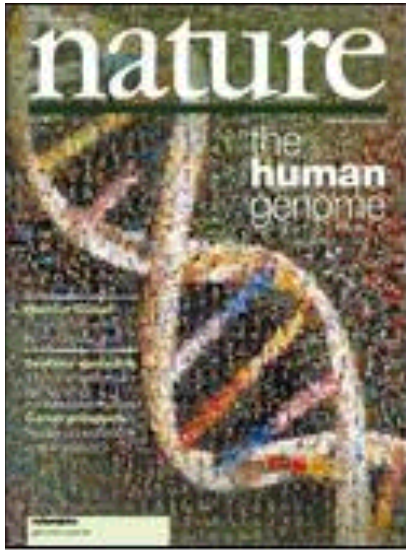
➔ **l'organizzazione delle macromolecole biologiche in insiemi complessi e articolati**

"*A prerequisite to understanding the complete biology of an organism is the determination of its entire genome sequence*" Fleischmann *et al*. 1995

A partire dagli anni '80-'90 si avviano progetti di sequenziamento di interi genomi (determinazione della sequenza lineare delle basi che compongono il DNA (A,T,C,G))

La conoscenza di interi genomi → aumento esponenziale sequenze di geni
→ **bioinformatica per archiviazione, l'organizzazione e la lettura dei dati**





2000-2001
Il genoma umano è composto da 3.12 miliardi di paia di basi

Genomica: è la disciplina che studia genomi completi

Dalla genomica sono derivate per assonanza numerosi termini che indicano lo studio d'insieme di vari aspetti degli esseri viventi:

Trascrittomica
Proteomica
Metabolomica
Glicosilomica
Farmacogenomica
….

→ NECESSITA' di un SISTEMA di ARCHIVIAZIONE (e di recupero/utilizzo) dei dati facile ed esaustivo

Le pagine seguenti sono estratte dalle dispense

dell' a.a. 2011/2012

Insegnamento di Bioinformatica

Prof. Ivano Zara

- Le prossime lezioni saranno incentrate sulla struttura delle principali macromolecole della cellula (DNA, RNA e proteine) e sul processo di traduzione della sequenza di DNA in proteine.
- Lo scopo è quello di dare una rapida visione a questi argomenti (che verranno poi ripresi molto più approfonditamente in corsi specifici) in modo da facilitare la comprensione dei successivi argomenti del corso.

Concetti Biologici basilari

Per approfondire http://www.ebi.ac.uk/2can/biology/index.html (in inglese)

Le -OMICS

La bioinformatica, in particolare, si occupa di gestire ed analizzare i dati che vengono prodotti in modo sistematico nelle numerose e più disparate discipline biologiche, quelle a cui spesso ci si riferisce, forse un po' esagerando, come **–OMICS**.

Ad esempio, la Genomica è la disciplina che si occupa di produrre, gestire ed analizzare i dati del genoma.

INSIEME DEI DATI	DISCIPLINA
GENOME	**GENOMICS**
PROTEOME	**PROTEOMICS**
TRASCRIPTOME	**TRASCRIPTOMICS**
METABOLOME	**METABOLOMICS**
BIBLIOME	**BIBLIOMICS**

Curiosità. Un sito che elenca tutte le –omics citate in letteratura:
http://www.genomicglossaries.com/content/omes.asp

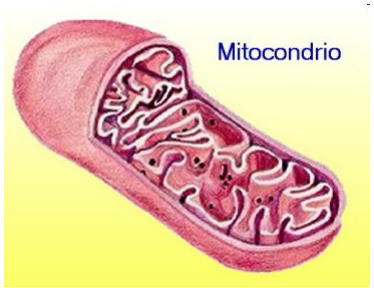
Procarioti ed Eucarioti

Procarioti (conosciuti anche come microbi) sono organismi unicellulari con una organizzazione relativamente semplice. Non contengono particolari organelli caratteristici degli eucarioti. Il materiale genetico (DNA) non è racchiuso in una particolare struttura.

Gli **Eucarioti** hanno un nucleo dove è contenuto il DNA ed hanno dei compartimenti interni, racchiusi da membrane, chiamati organelli, che assolvono a particolari compiti biologici (complesso del Golgi, lisosomi, **mitocondri** ecc.)

Mitocondri: Organelli cellulari, racchiusi da due membrane, posseggono un proprio DNA circolare a doppia elica (genoma mitocondriale). Una loro importane funzione è la produzione di energia (sotto forma di ATP) attraverso l'ossidazione di substrati organici.

Si pensa che, in origine, i mitocondri fossero dei batteri, inglobati dalle cellule eucariote con conseguente mutuo beneficio. Successivamente i batteri avrebbero trasferito gran parte del loro materiale genetico a quello cellulare, divenendo così, mitocondri (teoria endosimbiontica).



Principali molecole biologiche

-**Lipidi**: grazie alle loro caratteristiche chimiche formano membrane che racchiudono le cellule e gli organelli cellulari negli eucarioti.

-**Proteine**: svolgono quasi tutte le funzioni biologiche: formano strutture citoscheletriche; catalizzano alcune reazioni chimiche (enzimi); forniscono attività motorie alle cellule e nelle cellule; dirigono l'esportazione, l'importazione e lo spostamento di varie molecole; trasmettono particolari segnali tra i vari compartimenti cellulari, oppure come ormoni o fattori di crescita trasmettono segnali di controllo a differenti cellule; agiscono come anticorpi, enzimi digestivi; costituiscono tossine e veleni naturali; ecc.

-**Acidi Nucleici** (DNA e RNA) : Il DNA codifica l'informazione per costruire le proteine, l'RNA fornisce lo stampo per la sintesi delle proteine, interviene nella formazione di macromolecole complesse ed in particolari processi biologici.

Proteine in sintesi:

-Le proteine sono polimeri lineari di aminoacidi, uniti chimicamente l'uno all'altro tramite legami peptidici.

- Sono costituite essenzialmente da 20 possibili aminoacidi diversi

- La sequenza con cui gli aminoacidi si succedono l'uno all'altro determina le proprietà di ogni proteina.

- Le differenti combinazioni dei 20 aminoacidi consentono la formazione di innumerevoli sequenze proteiche (es. proteina di 50 aa può avere 20^{50} sequenze differenti)

-Esistono proteine di lunghezze molto diverse, da pochi aminoacidi (in questo caso sono generalmente chiamate peptidi) a diverse migliaia di aminoacidi (Le proteine più comuni sono lunghe 50-1000 aa)

Dove trovare informazioni 'gratis'

breve corso sulla struttura delle proteine: http://webhost.bridgew.edu/fgorga/proteins/default.htm
A Review of Amino Acids http://wbiomed.curtin.edu.au/teach/biochem/tutorials/AAs/AA.html

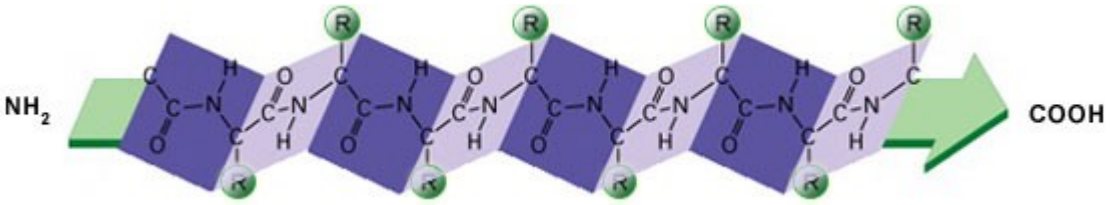
Sequenza di una proteina rappresenta la struttura primaria ed è l'ordine con cui gli aminoacidi si succedono nella molecola

La sequenza viene rappresentata con una 'stringa' di caratteri che rappresentano i simboli degli aminoacidi

es.: P R T W Q E R P R R T W C S S G R

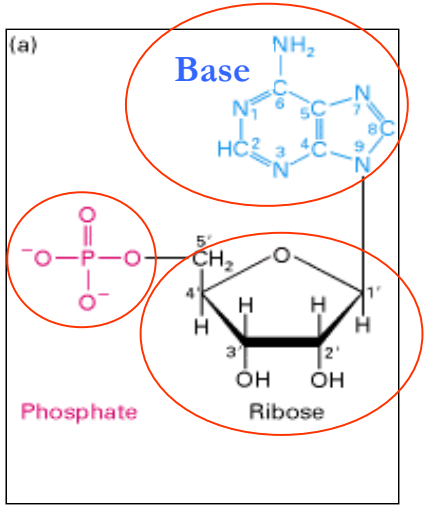
In una proteina la sequenza di aminoacidi ha una direzione. Per convenzione la sequenza si scrive a partire dall'estremità **NH2-(amino)terminale** all'estremità **COOH (carbossi)terminale** (corrisponde alla direzione di sintesi)

Quindi la sequenza 'ACDE' è diversa da 'EDCA'

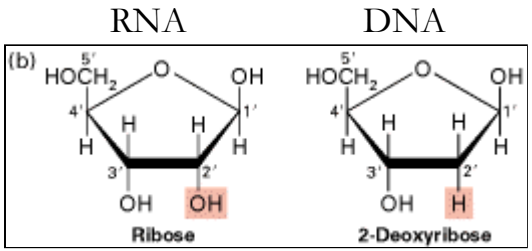


Acidi nucleici (DNA e RNA)

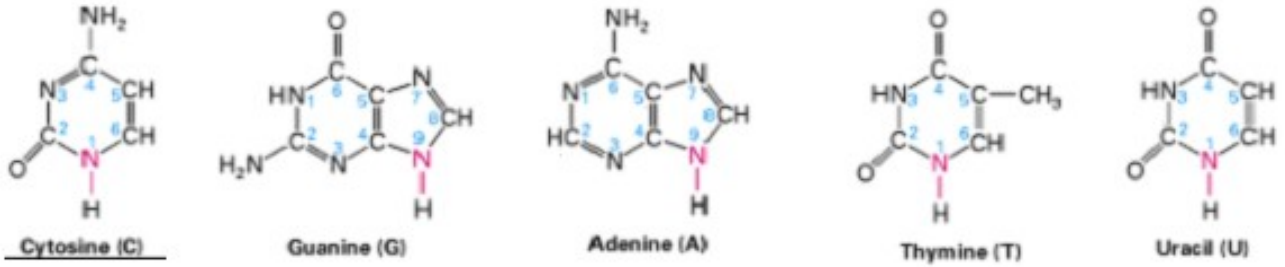
Gli acidi nucleici (DNA: *acido deossiribonucleico* e RNA: *acido ribonucleico*) sono dei polimeri organici costituiti da monomeri chiamati **nucleotidi**.



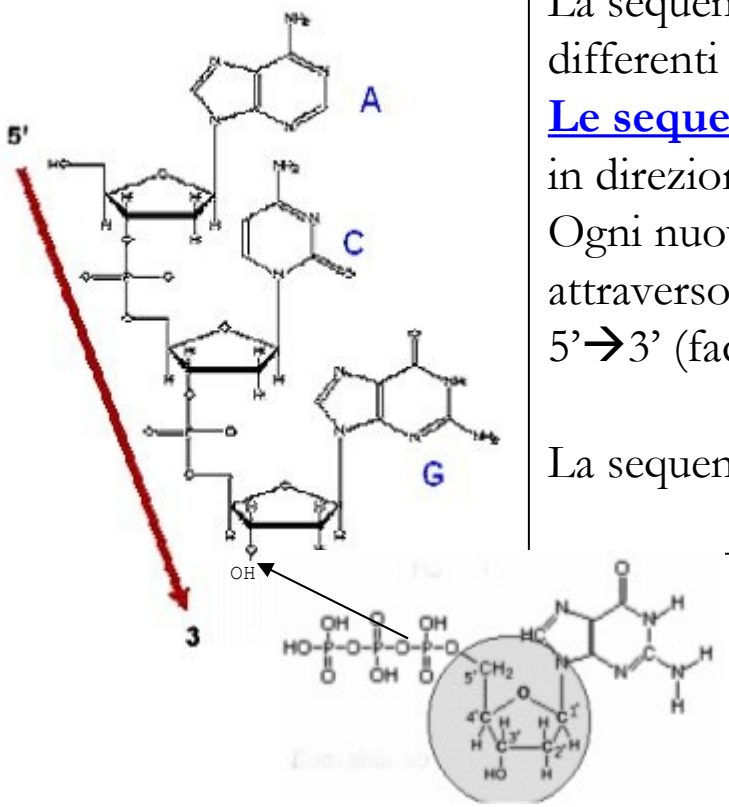
Tutti i nucleotidi sono costituiti da tre componenti fondamentali: un **gruppo fosfato**, una molecola di **zucchero pentoso** (**deossiribosio nel DNA o ribosio nell'RNA**) e una **base azotata** che si lega al carbonio 1' dello zucchero.



Gli acidi nucleici possono essere formati da solo quattro possibili basi azotate: **adenina**, **guanina**, **citosina** (comuni al DNA e all'RNA), la **timina** presente solo DNA mentre l'**uracile** solo nel RNA.



Direzionalità delle sequenze di acidi nucleici



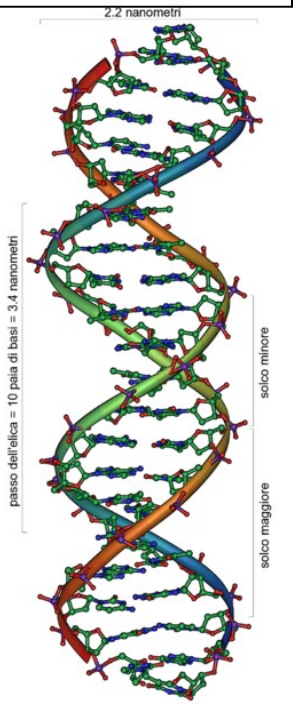
La sequenza di DNA (o RNA) rappresenta l'ordine con cui i differenti nucleotidi si succedono nella catena

Le sequenze hanno una direzione, sono generalmente scritte in direzione 5'- 3' (**direzione di sintesi**)

Ogni nuovo nucleotide viene aggiunto al 3' dello zucchero attraverso il gruppo fosfato. Quindi la catena cresce in direzione 5'→3' (facendo riferimento agli atomi di carbonio dello zucchero)

La sequenza in figura viene quindi scritta come segue:

5'-**ACG**-3'

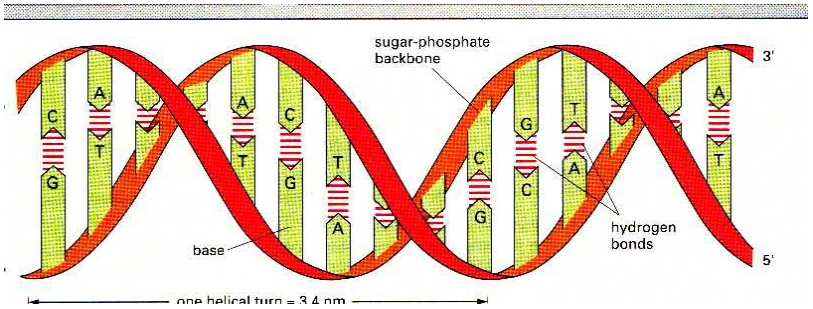


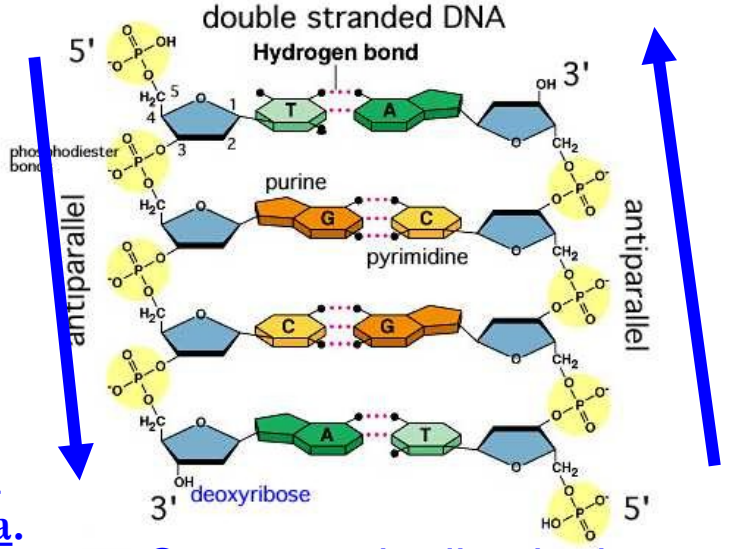
DOPPIA ELICA

Il DNA esiste prevalentemente in forma di doppia elica (2 molecole di DNA appaiate ed avvolte tra loro.

Questa forma rende il DNA chimicamente più 'stabile' e più facilmente 'compattabile' nelle cellule

Nel DNA la doppia elica si forma per l'appaiamento delle basi (mediante legami idrogeno) dei due differenti filamenti
In particolare:
-Una adenina si appaia con una timina (appaiamento A-T o T-A)
-Una citosina si appaia con una guanina (appaiamento C-G o G-C)
Le basi che si appaiano tra loro si dicono **complementari** (C è complementare a G e viceversa, ecc.)





Nota: Se si conosce la sequenza di un'elica si può ricavare anche la sequenza dell'altra elica.

Osservare le direzioni

Negli eucarioti, il DNA si dispone all'interno del nucleo in strutture chiamate **cromosomi.**

Nell'uomo ci sono 23 coppie di cromosomi, di cui 22 sono cromosomi somatici non sessuali (autosomi) mentre una coppia di cromosomi risultano diversi: i cromosomi sessuali (eterosomi).

Direzionalità e rappresentazione delle molecole di DNA

Una rappresentazione in formato testo di una doppia elica di DNA

```
5'    nnCGATGCTAGTAGTTGTACGCAnn -> 3'OH
        ||||||||||||||||||||
HO-3'<- nnGCTACGATCATCAACATGCGTnn - 5'
```

Normalmente è sufficiente rappresentare una sequenza di DNA con una sola stringa (per la complementarietà, se si conosce la sequenza di un'elica si può ricavare anche la sequenza dell'altra elica)

5'-CGATGCCACCAAGTTGTACGCA-> 3'OH

Esempi:

Sequenza d'origine	**5' CGATGCTAGTAGTTGTACGCA 3'**
Sequenza complementare	**3' GCTACGATCATCAACATGCGT 5'**
Sequenza inversa-complementare	**5' TGCGTACAACTACTAGCATCG 3'**

http://promix.cribi.unipd.it/cgi-bin/promix/promix_menu.pl --> **Sequence Utility**

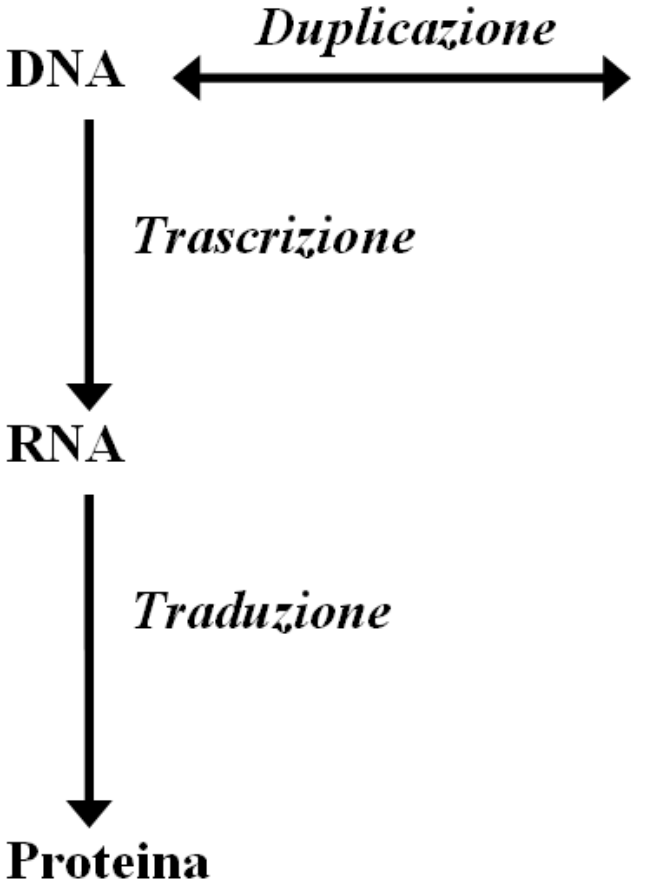
Quando la sequenza indica una singola elica lunga pochi nucleotidi, viene detta **OLIGONUCLEOTIDE**

Il DNA può essere costituito da catene lunghissime di basi, anche centinaia di milioni!!! (nell'uomo circa 3.3 miliardi di basi contenute in 46 cromosomi)

Genoma: indica l'intero patrimonio genetico (DNA) di un organismo vivente che si trova in tutte le sue cellule o in particolari organelli (mitocondri e cloroplasti).

Genomica: la scienza che studia, definisce e caratterizza il corredo genico nel suo complesso

DUPLICAZIONE, TRASCRIZIONE E TRADUZIONE



- Il processo di **DUPLICAZIONE** porta alla formazione di copie delle molecole di DNA ed al trasferimento del materiale genetico.

- Il processo di **TRASCRIZIONE** è il trasferimento dell'informazione dal DNA alle molecole di RNA.

- La **TRADUZIONE** è il processo mediante il quale dall'RNA si passa alla sintesi delle proteine

TRASCRIZIONE (Transcription)

Processo nel quale l'RNA è sintetizzato a partire dal DNA stampo.

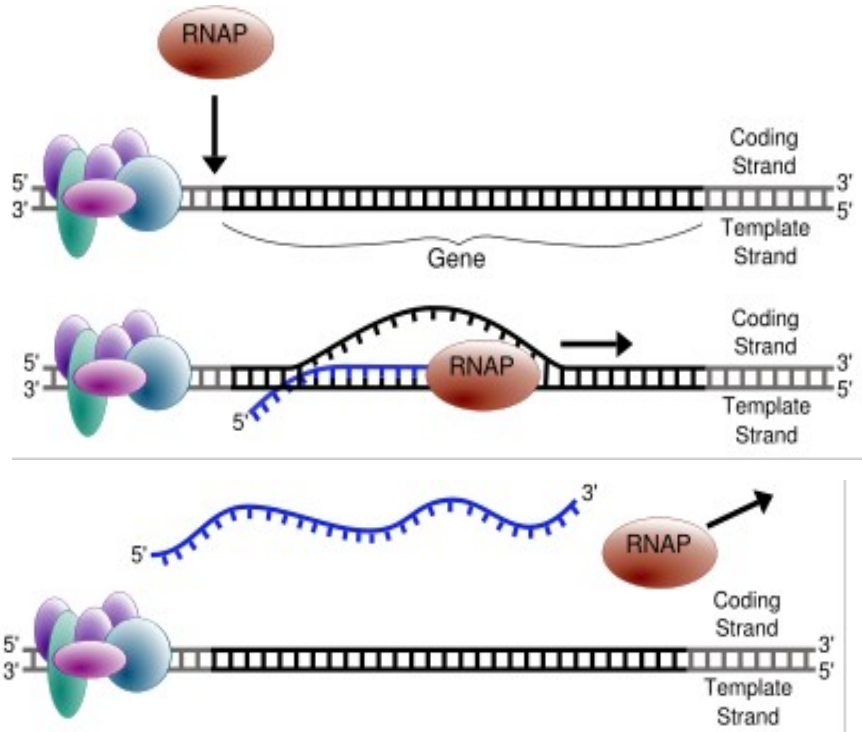
Con la trascrizione si ottengono molecole di RNA con differenti funzioni (es. intervengono nella sintesi proteica, nel silenziamento di geni, nella regolazione della trascrizione stessa):
mRNA (RNA messaggero), tRNA (RNA transfert), rRNA (RNA ribosomale), miRNA (micro RNA), snRNA (small nuclear RNA), ecc.

In particolare **l'mRNA (RNA messaggero)** rappresenta la molecola in cui è copiato il codice del DNA e che serve come stampo per la sintesi delle proteine

Gene (o cistrone): sequenza del DNA che determina la sequenza aminoacidica di una proteina.

DNA
↓
RNA
↓
Protein

Trascrizione dell'mRNA
(RNAP=RNA polmerasi, enzima che dirige la sintesi)



Junk-DNA? (DNA spazzatura?)

Non tutto il DNA viene trascritto in RNA. Alcune parti del DNA forniscono informazioni su:

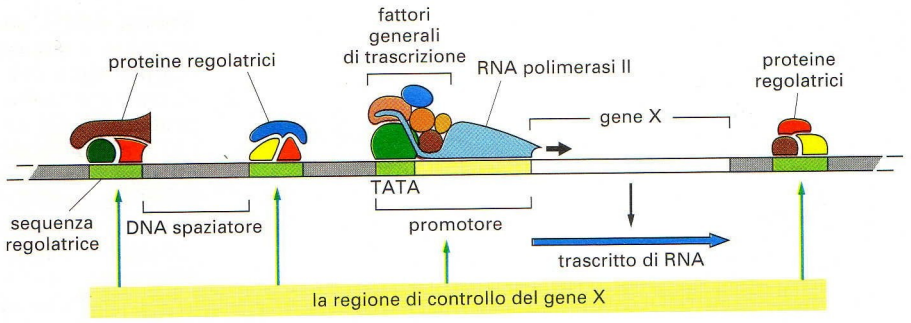
- inizio (segnale d'inizio) e fine (segnale di stop) della trascrizione,
- regolazione della trascizione (nello stesso organismo, non tutti i geni vengono trascritti in tutte le cellule) → **promotori**, **repressori** della trascrizione

Generalmente, particolari sequenze segnale indicano la fine della trascrizione.

Negli eucarioti, un particolare sito provoca il taglio dell'mRNA nascente e innesca una reazione di poli-adenilazione. Generalmente gli mRNA eucarioti finiscono con una sequenza di 'A' ripetute (**poli A**).
Noterete questo particolare in laboratorio quando effettuerete le ricerche di trascritti nei database.
……CTGCGCGAACTGCAAACAAAAAAAAAAAAAAAAAAAA (coda di poliA)

Regolazione della trascrizione (espressione genica)

La trascrizione della maggior parte dei geni è regolata in modo specifico



Dall'RNA alle proteine: la traduzione (translation)

Consideriamo la sequenza lineare di DNA: 5′-ATGATCAGAATCG……3′
Quante basi servono per poter definire 20 aminoacidi:

- 1 base (A, T, G, A, T, C,…….) : solo 4 aminoacidi
- 2 basi (AT, GA, TC, AG,……): 4^2 combinazioni = 16 aminoacidi, non basta!
- 3 basi (ATG, ATC, AGA,…...): 4^3 combinazioni = 64 aminoacidi, anche troppi,

ma è proprio così.

Il codice genetico:
fu decifrato negli anni '60.
Tutti gli organismi hanno essenzialmente lo stesso codice genetico con qualche piccola eccezione in casi molto particolari (ad esempio i mitocondri).
Il codice genetico viene perciò definito universale.

Il codice genetico è letto interpretando tre basi alla volta, senza sovrapposizioni: ogni gruppo di tre basi viene chiamato tripletta o più propriamente **codone**.

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC TAA **Ter** TAG **Ter**	TGT Cys (C) TGC TGA **Ter** TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " **ATG** Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

Dei 64 possibili **codoni**, 61 sono detti **codoni senso**, in quanto specificano degli aminoacidi, gli altri 3 codoni('Ter') indicano la terminazione della sintesi proteica.

Ci sono 61 codoni per 20 aminoacidi; questo comporta che la maggioranza degli aminoacidi è rappresentata da più di un codone:

→ **il codice genetico è detto degenere**

Metionina (Met, M) e triptofano (Trp, W) dispongono ciascuno di un solo codone; rappresentano gli aminoacidi meno abbondanti nelle proteine.

Importante: il codone ATG (AUG nell'RNA), della metionina, è il codone di inizio più comune, specifica l' AA all' N-terminale della catena proteica. Questa conoscenza è importante per individuare sequenze codificanti all'interno di lunghe sequenza di DNA

Sintesi delle proteine

Proteina nascente