

Verona - 18 Maggio 2007

(Seminario integrativo per il corso di  
Modelli Matematici per la Biologia,  
titolare M. Squassina, A.A. 2006/2007)

*Macromolecole come sistemi dinamici:  
applicazioni al DNA*

Antonio Ponno\*<sup>†</sup>

Università degli Studi di Padova  
Dipartimento di Matematica Pura ed Applicata

\*E-mail: [ponno@math.unipd.it](mailto:ponno@math.unipd.it)

<sup>†</sup><http://www.math.unipd.it/~ponno>

- **Struttura e funzione** del DNA
  - trascrizione
  - denaturazione
- **Un modello matematico** *iper*-semplificato
- **Onde solitarie** nel DNA ?
- Problemi aperti

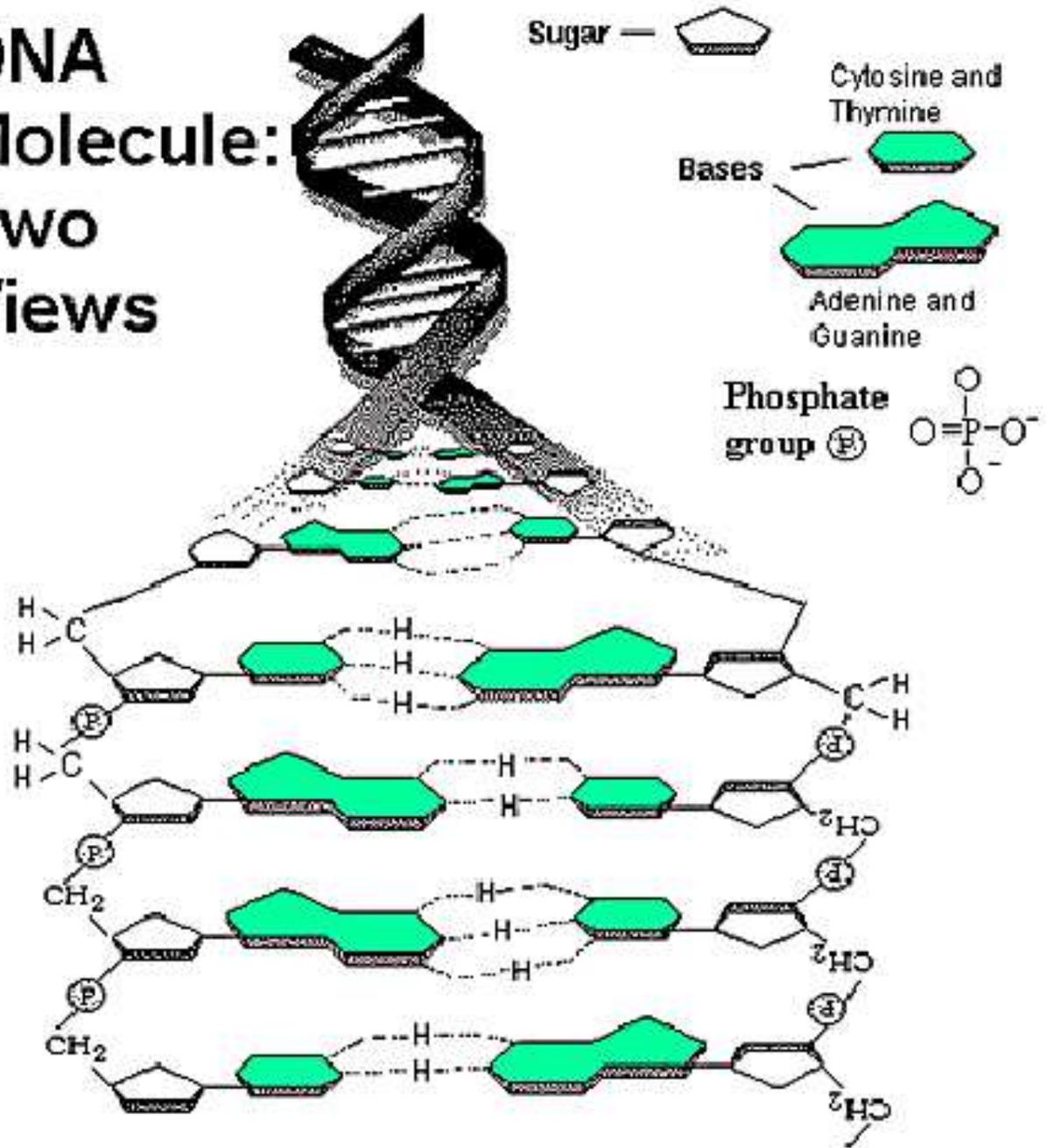
# II DNA

Il DNA (DeoxyriboNucleic-Acid) è una macromolecola organica costituita da due catene polimeriche interagenti tramite legami ad idrogeno (dipolari del tipo H-O e H-N). Una singola molecola di DNA umano contiene circa  $10^8$  unità di base o monomeri, per una massa complessiva dell'ordine di  $10^{10}$  u.m.a..

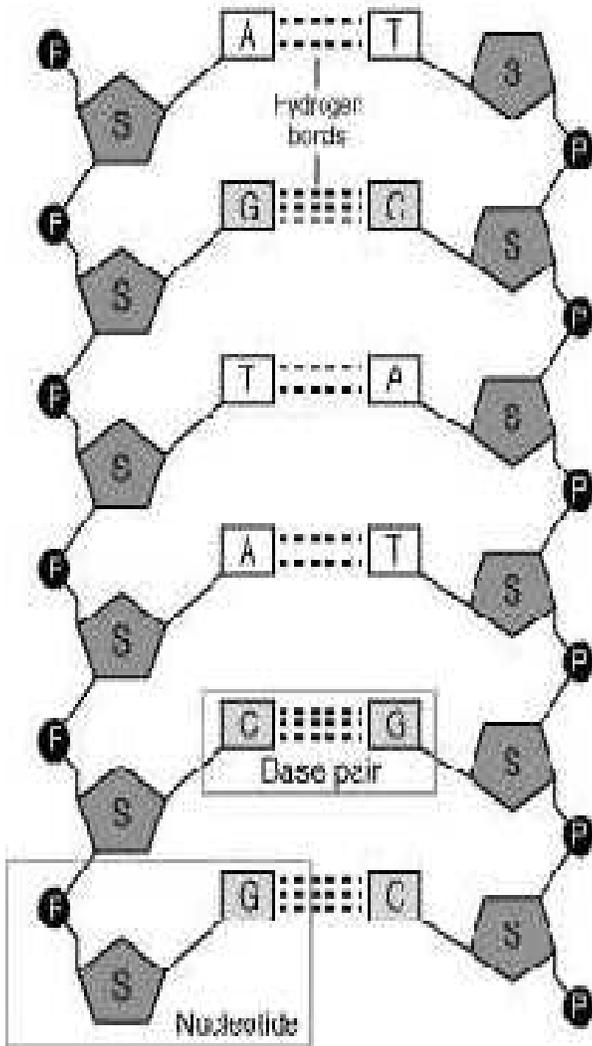
I monomeri costituenti le due catene polimeriche, detti **nucleotidi**, sono costituiti a loro volta da tre unità fondamentali: una molecola di **acido fosforico**, una molecola di **desossiribosio** (zucchero) e una **base azotata** tra quattro possibili: **A**denina, **G**uanina (purine, 2 anelli), **T**imina e **C**itosina (pirimidine, 1 anello).

Le due catene polimeriche sono intrecciate a doppia elica e tenute assieme da legami ad idrogeno tra coppie di basi complementari: **A = T e C ≡ G** (Watson e Crick '53, premio Nobel '62).

# DNA Molecule: Two Views



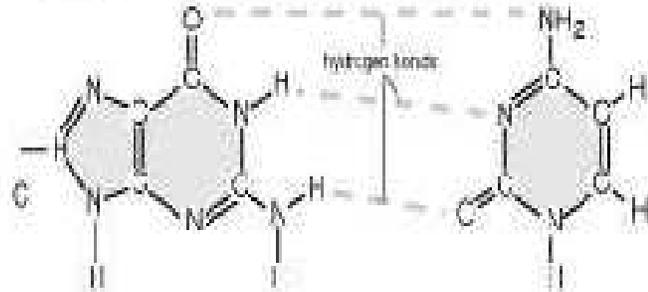
## Deoxyribonucleic Acid (DNA)



## Nitrogenous Bases

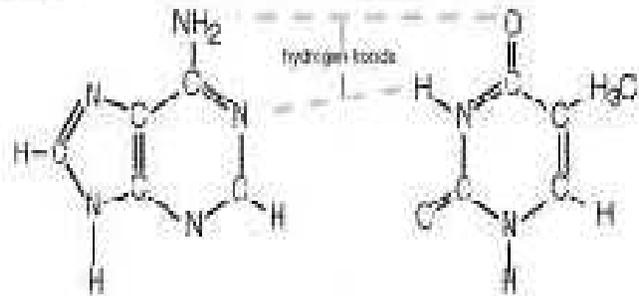
**G** Guanine

**C** Cytosine

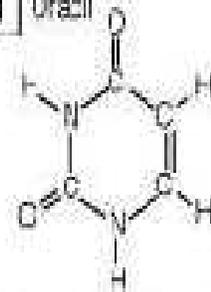


**A** Adenine

**T** Thymine



**U** Uracil



replaces Thymine in RNA

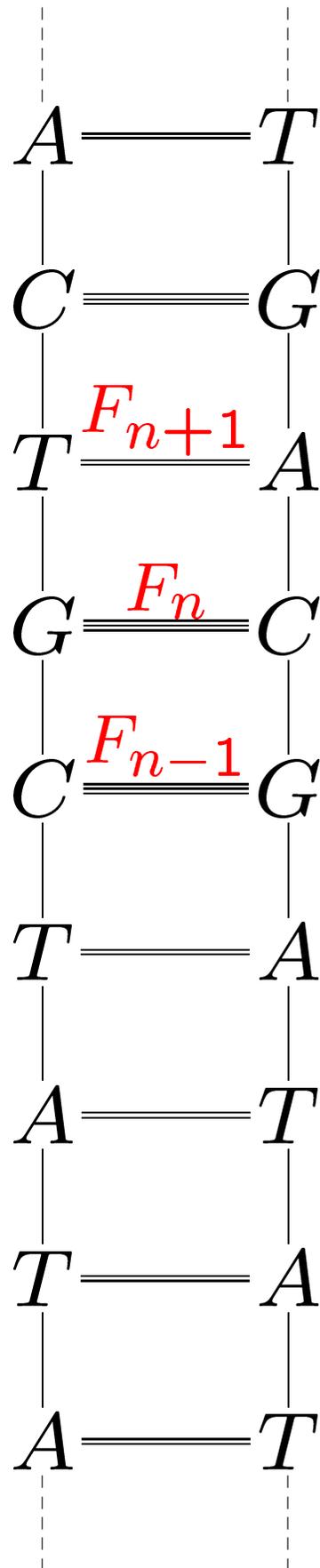


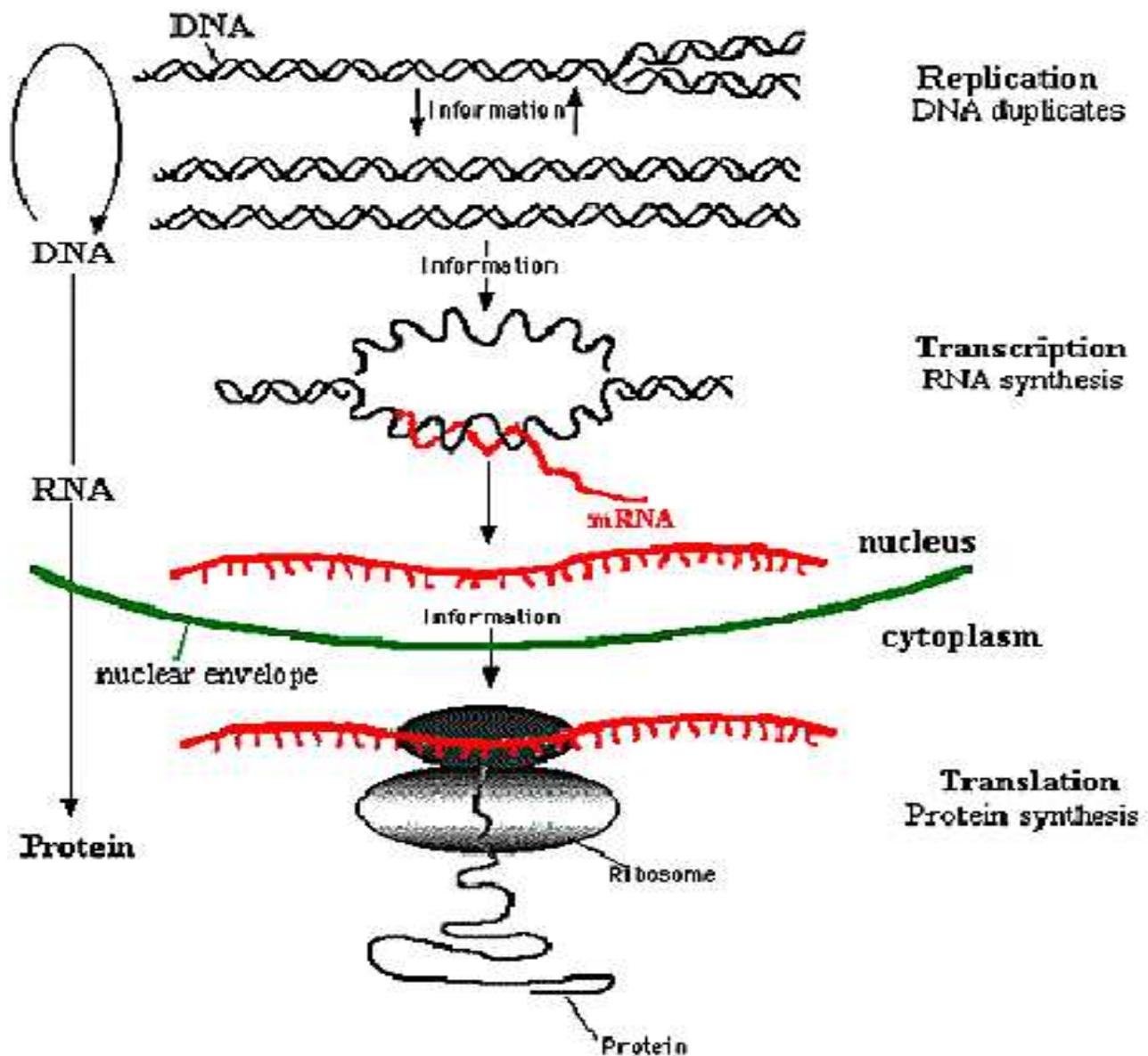
due legami ad idrogeno:  
1 H-O, 1 H-N



tre legami ad idrogeno:  
2 H-O, 1 H-N

la forza  $F_n$  al "sito"  $[n]$   
del legame ad idrogeno  
dipende dalla coppia  
di basi





## The Central Dogma of Molecular Biology

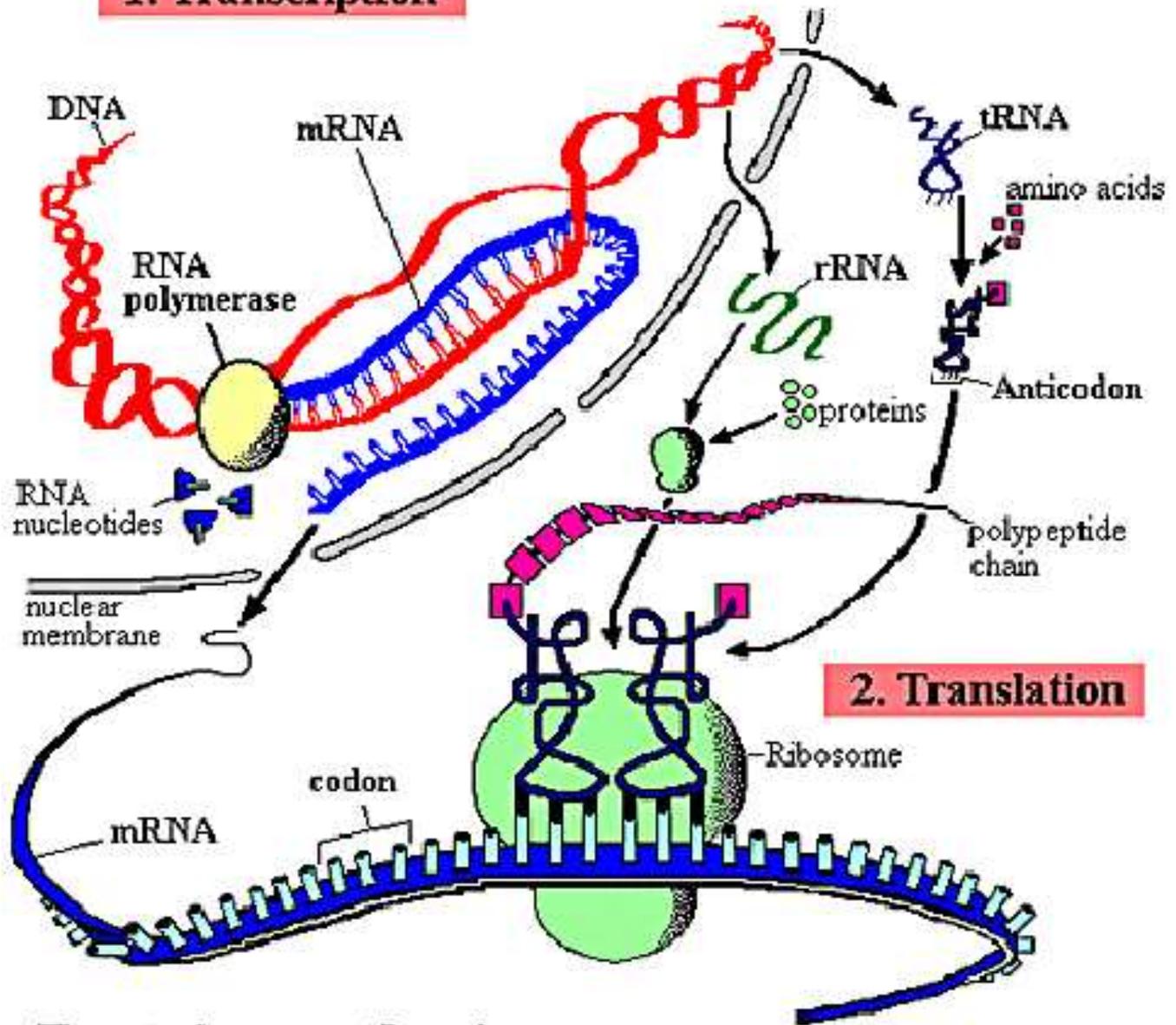
## Codice genetico

Il **patrimonio genetico** individuale è determinato univocamente dalla **sequenza complessiva di basi** nel DNA, e contiene l'istruzione per la **sintesi delle proteine** a livello cellulare.

Ogni dato cromosoma nel nucleo cellulare contiene una catena specifica di DNA (46 cromosomi nell'uomo, 22 in doppia copia piú i sessuali XX nella femmina e XY nel maschio, per un totale di 24 catene di DNA e  $3 \times 10^9$  basi coppie di basi).

I **geni** sono tratti specifici di basi consecutive codificanti la **sintesi di una singola proteina** (il DNA umano contiene, complessivamente, circa 30.000 geni). Ogni tripletta consecutiva di basi, o **codone**, presente nella regione codificante di un gene, codifica la **sintesi di uno specifico aminoacido**, l'unità costituente della catena proteica (gli aminoacidi costituenti le proteine sono 20).

## 1. Transcription



# Protein synthesis

**Il codice genetico è degenere:** le triplette possibili con un alfabeto di quattro basi  $\{A, C, T, G\}$  sono  $4^3 = 64$ , contro 20 aminoacidi. Una proteina "tipica" contiene 3-400 aminoacidi, ed è quindi codificata da un gene contenente *almeno* 3-400 codoni, ovvero circa 1000 basi nucleotidiche. Con 30.000 geni si ottiene una **frazione codificante del patrimonio genetico** pari all'**1%**.

Le regioni "non codificanti" dei geni contengono sia istruzioni quantitative sulla produzione delle proteine sia istruzioni deputate al processo di trascrizione.

**ATTENZIONE:** non codificante NON implica privo di informazione!

# Trascrizione

La sintesi proteica avviene grazie al processo di trascrizione del DNA: una copia conforme (molecola di RNA messaggero) di un intero gene viene fabbricata "in situ" e poi portata fuori dal nucleo fino ai ribosomi, dove la proteina viene materialmente costruita (traduzione).

Durante il processo di trascrizione la doppia elica si apre localmente ed espone il gene alla lettura (il processo è catalizzato da un enzima, l'RNA polimerasi).

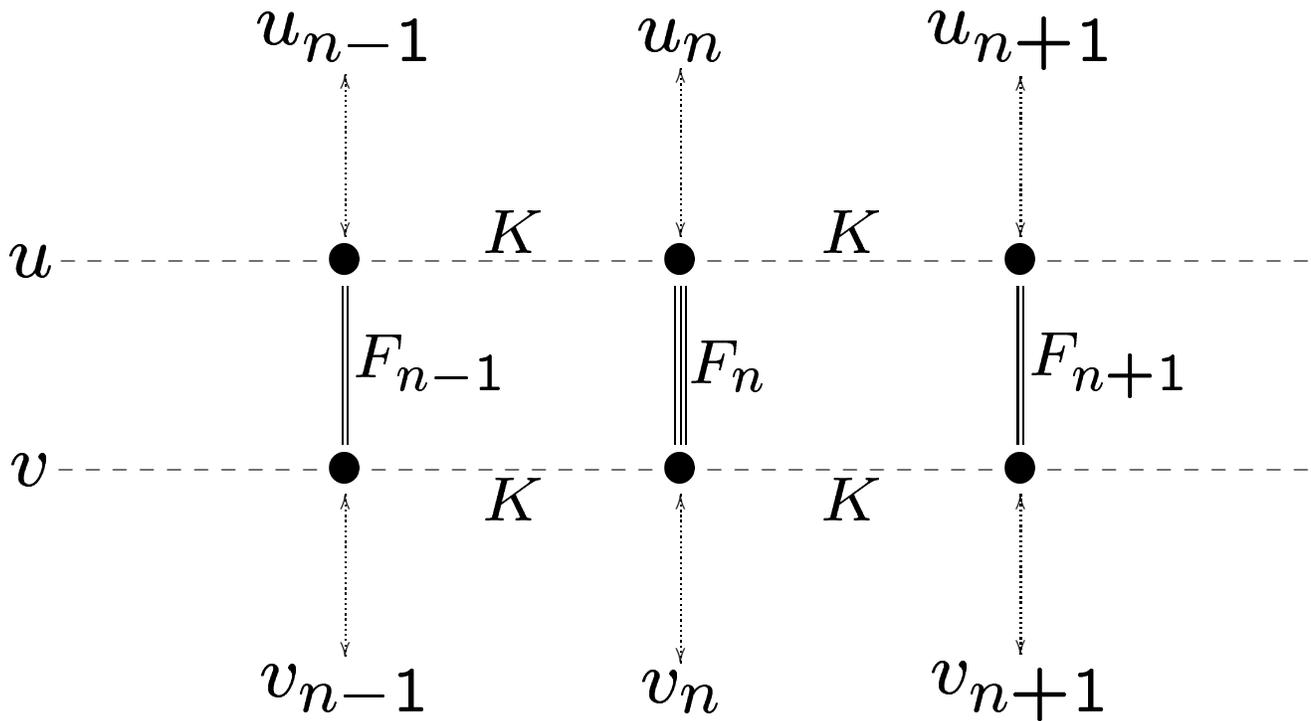
La "bolla" di trascrizione si forma all'inizio del gene, e si propaga fino alla fine di questo, dove si estingue. La larghezza media di una bolla tipica può essere di alcune decine di basi.

# Denaturazione

È l'insieme dei processi che portano all'**apertura irreversibile della doppia elica tramite rottura dei legami ad idrogeno**, con conseguente perdita di funzionalità del DNA e morte cellulare (osservabile a temperature di poco più elevate della temperatura corporea).

Fondamentale: il ruolo delle fluttuazioni dovute al bagno termico con il quale il DNA è a contatto. **Le fluttuazioni termiche locali a temperatura ordinaria potrebbero costituire anche il meccanismo di base per la formazione di bolle in seme**, gonfiate durante il processo di trascrizione.

# Modello matematico



$$M_n^{(u)} \ddot{u}_n = K(u_{n+1} + u_{n-1} - 2u_n) + F_n(u_n - v_n)$$

$$M_n^{(v)} \ddot{v}_n = K(v_{n+1} + v_{n-1} - 2v_n) - F_n(u_n - v_n)$$

con  $n \in \mathbb{Z}$  (catena infinita). La forza intracoppia deve soddisfare  $F_n(0) = 0$ :  $u_n$  e  $v_n$  sono gli scostamenti delle basi complementari rispetto alla configurazione di equilibrio  $\{(u_n, v_n) = (0, 0)\}_n$ .

## SEMPLIFICAZIONI

- $M_n^{(u)} = M_n^{(v)} \equiv M$  indipendente dal sito  $n$
- $F_n(\cdot) = F(\cdot)$  indipendente dal sito  $n$
- "limite al continuo":

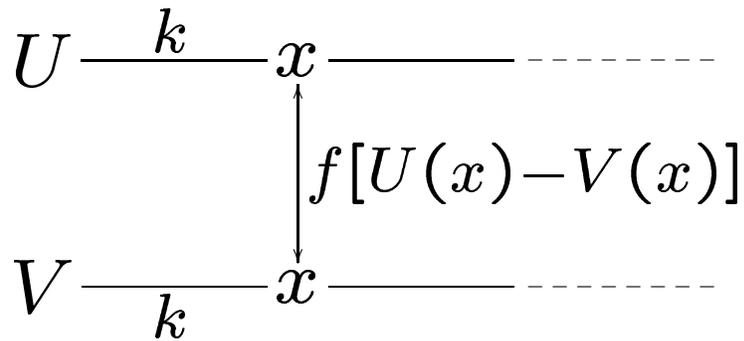
$$u_n = U(x) , v_n = V(x) , x = na$$
$$u_{n+1} + u_{n-1} - 2u_n = a^2 U_{xx} + O(a^4)$$
$$v_{n+1} + v_{n-1} - 2v_n = a^2 V_{xx} + O(a^4)$$

$$\lim_{a \rightarrow 0} K a^2 / M = c_o^2 \quad (\text{finito})$$

$$\lim_{a \rightarrow 0} F(x) / M = f(x) \quad (\text{finito})$$

Con le prime due ipotesi **passiamo a considerare una catena di DNA completamente omogenea**: l'informazione genetica è persa!  
Con il limite al continuo **modellizziamo la catena con due elastici legati trasversalmente da una forza per unità di massa  $f$** .

...in definitiva



Si ha una **coppia di EDP**:

$$U_{tt} = c_o^2 U_{xx} + f(U - V)$$

$$V_{tt} = c_o^2 V_{xx} - f(U - V)$$

per la **coppia di incognite**  $U(x, t)$  e  $V(x, t)$  definite su  $\mathbb{R} \times \mathbb{R} \ni (x, t)$ . Notare che non si è ancora specificata la  $f$ .

Introduciamo il **cambiamento di variabili**

$$R = (U + V)/2 \quad D = U - V$$

$R(x, t)$  è la posizione trasversale del centro di massa locale dei due filamenti;  $D(x, t)$  è lo scostamento trasversale locale dei due elastici rispetto alla distanza di equilibrio.

Si ottiene il sistema equivalente costituito da due equazioni indipendenti:

$$R_{tt} = c_0^2 R_{xx}$$

$$D_{tt} = c_0^2 D_{xx} + 2f(D)$$

$R$  soddisfa l'equazione delle onde standard, la cui *soluzione generale* è (Teorema di d'Alembert)

$$R(x, t) = r(x - c_0 t) + l(x + c_0 t)$$

Le funzioni  $r$  ed  $l$  sono univocamente determinate dal dato iniziale  $R(x, 0)$  e  $R_t(x, 0)$  e descrivono, rispettivamente, un profilo d'onda che si propaga verso destra e un profilo d'onda che si propaga verso sinistra, entrambi aventi velocità  $c_0 > 0$ .

Per quanto riguarda  $D$  ci si limita a cercare soluzioni particolari, aventi la forma di *onde solitarie* o (dizione impropria) solitoni.

# Onde Solitarie

Cerchiamo soluzioni dell'equazione per  $D$  della forma

$$D(x, t) = S(x - ct) \quad , \quad c \in \mathbb{R}$$

Poiché  $S_t = -cS_x$ ,  $S(y)$  ( $y \equiv x - ct$ ) soddisfa l'EDO conservativa

$$(c^2 - c_0^2)S''(y) - 2f(S(y)) = 0$$

Notare che per  $c = \pm c_0$  si ha  $f(S(y)) = 0$ , che ammette in generale un numero finito di soluzioni della forma  $S(y) = \text{costante}$ , tra cui, sicuramente,  $S(y) \equiv 0$ . Per  $c \neq \pm c_0$  si moltiplica l'equazione per  $S'(y)$ , si introduce *un* potenziale  $V(S)$  tale che  $V'(S) = -2f(S)$  e si ottiene la *legge di conservazione dell'energia*

$$H(S, S') \equiv \frac{c^2 - c_0^2}{2} (S')^2 + V(S) = E$$

Dunque si devono studiare gli insiemi di livello della funzione *energia*  $H(S, S')$ : ad ogni valore della costante di integrazione  $E$ , detta *valore dell'energia*, corrisponde un insieme di livello della  $H$ , contenente in generale varie componenti connesse, ad ognuna delle quali corrisponde una soluzione diversa della EDO di partenza.

Notare che se  $c^2 > c_0^2$  l'equazione di conservazione dell'energia è analoga a quella di un punto materiale di massa  $c^2 - c_0^2$  che si muove sulla retta sotto l'azione di una forza conservativa di potenziale  $V$ . Nel caso  $c^2 < c_0^2$  l'analogia meccanica vale ancora ma con massa  $c_0^2 - c^2$ , potenziale  $-V$  e valore dell'energia  $E' = -E$ .

Una scelta *plausibile* per il potenziale  $V(S)$ :

$$V(S) = m^2 S^2 - gS^4$$

(ovvero  $f(S) = -V'(S)/2 = -m^2 S + 2gS^3$ ).

Nel caso  $c^2 > c_0^2$  non ci sono soluzioni spazialmente localizzate. Nel caso  $c^2 < c_0^2$  l'insieme di livello  $\{H = 0\}$  contiene due connessioni omocline che corrispondono a due soluzioni localizzate, una con  $S > 0$  e l'altra con  $S < 0$ . L'equazione  $H = 0$  per  $S$  non identicamente nulla (soluzione banale) è separabile,

$$\sqrt{\frac{c_0^2 - c^2}{2m^2}} \frac{dS}{\sqrt{S^2 - (g/m^2)S^4}} = \pm dy$$

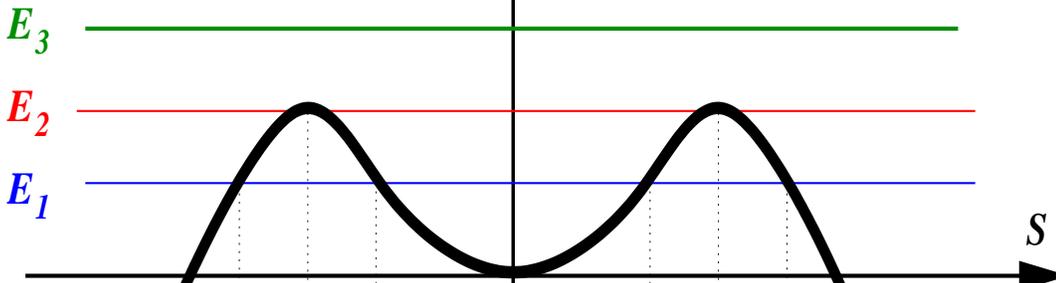
ed ammette le due soluzioni simmetriche

$$S(x - ct) = \pm \frac{m}{\sqrt{g}} \frac{1}{\cosh \left[ \sqrt{\frac{2m^2}{c_0^2 - c^2}} (x - x_0 - ct) \right]}$$

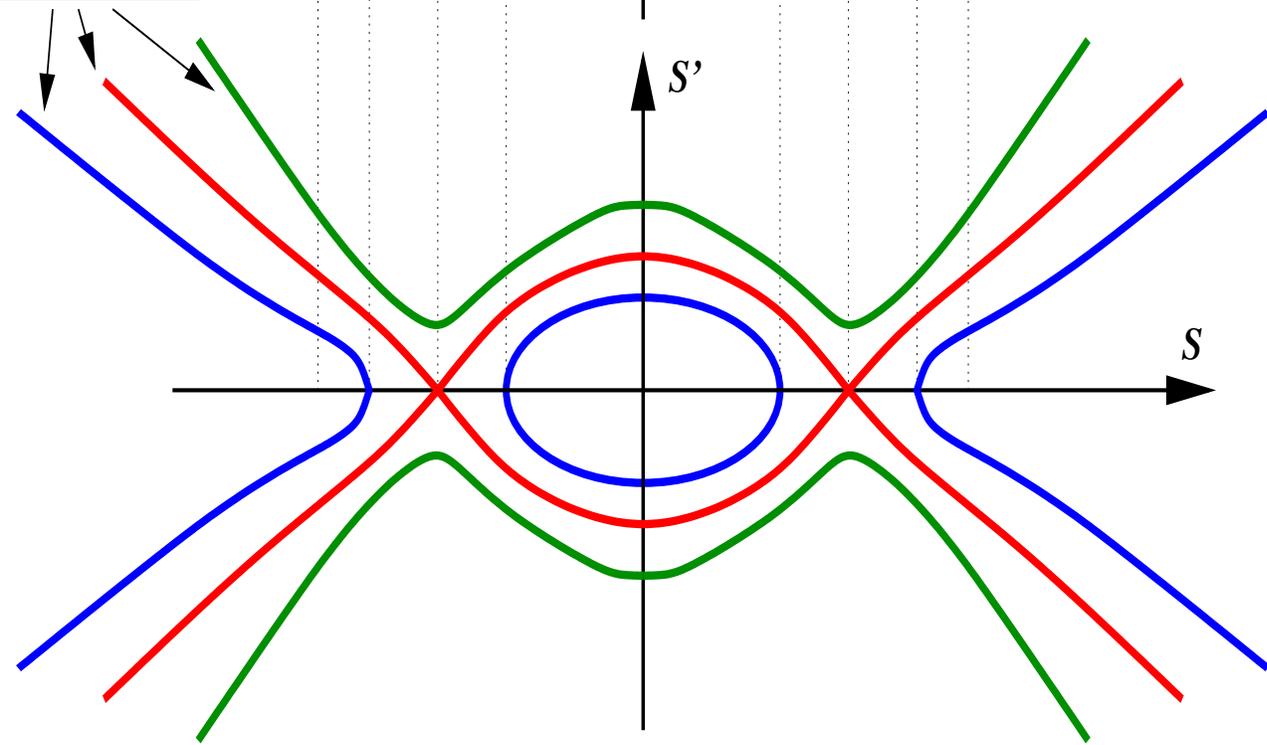
Onde solitarie (profilo indeformato, velocità costante) di questo tipo sono candidate a rappresentare le bolle di trascrizione anche in presenza di disomogeneità strutturali e di temperatura finita.

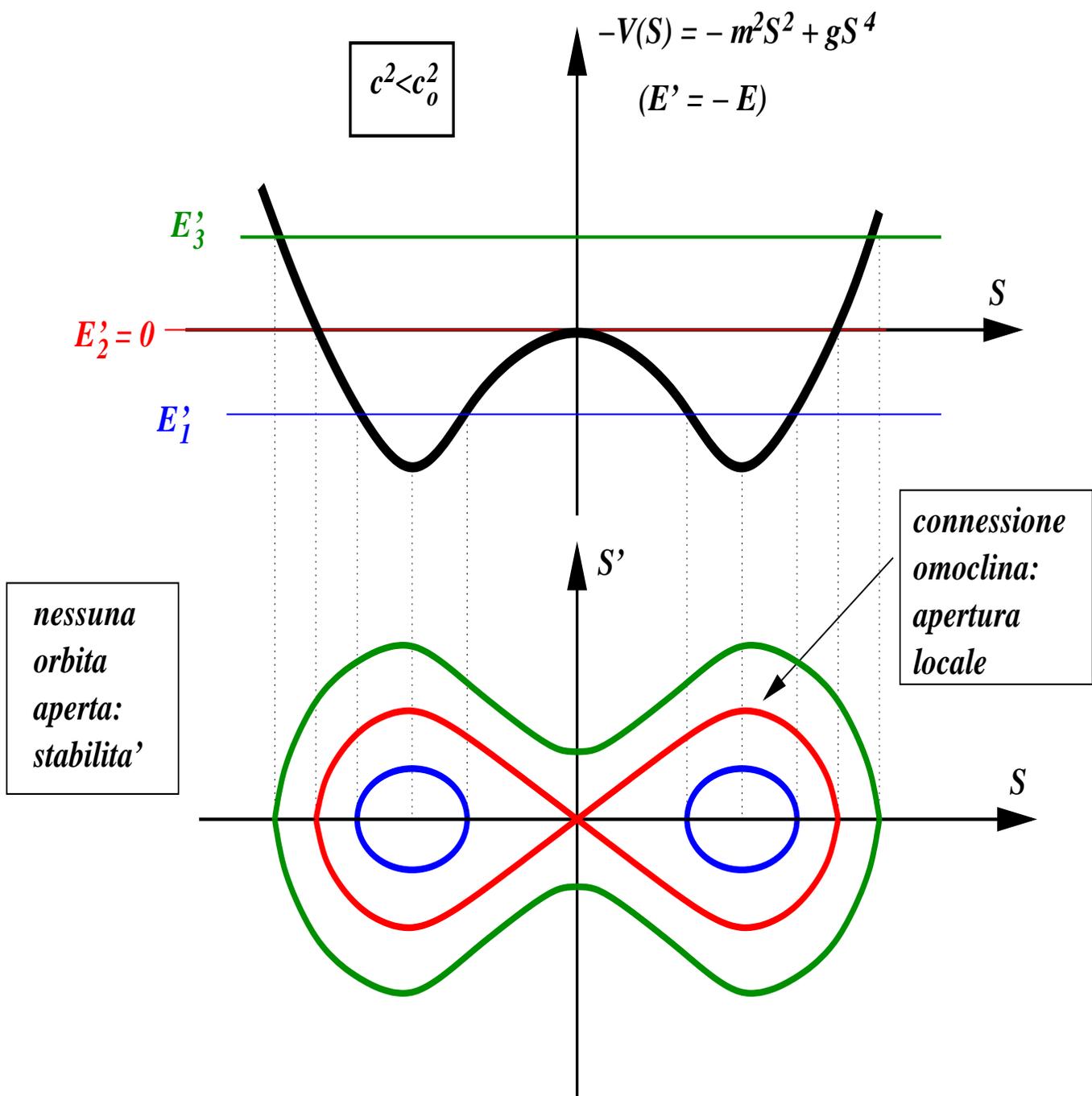
$$c^2 > c_0^2$$

$$V(S) = m^2 S^2 - g S^4$$



*orbite aperte:  
possibili stati  
denaturati*





# Problemi aperti

- Reintrodurre nel modello la **disomogeneità microscopica** dovuta alle masse diverse delle basi ed ai diversi legami ad idrogeno tra coppie di basi complementari
- Tenere conto della **temperatura**: il DNA è a contatto con un bagno termico ad una certa temperatura (37° per l'uomo). Necessario un modello con forze random agenti su ogni nucleotide.
- Introdurre la **geometria**: i due filamenti sono intrecciati a doppia elica e quest'ultima assume una struttura compatta molto complessa

## **BIBLIOGRAFIA**

- Michel Peyrard, Nonlinear dynamics and statistical physics of DNA, Nonlinearity **17** (2001) R1-R40
- Giuseppe Gaeta, Results and limits of the soliton theory of DNA transcription (1998): <http://users.mat.unimi.it/~gaeta>