

MATEMATICA E BIOLOGIA (2)

RACCONTO IN QUATTRO ATTI E UNA PREMESSA

di Giovanni Naldi*

Che nesso c'è tra la biologia e la matematica? Una prima risposta è stata presentata, nel primo atto, sul numero 12 di Emmeciquadro. Là l'autore ha messo a fuoco l'importanza dello strumento matematico nella comprensione di fenomeni biologici come la dinamica di popolazioni, l'evoluzione di un sistema ospite-parassita, la complessa fisiologia dei neuroni. Qui approfondisce il dibattito sull'utilità e applicabilità dei modelli matematici in biologia attraverso esempi relativi al livello molecolare: il protein folding, la TAC e il DNA computing.

Nella prima parte di questo articolo abbiamo introdotto il tema e accennato al caso dei sistemi dinamici. Abbiamo solo «sfiortato» alcuni argomenti mentre altri sono solo stati citati cercando di darne almeno una descrizione intuitiva. Per esempio abbiamo parlato di comportamento caotico senza darne una definizione formale ma solo una vaga idea¹ sperando di avere comunque suscitato la curiosità del lettore. Riprendiamo il racconto dal secondo atto, la scenografia è cambiata e si passa al livello molecolare; anche qualche attore in più fa la sua comparsa.

Atto II: il *protein folding*

Parleremo ora, in modo molto sintetico, di un problema estremamente importante, il *protein folding*: predire quale configurazione tridimensionale (e come viene raggiunta) verrà assunta da una proteina con una determinata struttura definita dalla sequenza di amminoacidi che la compongono. Le proteine (πρωτεΐνη, primo) hanno un nome quanto mai appropriato: infatti, esse rappresentano, per così dire, la sostanza della vita. Sono presenti con funzione di sostegno in gran parte del corpo animale; si trovano come costituenti essenziali in tutte le cellule viventi, compongono la parte principale dei muscoli, della pelle, dei tendini, dei nervi e del sangue, formano gli enzimi, gli anticorpi e numerosi ormoni. Potremmo suddividere le proteine in due categorie: proteine «di struttura» che servono come blocchi di base costitutivi degli organismi, proteine «funzionali» che intervengono come cataliz-

*Professore Associato di Analisi Numerica presso il Dipartimento di Matematica e Applicazioni dell'Università degli Studi di Milano-Bicocca.

L'autore ringrazia Giancarlo Travaglini per i commenti utili per la stesura finale del presente articolo.

¹Uno dei primi indizi di un sistema caotico è la sensibilità ai dati iniziali: traiettorie che partono da punti vicini tendono ad allontanarsi in modo esponenziale con il tempo. Questo significa che non è possibile prevedere con una certa accuratezza lo stato del sistema anche partendo da un'ottima approssimazione del suo stato iniziale. In altre parole non possiamo fare previsioni che per tempi molto brevi. Per una introduzione non eccessivamente tecnica, cfr.: R.L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, 1989.

zatori nei processi biochimici (per esempio enzimi).

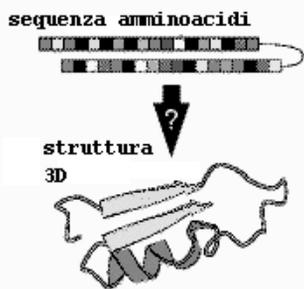
Chimicamente le proteine sono dei grossi polimeri composti da una catena di amminoacidi uniti tra loro da legami peptidici. Ogni molecola di proteina contiene centinaia e anche migliaia di unità di amminoacidi e, dato che gli amminoacidi naturali sono venti, il numero delle diverse combinazioni è enorme. Le proteine si dividono in due grandi classi: le proteine fibrose (insolubili in acqua) e le proteine globulari (solubili in acqua o nelle soluzioni acquose di acidi, basi e sali).

Lo studio della costituzione delle proteine presenta diversi livelli: la struttura primaria, cioè la sequenza di amminoacidi, legati covalentemente con legami peptidici a formare la catena polipeptidica; la struttura secondaria, cioè l'organizzazione locale che può assumere la catena lineare di amminoacidi, mediante interazioni tra amminoacidi vicini tra loro (per esempio queste catene sono sistemate nello spazio formando spirali, lamine o sferoidi compatti); la struttura terziaria: la configurazione tridimensionale della catena polipeptidica che compone la proteina (risulta anche dalla interazione tra amminoacidi lontani nella sequenza lineare).

Il problema del *protein folding* consiste, come abbiamo detto, nella predizione dell'architettura tridimensionale della proteina (struttura terziaria) partendo dalla catena lineare di amminoacidi. Nell'immagine a lato si mostra una schematizzazione grafica del problema.

Per certe proteine vi è anche la struttura quaternaria data dall'associazione di due o più catene polipeptidiche che concorrono a formare la molecola completa: di questo non ci occuperemo. Inoltre è stato dimostrato di recente che alcune grosse proteine, per potersi avvolgere e formare la struttura tridimensionale, hanno bisogno dell'aiuto di altre proteine, le chaperonine (dal francese *chaperon*).

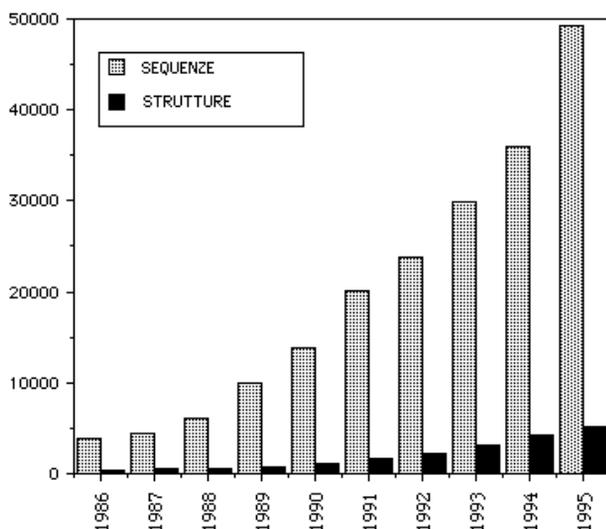
Ma è importante conoscere la configurazione tridimensionale delle proteine? Non solo è importante ma essenziale per capire il funzionamento della proteina stessa. Il premio Nobel 1962 per la chimica venne assegnato a M.F. Perutz² e J.C. Kendrew dell'Università di Cambridge per la loro determinazione della struttura della emoglobina e della mioglobina, una molecola strettamente legata alla prima. Usando la tecnica dei raggi X e conoscendo la sequenza degli amminoacidi, essi determinarono la forma tridimensionale di queste molecole estremamente complicate: in modo «più preciso» per la mioglobina e con buona approssimazione per l'emoglobina. L'emoglobina è detta anche «la molecola che respira» e rende possibile il trasporto dell'ossigeno ai tessuti. Questa proteina cattura, in modo reversibile, l'ossigeno per trasportarlo dal polmone ai tessuti. Inversamente l'emoglobina permette il trasporto dell'anidride carbonica dai tessuti al polmone. In un certo senso, con una immagine molto rozza, la proteina si apre, lega l'ossigeno, si richiude e viene trasportata. Senza la sua struttura tridimensionale non potrebbe svolgere tale compito di legame, trasporto e rilascio dell'ossigeno.



Una assunzione fondamentale del *protein folding* consiste nel fatto che la sequenza di amminoacidi determina completamente la struttura tridimensionale della proteina. Alla fine degli anni Cinquanta, ai *National Institutes of Health*, Christian B. Anfinsen e collaboratori, studiando il ripiegamento di alcune proteine in vitro, mostrò che molte proteine si possono spontaneamente ripiegare (in vitro) dopo essere state completamente denaturate, concludendo che la struttura tridimensionale deve essere completamente determinata dalla struttura primaria: non era necessario chiamare in causa alcun fattore esterno. Anfinsen³ dedusse anche che la stabilità termodinamica della struttura ripiegata correttamente rappresenta la guida del processo di *folding*. La conformazione più stabile di una proteina è detta stato nativo. Disfunzioni nella struttura tridimensionale di una proteina possono trasformare la proteina stessa da innocua a possibile fonte di malattie.⁴

Nel grafico riportato a lato si mostra l'andamento della conoscenza delle strutture primarie (denominate sequenze) e strutture terziarie (denominate strutture). I grafici si riferiscono a due banche dati di sequenze e di strutture (*Protein Data Bank*). Il grafico evidenzia la differenza tra numero di sequenze e strutture note.⁵ La metodologia per la determinazione della sequenza amminoacidica delle catene polipeptidiche si deve essenzialmente al lavoro di Frederick Sanger sull'ormone insulina. Per l'analisi della conformazione tridimensionale si utilizza la tecnica dei raggi X e la risonanza magnetica nucleare (NMR): il primo approccio è limitato alla possibilità di cristallizzare la proteina stessa e di studiarla quindi in un reticolo cristallino, il secondo approccio studia la proteina in soluzione ed è limitato soprattutto dalla dimensione della proteina stessa.

È evidente che mentre il sequenziamento delle proteine procede in modo rapido (le informazioni sono messe a disposizione della comunità scientifica attraverso banche dati) le conoscenze sulle strutture terziarie corrispondenti non hanno il medesimo incremento. Di qui la necessità di poter studiare algoritmi efficaci per predire la struttura tridimensionale di una proteina partendo dalla struttura primaria. In sostanza una via per il *protein folding* consiste nel prevedere il processo di avvolgimento delle proteine tramite modelli matematici e con l'utilizzo del computer. Gli organismi viventi sono capaci di sintetizzare proteine dotate di conformazione determinata e tuttavia, spesso, metastabili (questa espressione indica strutture che non si trovano nello stato più stabile possibile). Per azione del calore o di vari rea-



²Cfr.: *Agli albori della biologia molecolare, Intervista a M. Perutz*, in: *Emmeciquadro* n. 11, aprile 2001.

³«While many other substances may assist in the folding process...all genetic information required for proper folding is contained within the amino acid sequence.» C. Anfinsen (1964).

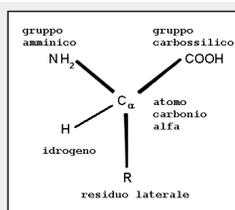
⁴Pare che le malattie da prioni siano dovute alla variazione del ripiegamento di alcune proteine che da innocue diventano infettive.

⁵Grafico ripreso da informazioni *on line* di M.Scarselli, Università di Siena.

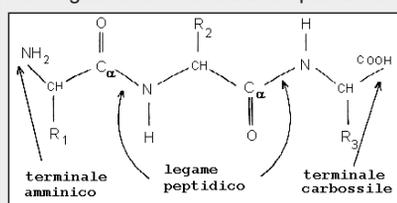
genti o se poste in ambiente fortemente acido o basico, le proteine perdono spesso la conformazione ottimale dal punto di vista biologico e si trasformano in strutture prive di ordine e orientamento, incapaci di estrinsecare la normale attività biologica. Si tratta di un processo, noto come denaturazione, che si verifica, per esempio, nell'albume d'uovo che muta aspetto e struttura durante il riscaldamento e nella carne durante la cottura. In quest'ultimo caso la cottura è alla base della maggiore digeribilità della carne, perché proprio la denaturazione espone all'attacco degli enzimi del tubo digerente quei legami proteici che nel materiale originario sono invece difficilmente accessibili. Anche i procedimenti di «messa in piega» dei capelli sono un esempio di interferenza con la struttura secondaria e terziaria delle proteine.

Come è fatta una proteina?

È composta da una sequenza di amminoacidi. Tutti i 20 amminoacidi presenti nelle proteine contengono un gruppo amminico e un gruppo carbossilico legati allo stesso atomo di carbonio (il carbonio α). Essi differiscono l'uno dall'altro per la catena laterale (residuo) o gruppo R che ha dimensioni, struttura e carica diversa. L'immagine a lato mostra la tipica struttura di un amminoacido.



Due molecole di amminoacidi possono unirsi mediante un legame covalente, per condensazione di un gruppo carbossilico di uno con il gruppo amminico dell'altro: per eliminazione di una molecola di acqua il carbonio del primo gruppo si lega all'azoto del secondo gruppo come rappresentato nell'immagine a lato. Il legame peptidico limita la configurazione della proteina perché gli atomi coinvolti da questo legame sono rigidamente vincolati in un piano. Come può avvenire l'avvolgimento? Soprattutto (ma non solo) per la rotazione

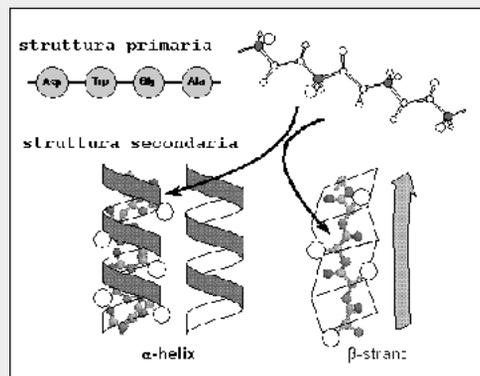


intorno agli assi che collegano l'atomo di carbonio α centrale con l'azoto amminico e, rispettivamente, con il carbonio carbossilico (come mostrato

nell'immagine a destra): gli angoli corrispondenti a tali rotazioni sono solitamente indicati con ϕ e ψ . Questi angoli possono assumere valori nell'intervallo $[-\pi, \pi]$ anche se certi valori sono più probabili di altri mentre alcuni non

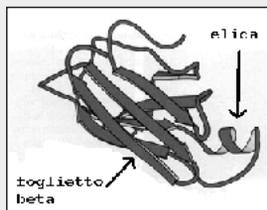
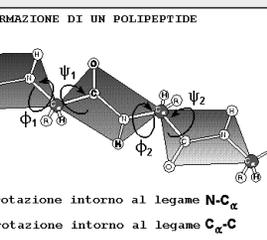
sono permessi. I valori possibili degli angoli ϕ e ψ sono stati determinati per la prima volta da Ramachandran e collaboratori (i valori permessi degli angoli per una proteina sono rappresentati in un diagramma detto diagramma di Ramachandran in cui ogni punto rappresenta una coppia di angoli ϕ, ψ).

La struttura tridimensionale della proteina nativa (che ha raggiunto lo stato finale biologicamente attivo) può essere localmente organizzata in strutture secondarie. Gli elementi della struttura secondaria rientrano in tre famiglie principali:



eliche (α -elica), filamenti e foglietti beta, ripiegamenti che collegano eliche e foglietti. Nell'immagine a lato sono schematizzate le due conformazioni principali della struttura secondaria di una proteina. Le strutture secondarie rappresentano i «mattoni» fondamentali con cui assemblare la struttura terziaria finale che potrebbe essere composta solo da α -eliche o solo da foglietti beta (β -strand) o dalla combinazione di eliche e foglietti.

Vi sono vari modi per rappresentare la struttura terziaria; un esempio (virtuale) è mostrato nell'immagine a lato dove i foglietti sono delle frecce, le eliche degli avvolgimenti e i collegamenti come dei fili spessi.



Diversi approcci al problema

Varie sono le questioni sul *protein folding*. Attraverso quale percorso la proteina arriva alla sua conformazione nativa ripiegata? Perché una particolare sequenza di amminoacidi determina una certa struttura tridimensionale piuttosto che un'altra? Come predire il passaggio dalla struttura primaria a quella terziaria?

Se conosciamo la struttura di una proteina P (per esempio tramite la cristallografia a raggi X) e una proteina Q è «molto simile» a questa a livello di sequenza di amminoacidi, sembra ragionevole supporre che la struttura di Q sia simile alla struttura di P. Sfortunatamente questo non è vero in generale: per esempio sono state sintetizzate proteine con sequenze simili (più del 50%) ma topologia strutturale molto diversa. Anche il contrario non è vero: vi sono proteine con strutture terziarie che si assomigliano ma che derivano da sequenze primarie con non più del 20%-30% di amminoacidi in comune. In ogni caso è stato anche osservato che proteine con sequenze anche molto diverse presentano alcune caratteristiche strutturali comuni.

Queste osservazioni hanno suggerito ai ricercatori, invece di predire la struttura terziaria dalla primaria, di cercare di far corrispondere a una certa struttura (nota) una sequenza di amminoacidi che possa produrla (*inverse protein folding*). La disponibilità di banche dati sempre più ricche e aggiornate ha incoraggiato la ricerca di algoritmi per la predizione delle strutture terziarie partendo da sequenze primarie omologhe: occorre naturalmente una definizione di sequenze omologhe e di struttura (problemi di questo tipo si chiamano di *protein threading*). Questo approccio parte dalla considerazione che molti *fold* di proteine sono simili, e si assume che vi sia un limitato numero di *fold*. Per una sequenza proteica si cerca allora la struttura che meglio si accorda in base alla minimizzazione di una certa funzione potenziale (usualmente costruita in termini statistici).

Riassumendo, tra gli approcci al problema del *protein folding* possiamo trovare: la minimizzazione di una certa funzione energia calcolata per una certa configurazione (spesso si utilizzano modelli semplificati per la proteina abbinati ad algoritmi di ottimizzazione combinatoria); la predizione delle strutture secondarie che compongono la proteina; metodi che utilizzano proteine omologhe; metodi di dinamica molecolare (esame di strutture di molecole dal punto di vista energetico e dinamico); metodi basati sulla «conoscenza» (per esempio *protein threading*).

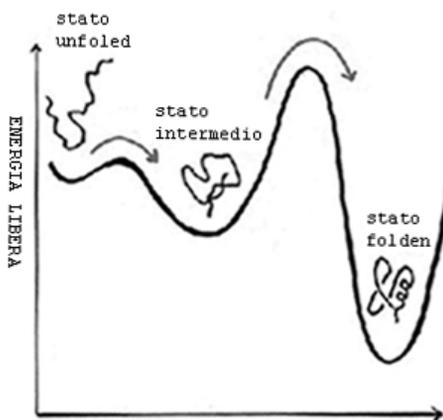
La funzione energia

Una delle possibili vie per arrivare alla conformazione nativa consiste nel ricercare il minimo globale di una «funzione energia libera» associata alla proteina. La funzione energia deve tener conto delle varie

forze in gioco: forze elastiche di legame, forze di torsione, forza di van der Waals, forze elettrostatiche eccetera. La ricerca di un minimo globale è un problema numerico estremamente più complicato rispetto alla ricerca di un minimo locale (dovrebbe essere intuitivo). Algoritmi sia di natura stocastica sia deterministici sono utilizzati, la complessità del problema rimane tremenda (le proteine possono avere normalmente da 100 a 500 amminoacidi e da 1000 a 5000 atomi). Le simulazioni considerano il *fold* della proteina in un solvente (e quindi occorre considerare anche l'interazione proteina/solvente).

Un recente risultato computazionale di questo tipo è stato ottenuto da Duan e Kollman nel 1998 che hanno seguito il *fold* di una piccola proteina per 1ms (ben poco rispetto all'intero processo). Il tempo impiegato è stato di 4 mesi per 256 processori (la simulazione ha mostrato stati intermedi nel cammino di avvolgimento). Questo lavoro è indicativo di quanto siamo lontani dalla simulazione della dinamica di una proteina: la maggior parte delle proteine sono più complesse di quella utilizzata e dovrebbe richiedere un tempo macchina di 10^3 - 10^6

milioni di volte superiore. Ma una proteina in natura esplora tutte le conformazioni possibili fino a trovare quella più favorevole? Sembrerebbe di no. Supponiamo, per esempio, che una proteina sia composta da 100 amminoacidi (una proteina piccola) e ogni singolo amminoacido possa assumere 10 configurazioni differenti: ci sarebbero 10^{100} configurazioni possibili (un numero superiore al numero di atomi che si ipotizza siano nell'intero Universo). Questa considerazione, nota come «paradosso di Levinthal», ci spinge a immaginare che debba avvenire qualche forma di campionamento delle conformazioni dato che le proteine si avvolgono in pochi secondi. Nell'immagine a lato si illustra l'idea intuitiva della ricerca della conformazione nativa come minimo della funzione energia.



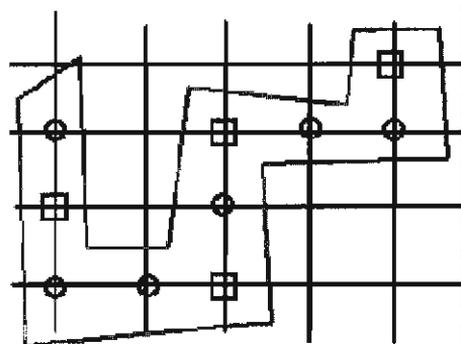
Minimizzare la funzione energia in modelli semplificati

Per capire meglio il meccanismo del *protein folding* sono stati introdotti alcuni modelli semplificati: pur non avendo la pretesa di trovare la configurazione nativa servono per studiare la dinamica e la termodinamica del processo. Una classe importante di questi modelli semplificati consiste nei *lattice model*. In questi modelli si rappresentano tutti i residui con la stessa lunghezza, la lunghezza dei legami è uniforme e i singoli amminoacidi possono essere posizionati nei nodi di una griglia regolare. Un esempio significativo di *lattice model* è il modello HP. Per semplicità consideriamo il caso di un avvolgimento che avvenga in un piano e la griglia sia ottenuta con nodi con coordinate intere (un foglio a quadretti con quadretti di lato unitario). In questo modello i 20 amminoacidi sono suddivisi in due categorie: amminoacidi H (idrofobi,

che si mescolano poco con l'acqua e preferiscono associarsi tra loro), e amminoacidi P (polari o idrofili, che interagiscono fortemente nelle interazioni elettrostatiche e tendono ad attrarre le molecole d'acqua). L'alfabeto di 20 lettere è diventato un alfabeto con due sole lettere.

Si ipotizza che la forza predominante nel processo sia la tendenza idrofoba ovvero sia la tendenza dell'acqua e degli amminoacidi non polari a evitarsi⁶ (si ricorda che si simula l'avvolgimento della proteina in soluzione). Si è anche osservato che l'interno delle proteine nello stato nativo è in gran parte privo di acqua e contiene soprattutto amminoacidi idrofobi. I residui polari sono invece situati principalmente sulla superficie a contatto con l'acqua. Questa regola non è però sufficiente a prevedere quali specifici residui (amminoacidi) saranno presenti in una determinata zona.

L'immagine a lato mostra un esempio di *protein fold* (di avvolgimento, di conformazione) per un modello HP bidimensionale rappresentato con una griglia regolare nel piano a coordinate intere. Ricordiamo che: ogni amminoacido può stare in un nodo; amminoacidi distinti occupano nodi distinti; amminoacidi adiacenti occupano nodi adiacenti (cioè a distanza unitaria); una conformazione di avvolgimento corrisponde a un cammino che non si «auto-interseca» (*self-avoiding walk*) nei nodi della griglia con gli amminoacidi. Pertanto, nell'immagine, a ogni pallino corrisponde un amminoacido di tipo P, a ogni quadratino un amminoacido di tipo H (la sequenza rappresentata è PPHPHPHPH).



Esaminiamo ora una funzione energia che consideri solo contatti locali tra nodi adiacenti. Le riflessioni precedenti sono tradotte nei seguenti valori di energia per i contatti tra residui: contatto HH valore -1, contatto HP, PH, PP, valore 0. La funzione energia è quindi:

$$E = \sum_{1 \leq i+1 < j \leq n} E_{i,j} \delta(r_i, r_j)$$

dove n indica il numero di residui (amminoacidi), r_i l' i -esimo residuo della struttura primaria (sequenza di amminoacidi), $E_{i,j}$ indica il valore dell'energia di contatto (è uguale a -1 se r_i e r_j sono di tipo H), $\delta(r_i, r_j)$ individua se i residui sono adiacenti: è uguale a 1 se la distanza euclidea tra r_i e r_j è uguale a 1 e $i \neq j \pm 1$, vale 0 altrimenti. Si noti che tramite la funzione δ si contano solo i contatti HH che non sono covalenti cioè dei contatti HH già presenti nella sequenza dei residui che compongono la proteina e che non vengono cambiati dall'avvolgimento: la somma delle loro energie è costante qualsiasi sia l'avvolgimento mentre per contatti non covalenti dipende da quanti contatti HH si formano nel percorso.

Il problema consiste nel determinare la conformazione che minimizza tale energia (e quindi quella che massimizza il numero di contatti HH non covalenti). Il difetto principale di questo modello consiste

⁶Oltre a quello riguardante la polarità, scelto nell'esempio, si potrebbero ipotizzare altri principi di base, anche se non definitivi, per esempio la tendenza della proteina a riempire lo spazio, a impacchettarsi nell'avvolgimento.

nel fatto che è degenerare: conformazioni differenti possono dar luogo a minimi uguali. Per ovviare a questo, possono essere adottati vari tipi di normalizzazione.

Il modello HP è concettualmente semplice e permette diversi miglioramenti e raffinamenti (per esempio modelli HPNX che comprendono anche le forze elettrostatiche). Pur sembrando computazionalmente più trattabile del problema generale di dinamica molecolare il problema del calcolo della conformazione nativa (che minimizza l'energia) è un problema di tipo NP (*Nondeterministic Polynomial*) per i quali non si conosce un algoritmo efficiente di soluzione. In ogni caso il modello HP, e sue varianti, può essere utile per l'analisi teorica del problema del *protein folding*. Esistono inoltre legami tra il modello HP e la teoria delle catene di Markov in probabilità e questo aiuta nell'analisi del modello. Infine esistono algoritmi approssimanti che, anche se non forniscono la conformazione ottimale, riescono a dare efficientemente la miglior conformazione, entro certi limiti e rispetto a una data tolleranza per la funzione energia, rispetto alla conformazione nativa: per esempio l'algoritmo di Hart-Istrail. Nel processo di minimizzazione dell'energia occorre stabilire le «mosse» utili a cambiare la conformazione e verificare che il loro effetto sia effettivamente quello di diminuire il valore attuale della funzione energia. L'analisi teorica di questi procedimenti è piuttosto complessa e spesso basata su considerazioni empiriche. Molte cose non sono state dette (per esempio come due proteine interagiscono tra loro) ma speriamo il lettore si sia incuriosito e sia pronto per ulteriori letture.

Concludendo vorrei ribadire che il problema del *protein folding* non è un problema delle proteine ma nostro: le proteine in natura si avvolgono e lo fanno efficacemente, siamo noi che non sappiamo come. L'esperimento ideale sarebbe poter seguire con un microscopio ogni particolare atomico e molecolare del processo e fotografare i vari passaggi: purtroppo non è possibile (almeno per ora). Si possono raccogliere indizi, formulare ipotesi e costruire modelli.

Problemi P e NP

Un problema è di tipo P (o appartiene alla classe P) se può essere risolto con un algoritmo il cui «tempo di esecuzione» cresce non più di una certa potenza x^n rispetto al numero x di simboli necessari per codificare l'ingresso. Si ritiene, ed è intuitivo, che problemi in P possano essere risolti efficientemente mentre problemi non in P non possono essere risolti in tempi ragionevoli in modo algoritmico (richiederebbero tempi enormi e impraticabili). Per dimostrare che un problema è in P occorre fornire un algoritmo esplicito che lo risolva in tempo polinomiale.

Un problema è di tipo NP (*Nondeterministic Polynomial*) quando esiste un algoritmo polinomiale che però può lavorare considerando più alternative contemporaneamente. In altre parole un problema appartiene alla classe NP se esiste un algoritmo che impieghi un tempo polinomiale per verificare se una soluzione proposta è effettivamente una soluzione.

Molti problemi nella classe NP sono equivalenti dal punto di vista del tempo di esecuzione. Problemi di questo tipo si dicono NP-completi: risolvere in tempo polinomiale uno di questi significa poter risolvere in tempo polinomiale tutti i problemi della famiglia NP. Ovviamente P è incluso in NP, il problema «P = NP» consiste nel chiedersi se le due classi sono la stessa cosa oppure no. Nel caso in cui P = NP (basterebbe trovare un algoritmo polinomiale per un problema NP completo) molti problemi di grande interesse applicativo potrebbero essere risolti velocemente.

Forse le proteine conoscono il segreto dato che il *protein folding* è un problema (anche nei casi semplificati) NP?

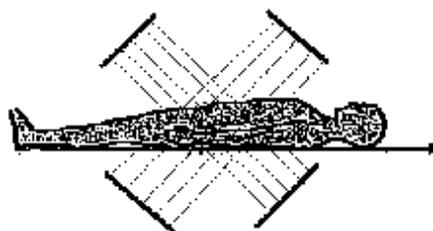
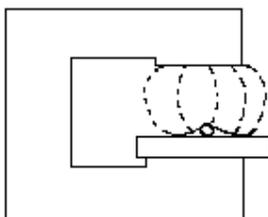
Atto III: immagini biomediche, la tomografia

L'analisi dei segnali e delle immagini utilizzate in medicina è sicuramente un ambito in cui vengono utilizzate idee e tecniche matematiche profonde e all'apparenza molto lontane dalle applicazioni. Nel precedente Atto abbiamo già accennato all'utilizzo della cristallografia a raggi X e alla risonanza magnetica per la determinazione sperimentale della struttura delle proteine.

Per quanto riguarda la Tomografia Assiale Computerizzata (TAC), la parte algoritmica si basa essenzialmente sulla «bella» teoria della trasformata di Radon, ideata nel lontano 1917 dal matematico Johan Radon.

La parola tomografia viene dal greco *τομος* che significa taglio o fetta. Il funzionamento della TAC si basa infatti sull'elaborazione di una sequenza di immagini a raggi X del paziente al

fine di ricostruire una serie di sezioni trasversali che contengono informazioni riguardanti i vari tessuti presenti: ossa, vasi sanguigni, tessuti adiposi, tessuti muscolari, eccetera (immagine a lato).



Contrariamente alle normali radiografie, i raggi X sono confinati su piani. Differenti sezioni, con angoli diversi, sono combinate e permettono una accurata ricostruzione della sezione. Rispetto alla normale radiografia la TAC ha la capacità di distinguere tra regioni con piccole differenze di assorbimento e permette la separazione tra strutture differenti sovrapposte.

Fu il medico statunitense Alan M. Cormack a proporre di impiegare in medicina il principio della tomografia, mentre attorno al 1970 fu l'ingegnere inglese Godfrey N. Hounsfield a costruire la prima macchina per la tomografia computerizzata (il lavoro valse ai due il premio Nobel per la medicina nel 1979). Le prime analisi di Sir Godfrey Hounsfield sono state ottenute con raggi X a densità molto bassa e richiedevano diverse ore di esposizione per ottenere i dati. L'attenuazione di 64 raggi era misurata ottenendo una figura 64x64. Le analisi moderne si realizzano con molti più dati ma richiedono meno tempo di esposizione e meno tempo di elaborazione. In particolare vengono selezionate griglie 148x148 o maggiori. La sorgente della radiazione viene fatta ruotare secondo diversi angoli di incidenza, tipicamente i 148 raggi vengono fatti passare per 180 angoli differenti per un totale di 26 640 raggi passanti. I dati vengono raccolti in sistemi lineari 26 640x26 640. Macchine TAC più accurate possono utilizzare 1024 raggi e 1200 angoli differenti. In altre tecniche si utilizzano metodologie e fenomeni fisici diversi. Per esempio nella PET (*Positron Emission Tomography*), realizzata negli anni Settanta da E. Hoffman, M.E. Phelps, M.M. Ter Pogossian della Washington University, il paziente ingerisce un preparato contenente una sostanza chimica che emette positroni, i positroni annichiliscono con gli elettroni del corpo del paziente producendo raggi gamma che sono rilevati dall'apparato.

Il problema della ricostruzione si situa nell'ambito più generale dei problemi inversi. La nozione di problema inverso sfugge a una definizione avente un certo carattere di generalità.

Solitamente per un particolare processo fisico o un particolare modello matematico è possibile precisare cosa si intende per problema diretto e cosa si intende per problema inverso. Di solito si passa da uno all'altro mediante uno scambio di ruolo tra soluzione e dati, inoltre il problema diretto è solitamente ben posto, quello inverso è mal posto. Uno dei primi esempi fu brevemente discusso da Lord Rayleigh nel 1877 riferendosi al moto di una corda vibrante di densità variabile.

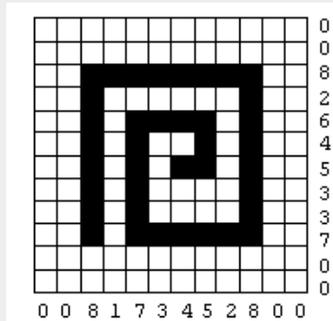
Il problema diretto è allora la determinazione delle frequenze di vibrazione della corda essendo nota la densità (si tratta di un problema di Sturm-Liouville).

Il problema inverso è invece quello di determinare la densità della corda essendo assegnate le sue frequenze di vibrazione. Una generalizzazione di questo problema è stata data da Kac nel 1966 in un articolo divenuto celebre anche per il titolo: *È possibile ascoltare la forma di un tamburo?.*⁷

Una versione giocattolo della TAC

Consideriamo immagini bidimensionali formate da quadrati bianchi e quadrati neri in una griglia. L'immagine è attraversata da ipotetici raggi X che penetrano l'oggetto da sinistra verso destra e dal basso verso l'alto. Viene registrata una successione di numeri, ciascuno indica il numero di quadrati neri che si trovano sulla colonna o sulla riga attraversata.

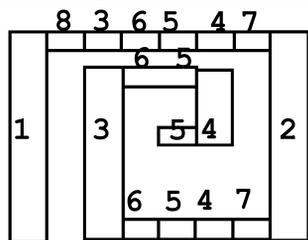
Vediamo un esempio nella figura che segue.



Il problema consiste nel ricostruire l'immagine contenuta nel quadrato dalla conoscenza delle letture per riga e colonna. Ovviamente un problema preliminare consiste nello stabilire se due configurazioni differenti possono presentare le medesime letture. La ricostruzione dell'immagine dipenderà dalla particolare configurazione dei quadrati neri.

Nell'esempio grafico la ricostruzione è possibile. Si osservi che per questa configurazione a spirale esiste almeno una lettura, per esempio la terza colonna con valore 8, che è uguale al numero di letture diverse da zero nell'altra direzione (8 letture diverse da zero per le righe). Questo significa che si ha un quadrato nero per ogni posizione che risulti nella colonna (riga) selezionata, e in tutte le righe (colonne) con lettura diversa da zero; per esempio si ha un quadrato nella terza colonna e in tutte le righe: 3,4,5,6,7,8,9,10. Si può quindi estrarre una parte della figura, nell'esempio la terza colonna. A questo punto si sottrae 1 a tutte le letture interessate dall'estrazione, un quadrato nero corrispondente alla

parte identificata, e si procede alla ricerca di una nuova parte estraibile. Ovviamente potrei avere più colonne (righe) nella medesima condizione, cioè identificabili, per questo occorre decidere (se si vuole predisporre un algoritmo), se si estraggono prima colonne o righe e nel caso di più colonne o righe estraibili quale identificare (la prima da destra, la prima da sinistra, ...).

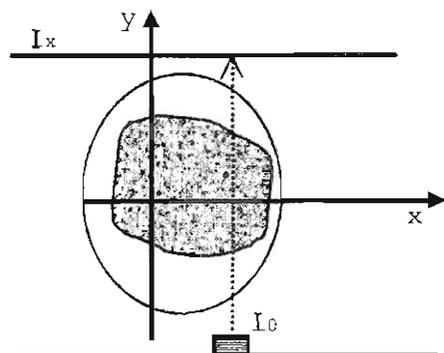


Nella configurazione a spirale disegnata sopra potremmo, per esempio, procedere all'identificazione delle parti rappresentate nell'immagine a lato dando precedenza alle colonne e scegliendo la colonna più a sinistra.

Esistono però configurazioni con la medesima sequenza di letture ma differenti come disposizione di quadrati neri: per queste configurazioni l'algoritmo produrrà una possibile configurazione. Per ovviare a questo si potrebbe pensare di introdurre nuove letture. Per esempio aggiungere alle due precedenti sequenze anche letture fatte «per diagonali» aumentano le informazioni e la possibilità di ricostruzione.

Accenniamo ora al caso continuo limitandoci a un semplice esempio: un corpo nel piano racchiuso in un cerchio di raggio R .

Si suppone inoltre che il corpo sia attraversato da una radiazione di intensità nota I_0 , quando la radiazione attraversa il corpo viene misurata: I_x .



Sia $f(x,y)$ la densità del corpo nel punto (x,y) del piano, le ipotesi di base consistono nel fatto che i raggi X si propagano in linea retta e che l'intensità I si attenui proporzionalmente all'intensità stessa, alla lunghezza del cammino percorso entro il corpo e alla densità di quest'ulti-

mo in base alla legge di Bouger:

$$dI/dy = -f I \quad (\text{LB})$$

Nel nostro semplice caso, risolvendo l'equazione (LB) e indicando con I_0 l'intensità del raggio prima di entrare nel corpo, si ottiene:

$$\log(I/I_0) = \int_{-r(x)}^{r(x)} f(x,y) dy$$

$$\text{dove } r(x) = \sqrt{R^2 - x^2}$$

Considerando una funzione $f=f(r)$, si ottiene la seguente equazione integrale

$$\log(I/I_0) = \int_x^R \frac{2r}{\sqrt{r^2 - x^2}} f(r) dr \quad \text{nell'incognita } f(r).$$

Il problema risulta mal condizionato, intuitivamente: piccole variazioni nei dati possono produrre grandi variazioni nelle soluzioni.

La trasformazione che associa ad una funzione f integrali come quelli descritti sopra è nota come trasformata di Radon. Il problema risulta allora: nota la trasformata di Radon, gli integrali, posso ricostruire la f ?

L'inversione della trasformata di Radon non è un problema semplice e, nelle applicazioni, si utilizzano tecniche numeriche adatte e "raffinate" e, per quanto è possibile, rapide.

Uno degli algoritmi attualmente più utilizzati è noto come metodo *filtered backprojection*.

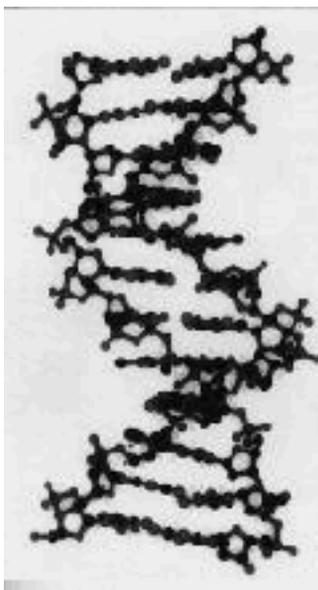
M. Kac, *Can one hear the shape of a drum?*, Am Math. Monthly 73, n.4 Part II, pp. 1-23.

Atto IV: calcolare con il DNA, *DNA Computing*

In questo ultimo atto parleremo dell'idea di base dei «calcolatori a DNA». Anche in questo caso la ricerca si è sviluppata molto e si è arrivati a notevoli modifiche e miglioramenti oltre a qualche critica. Inoltre siamo già passati a «calcolatori molecolari» o «calcolatori cellulari» in cui si implementano circuiti logici attraverso la biochimica di diverse componenti. Non occorre però pensare che sia disponibile un calcolatore «biologico» che possa sostituire il calcolatore elettronico. È comunque affascinante, almeno a mio parere, e interessante ricordare l'idea originale.

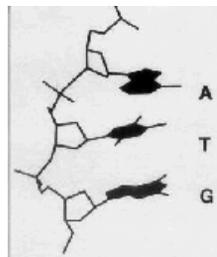
Nel 1994 Leonard Adleman⁸ ha proposto un algoritmo per la soluzione del problema del percorso di Hamilton in un grafo orientato basandosi sulla manipolazione di stringhe di DNA: un problema matematico era stato risolto attraverso tecniche di biologia molecolare. L'idea di base consisteva nell'osservazione che vi erano analogie tra un processo biologico e un algoritmo matematico: le complicate strutture biologiche presenti in natura sono il risultato dell'applicazione di semplici operazioni a partire da informazioni codificate nella sequenza del DNA; il risultato $f(x)$ di una «computazione complessa» per un dato argomento x spesso può essere ottenuto dalla combinazione di semplici funzioni in x . Nell'analisi di algoritmi un utile modello matematico di calcolo è infatti rappresentato dalla macchina di Turing. Una versione di tale macchina consiste in una coppia di nastri e in un meccanismo di controllo a «stati finiti» che muove una testina di lettura lungo il nastro di *input* e in base al suo stato attuale, al carattere letto (si immagina il nastro di *input* o di *output* come una sequenza monodimensionale di caselle ognuna delle quali può contenere un carattere di un opportuno alfabeto), al carattere presente sul nastro di *output*, il controllo stesso cambia stato, muove eventualmente la testina di lettura, scrive sul nastro di *output* e sposta la testina di scrittura. Non bisogna lasciarsi ingannare dalla apparente semplicità della macchina di Turing, in realtà può calcolare tutto quello che calcola il calcolatore elettronico che il lettore ha sulla scrivania. Adleman immaginò che una cosa simile si potesse fare anche per le sequenze di DNA, i nastri di *input* e di *output* sono composti da sequenze delle basi nucleotidiche: adenina, timina, guanina e citosina, abbreviate usualmente con le lettere A, T, G, C (ecco l'alfabeto di *input* e di *output*). In seguito osservò che in natura esiste, in un certo senso, un controllo a stati finiti che permette di leggere sequenze di A,T,G, C dal nastro di *input* e di scrivere contemporaneamente sul nastro di *output* la sequenza complementare secondo le regole trovate da Watson e Crick (da A a T, da T ad A, da G a C e da C a G): si tratta della DNA polimerasi. Per esempio, dalla sequenza CGTACATGC si produce la sequenza complementare GCATGTACG. La DNA polimerasi permette al DNA di riprodursi e, quindi, alle cellule di replicarsi. A questo punto Adleman si chiese se

«[ask] not Mathematics can do for Biology, but what Biology can do for Mathematics» (Stanislaw Ulam, 1976).



era possibile costruire un computer a DNA simile ad una macchina di Turing. La macchina di Turing è un calcolatore universale nel senso che potrebbe essere programmata per calcolare tutto ciò che si presta al calcolo. Potrebbe allora il computer DNA avere la medesima capacità? Che cosa è necessario per il calcolo di una funzione, per una elaborazione?

Riassumendo: è necessario avere a disposizione un modo per memorizzare le informazioni e avere la possibilità di effettuare operazioni (elementari) sui dati memorizzati. Sfortunatamente, mentre esiste un enzima capace di leggere sequenze e generare le complementari, è difficile che esistano enzimi per operazioni aritmetiche o per compiti diversi, per esempio la scomposizione di un intero in fattori primi. Adleman allora si basò sulle operazioni biochimiche disponibili e utilizzate anche nelle applicazioni biotecnologiche.



Complementarietà e struttura a doppia elica

Il DNA codifica l'informazione sotto forma di lunghe sequenze delle quattro basi fondamentali legate chimicamente allo scheletro di sostegno della molecola. Altra caratteristica tipica del DNA è quella di essere costituito da due filamenti complementari, avvolti elicoidalmente l'uno sull'altro. Ciascuna base di un filamento è prospiciente con una base complementare sull'altro filamento. Se un filamento di DNA in soluzione incontra una sequenza complementare, le due eliche si arrotolano fra loro per formare la famosa doppia elica. Le due eliche sono unite non da legami covalenti ma da forze più "deboli" come i legami di idrogeno. Se invece un filamento di DNA in soluzione incontra un filamento non complementare, o con una lunga sequenza non complementare, le due eliche non si appaiano.

Polimerasi

Le polimerasi copiano l'informazione da una molecola a un'altra. La DNA-polimerasi necessita di un «punto di partenza» per iniziare la copia complementare: il segnale è costituito dal *primer* che consiste in un tratto di DNA che si appaia all'elica copia grazie alla complementarietà. Dove si è formata una porzione di doppia elica fra il *primer* e la copia la DNA-polimerasi comincerà ad allungare il tratto *primer* aggiungendo nuove basi e creando la copia complementare.

Le nucleasi

Le nucleasi sono enzimi che permettono di tagliare la molecola di DNA quando incontrano una specifica sequenza. Per esempio c'è un enzima, noto come EcoRI dal nome del batterio *Escherichia coli*, che taglia il DNA immediatamente dopo la base G della sequenza GAATTC. L'azione di questo enzima è molto specializzata ed è estremamente raro che tagli il DNA in sequenze diverse.

Le ligasi

Le ligasi uniscono insieme due o più molecole. Per esempio la DNA-ligasi ha il compito di riparare le interruzioni che si creano nella molecola di DNA. Per fare questo, la DNA-ligasi riconosce due tratti di DNA vicini e li unisce in modo covalente per formare un unico filamento.

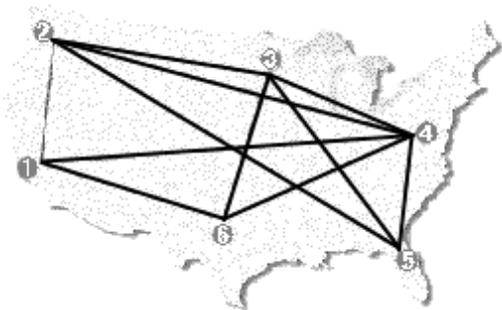
Gel elettroforesi

Questa tecnica, artificiale e non naturale, permette di separare filamenti di DNA in base alla lunghezza. Una soluzione contenente molecole di DNA viene posta in un pozzetto all'estremità di un sottile gel. Al pozzetto viene applicata una differenza di potenziale che produce una corrente elettrica. Le molecole di DNA, che hanno carica negativa, vengono attratte verso l'anodo. I filamenti più corti migrano più velocemente di quelli più lunghi ed è quindi possibile selezionarli.

Oltre alle operazioni brevemente descritte occorre osservare che oggi è possibile anche produrre artificialmente tratti di DNA contenenti sequenze prestabilite. È quindi possibile codificare un particolare problema attraverso l'alfabeto del DNA e memorizzare tale codifica. Vediamo ora come funziona il computer a DNA rifacendosi all'esempio di Adleman: la ricerca del cammino Hamiltoniano. Ribadiamo il fatto che l'algoritmo di Adleman non è implementato su un calcolatore universale a DNA ma si costruiscono esplicitamente dei filamenti di DNA in cui è codificata la risposta al problema specifico.

³L. Adleman, *Molecular Computation of solutions to combinatorial problems*, Science 226, No. 1994, pp. 1021-1024.

Consideriamo un grafo⁸ diretto G con vertici, v_1, v_2, \dots, v_N e fissiamo due vertici v_{in}, v_{out} , il grafo possiede un cammino Hamiltoniano se esiste una sequenza di lati, cioè un percorso, che inizi in v_{in} , finisca in v_{out} e passi in ogni vertice una sola volta.



Il problema può essere esemplificato considerando una mappa geografica in cui le città sono collegate tra loro da voli diretti.

Nell'immagine a lato riportiamo una mappa simile a quella considerata da Adleman.

La domanda è: esiste un percorso che parta da una certa città e finisca in un'altra passando una sola volta per tutte le città della mappa?

L'algoritmo proposto è il seguente:

Passo 1: Generare un insieme di percorsi casuali nel grafo.

Passo 2: Tenere solo i cammini che iniziano da v_{in} e finiscono in v_{out} .

Passo 3: Se il grafo ha N vertici, tenere solo i cammini che hanno esattamente N vertici.

Passo 4: tenere solo i cammini che passano per tutti i vertici.

Passo 5: se rimangono cammini si è trovato un percorso Hamiltoniano.

Il problema fondamentale è: come codificare il problema con il DNA? Che cosa vuol dire città (vertice), collegamento (lato) e che cosa vuol dire percorso? Si tratta quindi di scegliere la codifica del problema. Si abbina a ogni città una sequenza casuale diversa composta da otto⁹ basi fondamentali, per esempio a v_1 si associa ACTTGCAG mentre al vertice v_4 la sequenza TCGGACTG. A ogni collegamento (volo) diretto si associa la sequenza ottenuta prendendo le ultime quattro basi del vertice (città) da dove si parte e le prime quattro del vertice (città) dove si arriva. Con le codifiche precedenti, al lato (v_1, v_4) si associa la sequenza GCAGTCGC. Scelta la codifica si generano le sequenze complementari alle sequenze abbinate ai vertici e le sequenze dei collegamenti (voli). Si noti che questo avviene artificialmente in laboratorio e che si generano moltissimi filamenti con le caratteristiche volute (dell'ordine di 10^{14} - 10^{18} filamenti con la sequenza desiderata). Per completare il *Passo 1* dell'algoritmo occorre mettere in soluzione le sequenze complementari dei vertici con le sequenze dei lati orientati. L'idea è di far incontrare città e percorsi in modo che si leghino per complementarità. Per esempio, il cammino GCAGTCGC, cioè (v_1, v_4) e il nome complementare di v_4 , AGCCTGAC si possono incontrare per caso. Per come sono state costruite le sequenze, le estremità delle due sequenze sono complementari e si appaieranno tra loro. Se poi il primo tratto del codice del volo GCAG incontra il complementare giusto (cioè un altro vertice che arriva in v_1) la sequenza si allunga formando una catena in cui i codici corrispondenti a un lato (volo) sono

collegati e tenuti insieme dai nomi, con codifica complementare, delle città. Le ligasi presenti nella soluzione permettono poi di formare legami permanenti nelle catene di DNA che si sono create. In questo modo si producono molecole di DNA che codificano percorsi casuali attraverso le città. Si noti che tutti i percorsi si formano quasi contemporaneamente dalla interazione simultanea di migliaia di miliardi di molecole: è un buon esempio di elaborazione parallela.

Per il *Passo 2* Adleman si è basato su una catena di polimerasi utilizzando come *primer* le ultime quattro basi del nome di v_{in} e le complementari delle prime quattro di v_{out} . Il primo *primer* obbliga la polimerasi a fabbricare copie complementari che inizino con il giusto vertice, il secondo *primer* invece dà inizio alla duplicazione di sequenze che terminano con il giusto vertice. La catena di polimerasi viene effettuata variando la temperatura della soluzione. Aumentando la temperatura si stimola la DNA-polimerasi a iniziare la duplicazione. Per temperature sufficientemente alte si ha la scissione delle doppie eliche appena formate permettendo la replicazione dei singoli filamenti. Il risultato consiste nel fatto che le molecole con la giusta città di partenza e di arrivo vengono duplicate esponenzialmente mentre le molecole con la sola giusta città di partenza o la sola città di arrivo si duplicano con velocità lineare. Infine le sequenze che non possiedono né la corretta città di partenza, né quella di arrivo non si duplicano. Si crea quindi una popolazione contenente «moltissime copie» di molecole con la corretta partenza e il corretto arrivo.

Per il *Passo 3* si ricorre alla elettroforesi su gel che permette di selezionare solo le molecole con la giusta lunghezza (corrispondenti a cammini che collegano N vertici).

Per il *Passo 4* si sfrutta il principio di complementarità. Si introducono delle molecole di DNA attaccate a minuscole sferette magnetiche e che codificano il nome complementare a quello di un dato vertice. Solo le sequenze che contengono la codifica di quel vertice si legheranno a questa molecola DNA e, tramite un magnete, è possibile selezionarle. Ripetendo più volte questo procedimento, si possono selezionare i cammini che contengono tutti i vertici. Se al termine di tutti i passi rimangono delle molecole di DNA, queste codificano il cammino Hamiltoniano desiderato.

Dopo il lavoro di Adleman l'interesse per questo nuovo approccio è cresciuto enormemente soprattutto nell'ambito della soluzione di difficili problemi combinatori considerati intrattabili (e su cui si basano, per esempio, efficienti sistemi di crittografia). In particolare l'attenzione è rivolta alla soluzione di problemi NP (si veda il riquadro a pagina 30). La strategia consiste nel generare un enorme spazio di candidati a essere soluzione del problema (con un basso consumo energetico) codificato in un modo opportuno. In seguito si seleziona, tramite operazioni biochimiche e opportuni enzimi, una soluzione corretta (sempre che ci sia).

⁹Un grafo è descritto fornendo un insieme di punti (vertici) eventualmente collegati tra di loro da lati.

¹⁰I parametri qui riportati si rifanno all'articolo di Adleman già citato.

La ricerca è appena iniziata e richiede apporti sia dalla biologia molecolare, dalla logica, dalla matematica, dall'informatica in un ambito realmente multidisciplinare.

La ricerca di algoritmi efficienti per la soluzione di vari problemi è di importanza fondamentale per molte situazioni reali. Inoltre le moderne tecniche in crittografia basano la loro sicurezza proprio sul fatto che coinvolgono problemi per i quali non si ha un algoritmo rapido per la ricerca di una soluzione. Il problema dell'utilizzo del metodo originale di Adleman per problemi realistici consiste nel fatto che è stato stimato che sarebbe necessario manipolare qualche etto di DNA. Come dicevamo alternative e modifiche sono analizzate e introdotte. ▽

INDICAZIONI BIBLIOGRAFICHE

- V. Comincioli, *Metodi Numerici e Statistici per le Scienze Applicate*, Ambrosiana, Milano 1992.
 V. Comincioli, *Problemi e Modelli Matematici nelle Scienze Applicate*, Ambrosiana, Milano 1993.
 A. C. Capelo, *Modelli Matematici in Biologia*, Decibel, Padova 1989.
 F.C. Hoppensteadt, C. Peskin, *Mathematics of medicine and life sciences*, Springer-Verlag, New York 1992.
 J.D. Murray, *Mathematical Biology*, Springer-Verlag, New York 1989.
 J. Mazumdar, *An Introduction to Mathematical Physiology and Biology*, Cambridge University Press, 1989.
 D.S. Jones. B.D. Sleeman, *Differential Equations and Mathematical Biology*, George Allen & Unwin, London 1983.
 F.C. Hoppensteadt, *Analysis and Simulation of Chaotic Systems*, Springer Verlag, New York 1993.
 J. Smítal, *On Functions and Functional Equations*, IOP Publishing Ltd 1988.
 C.W. Groetsch, *Inverse problems in the Mathematical Sciences*, Vieweg 1993.
 M. Bertero, P. Boccacci, *Inverse Problems in Imaging*, Institute of Physics Publishing, 1998.
 G. Paun, G. Rozenberg, A. Salomaa, *DNA Computing*, Springer-Verlag Text in Theoretical Computer Science, 1998.

ALCUNE RISORSE WEB

Per il rapporto della *National Science Foundation* (NSF) su *Mathematics and Biology: The Interface challenges and opportunities*, <http://www.gdb.org/Dan/Mathbio/>

Per iniziare a «navigare» nel mondo non lineare: <http://amath.colorado.edu/appm/faculty/jdm/faq-Contents.html>

e per alcune applicazioni il sito della ricerca del *Center for Nonlinear Studies*: <http://cnls.lanl.gov/Frames/research-a>.

Un sito matematico di carattere generale suddiviso per argomenti: <http://archives.math.utk.edu/topics/>

Il sito della *Society for Mathematical Biology*: <http://www.smb.org/>

Sistemi dinamici: <http://www.math.sunysb.edu/dynamics/>

L'articolo di Adleman su *DNA Computing*: http://www.hks.net/~cactus/doc/science/molecule_comp.

Altri link su *DNA Computing*

<http://cominfo.chem.wisc.edu/writings/DNAcomputing>.

<http://www.english.cornell.edu/scitech/w96/DNA>.

Per corsi *on-line* di *Computational Molecular Biology* e *Protein Folding*

<http://cmgm.stanford.edu/biochem218>

<http://pps97.cryst.bbk.ac.uk/index>.

<http://cmb.rutgers.edu/>

<http://www.unisi.it/ricerca/centri/sisbio/Scarsell>

<http://theory.lcs.mit.edu/~bab/99-18.417-home>.

<http://solon.cma.univie.ac.at/~neum/>

The International Society for Computational Biology <http://iscb.org/>

Due argomenti, non trattati in questa breve esposizione, potrebbero essere interessanti e richiedono solo nozioni elementari di matematica: gli automi cellulari e «*Ant Algorithms*».

Per i primi indichiamo qualche sito (tra i numerosi disponibili per iniziare la ricerca):

<http://www.aridolan.com/ad/adbc/CA>.

<http://iinwww.ira.uka.de/ca/>

<http://staff.sk-kari.put.poznan.pl/siwak/>

Per il secondo argomento lasciamo la sorpresa al lettore (si tratta di algoritmi che prendono lo spunto dall'organizzazione sociale di insetti quali le formiche o le api), il sito di uno dei fondatori: <http://iridia.ulb.ac.be/~mdorigo/ACO/ACO>.