Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Camillo Tosi, Raffaella Pavani, Roberto Fusco, Filippo Aluffi-Pentini, Valerio Parisi, Francesco Zirilli

Ricerca di conformazioni di minima energia potenziale intramolecolare mediante un nuovo metodo di minimizzazione globale.

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **78** (1985), n.4, p. 149–156.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1985_8_78_4_149_0>

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

Articolo digitalizzato nel quadro del programma bdim (Biblioteca Digitale Italiana di Matematica) SIMAI & UMI
http://www.bdim.eu/

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1985.

SEZIONE II

(Fisica, chimica, geologia, paleontologia e mineralogia)

Strutturistica chimica. — *Ricerca di conformazioni di minima energia potenziale intramolecolare mediante un nuovo metodo di minimizzazione globale.* Nota di CAMILLO TOSI (*), RAFFAELLA PAVANI (*), ROBERTO FUSCO (*), FILIPPO ALUFFI-PENTINI (**), VALERIO PARISI (**) e FRANCESCO ZIRILLI (**), presentata (***) dal Corrisp. R. UGO.

SUMMARY. — A recent method for the global optimization of a real valued function defined in the n-dimensional euclidean space has been applied to the detection of the most stable conformation of a molecule. The method, inspired by the quantum mechanical theory of diffusion across potential barriers, finds the deepest minimum of the intramolecular energy hypersurface by following the asymptotic behaviour of the paths of a system of stochastic differential equations. As an example of its application. the search for the global minimum of the deoxyribose–phosphate–deoxyribose fragment (regarded as the smallest unit representative of polynucleotide backbones) is described,

1. INTRODUZIONE

Uno dei problemi più importanti nello studio teorico delle proprietà molecolari è costituito dalla ricerca delle strutture stabili di una molecola. Tali strutture corrispondono ai punti dell'iperspazio conformazionale, definito dal vettore delle coordinate interne $\boldsymbol{x} \equiv (x_1, x_2, \ldots, x_{3N-6})$ (dove N è il numero di atomi contenuti nella molecola), nei quali l'energia potenziale intramolecolare $E(\boldsymbol{x})$ è minima. Tutti questi punti sono caratterizzati dall'annullamento delle derivate parziali prime di E, e dal fatto che la matrice delle derivate parziali seconde sia una matrice non negativa; detta matrice determina, attraverso la legge di distribuzione di Boltzmann, la popolazione degli stati conformazionali intorno al minimo.

(*) Istituto Guido Donegani S.p.A., Via G. Fauser, 4 – 28100 Novara. Lavoro realizzato con il contributo del progetto finalizzato del C.N.R. Chimica Fine e Secondaria, contratto n. 83.00437.95.

(**) Filippo Aluffi-Pentini: Dipartimento di Matematica, Università di Bari–70125 BARI; Valerio Parisi: Dipartimento di Fisica, II Università di Roma (Tor Vergata) – 00173 ROMA; Francesco Zirilli: Istituto di Matematica, Università di Salerno – 84100 SALERNO. Lavoro realizzato con il contributo del Governo degli Stati Uniti d'America attraverso l'European Research Office della U.S. Army, contratto n. DAJA–37–81–C–0740.

(***) Nella seduta del 13 aprile 1985.

Tra i vari minimi della funzione $E(x)$ nell'iperspazio conformazionale il maggiore interesse in chimica o biologia è presentato dal minimo o dai minimi globali, cioè dai punti x^* tali che per ogni x sia:

$$E(x^*) \leq E(x) .$$

Data l'elevata complessità del problema, la ricerca dei minimi globali della funzione di energia potenziale non può essere eseguita per via analitica e richiede l'impiego di procedimenti numerici. Esistono numerosi programmi di calcolo per la minimizzazione di funzioni; ma, indipendentemente dalle loro caratteristiche peculiari e dalle condizioni ottimali di applicabilità, essi sono contraddistinti dalla proprietà di raggiungere un minimo locale, che generalmente è il più vicino al punto da cui si intraprende la minimizzazione, e di non poter fornire quindi alcuna garanzia che l'energia della conformazione finale trovata sia la più bassa possibile per la molecola in questione.

L'unico modo che, in linea di principio, assicura il raggiungimento del cosiddetto minimo globale è l'esplorazione « a tappeto » dell'iperspazio conformazionale: una possibilità puramente teorica, in quanto il numero di valutazioni della funzione $E(x)$ che dovrebbero essere eseguite richiederebbe tempi di calcolo proibitivi.

Questo discorso vale anche se si mantiene rigida la geometria di valenza (lunghezze di legame ed angoli di valenza costanti) e ci si limita all'esplorazione del sottospazio torsionale: si consideri ad esempio che, per una molecola nella quale l'energia $E(x)$ dipende da cinque angoli di rotazione interna, una esplorazione sufficientemente accurata da poter sperare di trovare tutti i possibili minimi richiede che l'energia sia calcolata ad intervalli angolari di al più 15° per ogni legame ruotato: questo comporta il calcolo dell'energia in circa 8×10^6 punti dell'iperspazio conformazionale. Se gli atomi contenuti nella molecola sono qualche decina il numero di valutazioni di E, ed i tempi di calcolo, anche su elaboratori (scalari) di notevole potenza, sono elevatissimi. La conclusione che ne consegue è che non è possibile trovare in modo diretto il minimo globale delle funzioni di energia potenziale intramolecolare.

Come verrà mostrato in questa comunicazione, abbiamo potuto colmare tale lacuna grazie all'applicazione di un nuovo metodo, ispirato alla termodinamica statistica, con il quale il minimo globale è ottenuto numericamente seguendo le traiettorie di un sistema di equazioni differenziali stocastiche.

Nel paragrafo 2 viene descritto il metodo di minimizzazione globale adottato; nel paragrafo 3 il metodo viene applicato allo studio dell'energia conformazionale delle molecole e se ne presenta un esempio particolare; nel paragrafo 4 vengono tratte alcune semplici conclusioni.

2. Descrizione dell'algoritmo per la ricerca del minimo globale

L'algoritmo di minimizzazione globale adottato, che si basa sul metodo proposto da Aluffi-Pentini, Parisi e Zirilli in [1], è descritto in dettaglio in [2], mentre la sua traduzione in un insieme di sottoprogrammi FORTRAN è in [3].

Il metodo ricerca un punto di minimo globale di una funzione $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$ di n variabili reali, seguendo le traiettorie generate da una opportuna discretizzazione numerica dell'equazione differenziale stocastica

$$d\boldsymbol{x} = -\nabla f(\boldsymbol{x})\, dt + \varepsilon\, d\boldsymbol{w} \tag{1}$$

[a partire da una condizione iniziale $\boldsymbol{x}(0) = \boldsymbol{x}_0$], essendo ∇f il gradiente di $f(\boldsymbol{x})$, $\boldsymbol{w}(t)$ un processo stocastico di Wiener standardizzato a n dimensioni, e ε un coefficiente positivo (« coefficiente di rumore ») che consideriamo variabile nel tempo.

L'equazione (1) – generalmente considerata con ε costante – è nota come equazione di Smoluchowski e Kramers, ed è usata ad esempio nello studio della diffusione degli atomi nei cristalli, o nello studio di certe reazioni chimiche.

In queste applicazioni l'equazione (1) rappresenta una diffusione attraverso barriere di potenziale sotto l'azione di una forza aleatoria $\varepsilon\, d\boldsymbol{w}$, essendo f il potenziale e $\varepsilon^2 = (2kT/m)$, ove k è la costante di Boltzmann, T la temperatura assoluta, e m un coefficiente di massa.

L'equazione (1) si può considerare come caso limite dell'equazione (del 2° ordine) di Langevin, quando si trascura il termine inerziale.

L'uso dell'equazione (1) è suggerito dal comportamento per t molto grande del processo aleatorio $\boldsymbol{x}(t)$ soluzione dell'equazione (1) con ε costante, a partire da un punto iniziale $\boldsymbol{x}_0$: infatti, sotto ipotesi molto poco restrittive sulla funzione f, la densità di probabilità di $\boldsymbol{x}(t)$ all'istante t tende, per $t \to \infty$, a una densità di probabilità limite

$$p(\boldsymbol{x}) = A\, e^{-\frac{2}{\varepsilon^2} f(\boldsymbol{x})}$$

indipendentemente dal punto iniziale $\boldsymbol{x}_0$ (A è una costante di normalizzazione), che risulta tanto più concentrata intorno ai minimi globali di f quanto minore è ε, fino a diventare, nel limite per $\varepsilon \to 0$, una somma pesata di funzioni δ di Dirac centrate sui minimi globali (ad esempio, in una dimensione ($n = 1$), se $f(x)$ ha due minimi globali a e b, la densità $p(x)$ tende, per $\varepsilon \to 0$, alla densità

$$\gamma\, \delta(x - a) + (1 - \gamma)\, \delta(x - b)$$

ove $\gamma = (1 + \sqrt{\beta/\alpha})^{-1}$, essendo $\alpha = f''(a) > 0$ e $\beta = f''(b) > 0$).

Dato il punto iniziale $\boldsymbol{x}_0$, la discretizzazione usata per la (1) ha la forma

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - h_k\, n\, \tilde{\eta}_k\, \boldsymbol{r}_k + \varepsilon(t_k) \sqrt{h_k}\, \boldsymbol{u}_k \qquad k = 0, 1, 2, \ldots \tag{2}$$

in cui h_k è la lunghezza del passo di integrazione temporale, $t_k = h_0 + h_1 + h_2 + \ldots + h_{k-1}$, $\boldsymbol{r}_k$ e $\boldsymbol{u}_k$ sono due vettori aleatori in n-dimensioni scelti il primo da una distribuzione uniforme sulla sfera unitaria e il secondo da una

distribuzione gaussiana standard, e $\tilde{\eta}_{lk}$ è una approssimazione a differenze finite della derivata direzionale nella direzione $\boldsymbol{r}_k$.

L'algoritmo considera un numero fisso di traiettorie generate dalla (2), che si sviluppano (simultaneamente ma indipendentemente una dall'altra), a partire dalle stesse condizioni iniziali, durante un « periodo di osservazione » in cui il coefficiente di rumore di ogni traiettoria è mantenuto costante, mentre h_k e il passo Δx_k usato per calcolare $\tilde{\eta}_{lk}$ sono aggiustati automaticamente per ciascuna traiettoria.

Al termine di ogni periodo di osservazione le traiettorie sono confrontate, una di esse viene scartata, tutte le altre continuano imperturbate nel periodo di osservazione seguente, e una di esse è prescelta per dare luogo a una « diramazione », cioè a una seconda continuazione della stessa traiettoria, che differisce dalla prima solo per i valori iniziali di ε e Δx, ma che si considera avere la stessa « storia passata » della prima.

Il numero totale di traiettorie simultanee rimane perciò invariato, e la seconda continuazione prende – dal punto di vista del programma di calcolo – il posto della traiettoria scartata.

L'insieme delle traiettorie simultanee è considerato come una singola « prova », e l'algoritmo completo è un insieme di prove ripetute.

Una singola prova è arrestata (al termine di un periodo di osservazione, e dopo aver eliminato la traiettoria peggiore) se tutti i valori finali della f nelle rimanenti traiettorie risultano – entro tolleranze numeriche, ed eventualmente in punti finali diversi – uguali tra loro (arresto « uniforme »).

La prova è in ogni caso arrestata, alla fine di un periodo di osservazione, se si è raggiunto un dato numero massimo di periodi di osservazione.

La prova è considerata un successo soltanto nel caso di un arresto uniforme su un valore finale che sia (numericamente) uguale al più basso valore trovato per f dall'inizio dell'algoritmo.

Le prove sono ripetute cambiando i valori di alcuni parametri, e l'intero algoritmo è arrestato, al termine di una prova, se si raggiunge un dato numero di arresti uniformi tutti al livello del migliore valore di f trovato, o in ogni caso se si raggiunge un dato numero massimo di prove.

L'algoritmo considera di aver trovato il minimo globale se si è avuto almeno un arresto uniforme al livello del migliore valore trovato per f.

3. Applicazione dell'algoritmo all'analisi conformazionale

L'algoritmo descritto nel paragrafo 2 è stato concepito per un uso del tutto generale.

Per lo studio delle posizioni di equilibrio delle molecole, nel caso di geometria di valenza rigida, la funzione da minimizzare è l'energia conformazionale

$$E(\varphi) = E(\varphi_1 \ldots \varphi_m),$$

in cui gli argomenti $\varphi_1 \ldots \varphi_m$ sono gli angoli di torsione liberi.

Il calcolo effettivo di E (φ) viene effettuato come segue.

Dati gli angoli $\varphi_1 \ldots \varphi_m$ si calcola direttamente il contributo torsionale all'energia E (φ), le coordinate cartesiane e successivamente i potenziali tra coppie di atomi non legati nella conformazione in oggetto (espressi ad esempio attraverso un termine di Lennard-Jones ed un termine coulombiano). L'energia E (φ) è ottenuta come somma delle energie torsionale e di non legame.

Nella fig. 1 è descritto schematicamente il diagramma di flusso relativo agli algoritmi di calcolo.

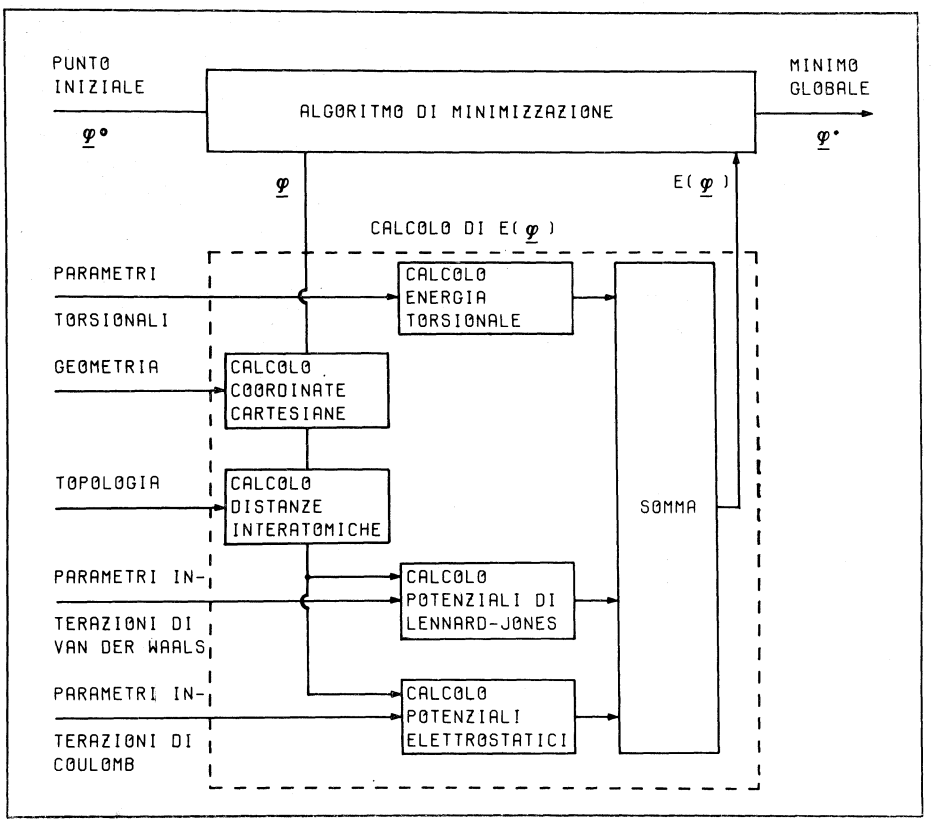

Fig. 1. – Descrizione schematica del diagramma di flusso relativo agli algoritmi di calcolo.

Come esempio di applicazione del nuovo algoritmo, consideriamo il frammento (di DNA) desossiribosio – fosfato – desossiribosio mostrato in fig. 2, che nel seguito indicheremo con la sigla SPS (sugar-phosphate-sugar). Per questo frammento Matsuoka, Tosi e Clementi calcolarono, con un metodo quantomeccanico ab initio [4, 5], le energie di un centinaio di conformazioni ottenute variando gli angoli di rotazione interna ε, ζ, α, β, γ, e mantenendo la geometria di valenza costante (con le lunghezze di legame e gli angoli di valenza risultanti dall'analisi cristallografica della citosina-3′-fosfato [6]). Mediante un

procedimento di *best fit* con tali energie fu trovata una funzione di potenziale costituita da un'espressione di Lennard-Jones per le energie d'interazione fra atomi non legati e da un'espressione di Pitzer per le energie torsionali intrinseche:

$$E(\varphi) = \sum_{i>j} \left(-\frac{A_{ij}}{r_{ij}^{6}} + \frac{B_{ij}}{r_{ij}^{12}} \right) + \sum_{i=1}^{5} \tfrac{1}{2} K_{\varphi_i} (1 + \cos 3\varphi_i) ,$$

dove i φ_i sono gli angoli di torsione ε, ζ, α, β, γ, e la prima delle due sommatorie è estesa alle coppie di atomi non legati. Per le ragioni esposte in [5], il contributo elettrostatico non fu incluso nel *fit*. I valori numerici dei parametri A, B e K_φ sono riportati nella Tabella I del Rif. [7]. Si osservi che per le interazioni fra l'idrogeno e gli ossigeni del gruppo PO_4^-, il coefficiente attrattivo A è sensibilmente più elevato, e il coefficiente repulsivo B sensibilmente più basso, che per le interazioni fra l'idrogeno e gli altri ossigeni: corrispondentemente tanto l'ascissa quanto l'ordinata del punto di minimo della curva energia-distanza sono nettamente inferiori nel primo caso che nel secondo (2,04 Å contro 2,72 Å e $-10{,}59$ *k*J mol^{-1} contro $-0{,}61$ *k*J mol^{-1}). C'è quindi da aspettarsi che il potenziale tenda a favorire strutture stabilizzate da legami di idrogeno intramolecolari con gli ossigeni del gruppo fosfato. L'applicazione del

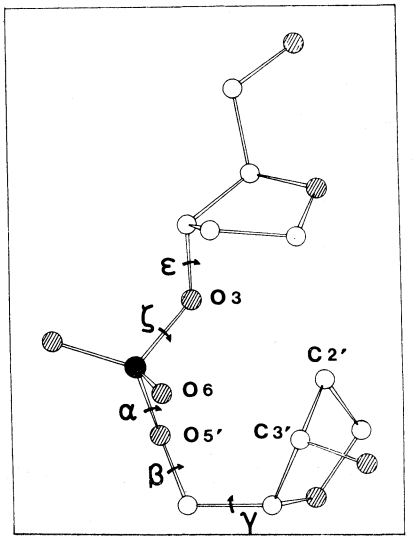

Fig. 2. – Il frammento C 2′ (endo) desossi SPS nella conformazione di minima energia trovata con il nuovo algoritmo. I cerchi tratteggiati corrispondono agli atomi di ossigeno e il cerchio pieno all'atomo di fosforo. Per maggiore chiarezza grafica, i 18 atomi di idrogeno sono stati omessi.

nuovo metodo conferma questa previsione, in quanto porta ad un minimo globale, rappresentato in fig. 2 (con $\varepsilon = 176{,}2^\circ$, $\zeta = 180{,}0^\circ$, $\alpha = 122{,}1^\circ$, $\beta = -96{,}7^\circ$, $\gamma = 55{,}6^\circ$, $E = -94{,}0$ *k*J mol^{-1}), caratterizzato da contatti

H (C3′)... 06 = 2,15 Å, H (C 2′) ... 03 = 2,12 Å, H (C 2′) ... 05′ = 2,47 Å. Un procedimento di ricerca « diretta » dei minimi di più bassa energia [8] aveva portato all'individuazione di due conformazioni, la prima con $\varepsilon = -75^o$, $\zeta = 180^o$, $\alpha = 70^o$, $\beta = -110^o$, $\gamma = 55^o$, $E = -91{,}9$ *k*J mol^{-1} e la seconda con $\varepsilon = -170^o$, $\zeta = 180^o$, $\alpha = 115^o$, $\beta = -100^o$, $\gamma = 55^o$, $E = -90{,}4$ *k*J mol^{-1}. Il minimo globale ottenuto tramite il metodo qui usato corrisponde alla seconda di esse, per la quale si ha un guadagno energetico di 3,6 *k*J mol^{-1}.

4. CONCLUSIONI

L'uso di un nuovo algoritmo di minimizzazione delle funzioni di energia potenziale intramolecolare, la cui concezione si distacca radicalmente da quella dei metodi finora proposti, pone una serie di problemi ai quali solo l'esperienza acquisita attraverso l'applicazione ad un elevato numero di casi potrà dare una risposta completa. Ci limitiamo qui ad indicarne i principali.

Data la natura non deterministica del nuovo algoritmo, la probabilità di individuare il minimo globale tende ad 1 al tendere all'infinito del numero di valutazioni della funzione, e, di conseguenza, del tempo di calcolo necessario alla loro esecuzione. Per tenere conto di questa peculiarità del programma, è possibile decidere a priori il numero di volte in cui si vuole che una prova finisca nel medesimo minimo prima che questo possa essere considerato come il minimo globale. Quanto più elevato è tale numero, tanto più la probabilità di aver trovato il minimo globale si avvicina alla certezza. Nel nostro esempio esso è stato posto uguale a 5.

Un altro problema di notevole importanza è l'accuratezza ottenibile nella valutazione del punto di minima energia. Proprio perché il suo scopo essenziale è l'individuazione del punto di minimo globale, il programma non raggiunge il grado di accuratezza raggiunto da altri metodi di minimizzazione, tanto che è opportuno, quando si sia individuato il minimo, applicare un metodo più rapido di minimizzazione locale, ad esempio un metodo a convergenza quadratica, quale quello di Newton-Raphson, per avere una minimizzazione numericamente più accurata. Nel caso della molecola di SPS discussa nel paragrafo precedente, si ottiene in questo modo un ulteriore guadagno di 0,6 *k*J mol^{-1}, corrispondentemente agli angoli di rotazione interna $\varepsilon = 176{,}7^o$, $\zeta = 176{,}4^o$, $\alpha = 121{,}1^o$, $\beta = -95{,}4^o$, $\gamma = 55{,}3^o$.

Su questi punti riteniamo di dover lavorare in futuro. Già nella attuale applicazione, peraltro, il nuovo algoritmo appare come uno degli strumenti più potenti nello studio delle strutture molecolari, in particolare nella ricerca di possibili correlazioni fra struttura e attività in molecole di interesse biologico.

BIBLIOGRAFIA

[1] F. ALUFFI-PENTINI, V. PARISI e F. ZIRILLI - *Global optimization and stochastic differential equations*. « Journal of Optimization Theory and Applications », *47*/1 (September 1985), pp. 1–16.

[2] F. ALUFFI-PENTINI, V. PARISI e F. ZIRILLI - *A global optimization algorithm using stochastic differential equations*, Technical Summary Report n. 2791, Mathematics Research Center, University of Wisconsin–Madison, Febr. 1985; inviato per la pubblicazione a ACM Transactions on Mathematical Software.

[3] F. ALUFFI-PENTINI, V. PARISI e F. ZIRILLI - *Algorithm SIGMA. A stochastic-integration global minimization algorithm*, Technical Summary Report n. 2806 Mathematics Research Center, University of Wisconsin–Madison, March 1985; inviato per la pubblicazione a ACM Transactions on Mathematical Software.

[4] O. MATSUOKA, C. TOSI e E. CLEMENTI (1978) - « Biopolymers », *17*, 33.

[5] C. TOSI, E. CLEMENTI e O. MATSUOKA (1978) - « Biopolymers », *17*, 51.

[6] M. SUNDARALINGAM e L.H. JENSEN (1965) - « J. Mol. Biol. », *13*, 914.

[7] C. TOSI e KJ. RASMUSSEN (1981) - « Biopolymers », *20*, 1059.

[8] C. TOSI e E. PESCATORI (1978) - « Gazz. Chim. Ital. », *108*, 365.