

Estratègies pel càlcul d'energia lliure al llarg d'una coordenada amb Umbrella Sampling

November 9, 2004

Contents

1	L'optimització de la PES dona punts al llarg d'un camí de reacció	1
2	Definició d'una coordenada de reacció	2
2.1	La funció s	2
2.2	Distància a un punt (r,p,q,t)	3
3	Aplicació a l'Umbrella Sampling	4
3.1	La funció s	5
3.2	Distància a un punt (r,p,q,t)	6
4	Resultats	6
4.1	La funció s	6
4.2	Distància a un punt (r,p,q,t)	6

1 L'optimització de la PES dona punts al llarg d'un camí de reacció

Hem trobat un camí de reacció s que ve representat per un conjunt de coordenades internes (r, p, q, t) ¹ En estudis previs ja hem vist que l'avanç de la reacció ha de ser descrit per una col·lecció de poques internes ja que el centre actiu del nostre enzim és suficientment flexible com perquè un conjunt de cartesianes sigui massa restrictiu.

$$\begin{array}{cccc|c} r_1 & p_1 & q_1 & t_1 & s_1 \\ r_2 & p_2 & q_2 & t_2 & s_2 \\ r_3 & p_3 & q_3 & t_3 & s_3 \\ (\dots) & & & & \\ r_n & p_n & q_n & t_n & s_n \end{array}$$

¹farem servir l'índex k per especificar una interna qualsevol i l'índex i pels punts discrets del camí de reacció

on s_i per donar-li un valor és la suma de canvi geomètric de totes les internes entre dos punts

$$s_1 = 0; \quad s_i = s_{i-1} + \sum_{k=r,p,q,t} |(k_i - k_{i-1})| \quad \text{tot en \AA} \quad (1)$$

el càlcul del valor d' s és bastant arbitrari ja que la geometria no queda completament definida amb el conjunt (r, p, q, t) i per tant no està definida la distància entre els punts.

Per a conèixer el valor de les internes a una s donada suposem un comportament lineal de les internes entre dos punts d' s s_1, s_2 consecutius

$$\begin{aligned} r(s) &= r_1 + \frac{r_2 - r_1}{s_2 - s_1}(s - s_1); & s &\in [s_1, s_2] \\ p(s) &= p_1 + \frac{p_2 - p_1}{s_2 - s_1}(s - s_1); & s &\in [s_1, s_2] \\ (\dots) & q(s), \quad t(s) \end{aligned} \quad (2)$$

2 Definició d'una coordenada de reacció

Basats en la taula de sobre

2.1 La funció s

L'objectiu és trobar l'expressió de la funció $s(r, p, q, t)$ que depèn del conjunt d'internes (r, p, q, t) .

Per a poder associar una geometria a un valor de la funció s primer definim una funció error pel segment $s \in [1, 2]$.

$$E_{rr} = (r(s) - r)^2 + (p(s) - p)^2 \dots \quad (3)$$

Volem saber quin és el valor d' s que més s'ajusta a un conjunt d'internes donat (per exemple durant una MD). Per tant fem la derivada de la funció error respecte a s i en busquem l'extrem

$$\begin{aligned} \frac{dE_{rr}}{ds} &= 2(r(s) - r)\alpha_{r_{12}} + 2(p(s) - p)\alpha_{p_{12}} \dots = 0 \\ \left\| \frac{dr(s)}{ds} = \frac{r_2 - r_1}{s_2 - s_1} \right\| &\equiv \alpha_{r_{12}} \end{aligned} \quad (4)$$

Troblem un mínim ja que E_{rr} és una paràbola de coeficient positiu i el màxim seria per valors infinits d' s . Substituint $r(s), p(s), q(s), t(s)$ per les seves expressions (eq. 2) i agrupant tenim

$$\sum_{k=r,p,q,t} k_1 \alpha_{k_{12}} + s \alpha_{k_{12}}^2 - s_1 \alpha_{k_{12}}^2 - k \alpha_{k_{12}} = 0 \quad (5)$$

d'aquesta manera donat el comportament lineal de les funcions $r(s), p(s), \dots$ per a un únic segment $[s_1, s_2]$ podem aïllar la variable s .

$$s = \left(\sum_{k=r,p,q,t} s_1 \alpha_{k_{12}}^2 + k \alpha_{k_{12}} - k_1 \alpha_{k_{12}} \right) \left(\sum_{k=r,p,q,t} \alpha_{k_{12}}^2 \right)^{-1}; \quad s \in [s_1, s_2] \quad (6)$$

Resumint, tenim en l'equació 6 una funció d' s que durant una Dinàmica Molecular si en mesurem les quatre internes implicades $k = r, p, q, t$ podem saber a quin punt del camí de reacció es troba. Ara bé, hem dit que donada una geometria la s és única, però això serà cert sempre i quan el valor de $s \in [s_1, s_2]$, fora d'aquest interval l'equació 6 no és vàlida i cal canviar de segment. Per tant hem de fer el càlcul de s per tots els $n - 1$ segments. Els casos patològics són o bé que hi hagi dos segments en què la s predita per l'equació 6 cau dins l'interval pel qual està definida i per tant tenim varis valors de s per a una única geometria (cas 1), o bé no n'hi ha cap (cas 2). Pel cas 1 es pot pensar en associar el valor d' s que dona el segment més proper al que centrarem l'umbrella. Pel cas 2 no hi veig solució.

2.2 Distància a un punt (r, p, q, t)

Podem pensar que per saber si avança la reacció no cal construir una funció s . Una alternativa és mesurar la distància en el sentit de la reacció entre el punt donat i un punt de referència. El terme "la distància en el sentit de la reacció" és el punt clau d'aquesta nova aproximació. Hi haurà dues maneres, a través del solapament sobre un vector instantani, o a través de la norma d'un vector de correcció.

Suposem que volem centrar la nostra simulació a un valor d' s , s_o . A partir de la taula i de la linearització en l'equació 2 podem saber quin és el conjunt d'internes al que volem centrar la nostra simulació (r_0, p_0, q_0, t_0).

Durant la simulació tindrem un punt en cartesianes X_1 on les internes seran (r_1, p_1, q_1, t_1) . Agafant el punt X_1 forcem la geometria perquè tinguí les internes exactament iguals a (r_0, p_0, q_0, t_0) a través d'un algoritme iteratiu tipus shake per internes i obtenim el punt en cartesianes X_0 .

Explicant una mica millor aquest procés, cada cartesiana dels àtoms implicats es corregeix de la següent manera (fent servir l'índex k per les internes i l pels àtoms):

$$x_l = x_l + \frac{1}{m_l} s_{x_l} \lambda_k \quad (7)$$

on s_{x_l} per distàncies tenen una expressió senzilla

$$s_{x_l} = \frac{\partial r}{\partial x_l} = \frac{\Delta x}{r_k}$$

on λ seria el multiplicador de Lagrange

$$\lambda_k = \frac{\Delta R_k}{g_k}$$

$$\Delta R_k = r_{k_0} - r_{k_1}$$

$$g_k = \sum_l^2 \frac{s_l^2}{m_l}; \quad s_l^2 = s_x^2 + s_y^2 + s_z^2$$

El procés es fa iterativament fins que al diferència entre el valor perseguit i l'obtingut cau per sota un llindar ($\Delta R_k = r_{k_0} - r_{k_1}$) per a totes les internes (veure Tobias and Brooks J. Chem. Phys., 89, 5115–5127, 1988)

El vector diferència entre el punt X_1 i X_0 li direm \vec{d} .

$$\vec{d} = \vec{X}_1 - \vec{X}_0 \quad (8)$$

Ara volem fer avançar el punt X_0 en el sentit d'avanç de la reacció, per tant tornem a les equacions 2 i trobem el valor de les internes per al punt $s_f = s_o + \partial s$ (a la pràctica fem un

increment finit de $\sim 0.1\text{\AA}$) és a dir (r_f, p_f, q_f, t_f) . Aplicant el mateix algoritme shake trobem el punt X_f . Definim el vector $f\vec{w}d$

$$f\vec{w}d = \frac{\vec{X}_f - \vec{X}_0}{\|\vec{X}_f - \vec{X}_0\|} \quad (9)$$

Per a fer retrocedir el punt X_0 en el sentit de la reacció tornem a l'equació 2 i trobem el conjunt d'internes (r_b, p_b, q_b, t_b) per al punt $s_b = s_o - \partial s$ que de nou aplicant el shake tenim el punt X_b i el vector $bc\vec{w}d$

$$bc\vec{w}d = \frac{\vec{X}_b - \vec{X}_0}{\|\vec{X}_b - \vec{X}_0\|} \quad (10)$$

Cal dir aquesta manipulació de coordenades és externa a la MD, que només es fa pels àtoms implicats en les internes i que no requereix cap càlcul d'energia addicional.

Tenint això en compte he pensat en dos indicadors del grau d'avanç de la reacció i que per tant es poden fer servir de coordenades de reacció.

La norma del vector \vec{d} : Com que una norma és sempre positiva calcularem el producte escalar de \vec{d} amb el vector $f\vec{w}d$ i amb el $bc\vec{w}d$

$$norm = \begin{cases} abs(\|\vec{d}\|) & \text{si } \vec{d} \cdot f\vec{w}d > \vec{d} \cdot bc\vec{w}d \\ -abs(\|\vec{d}\|) & \text{si } \vec{d} \cdot f\vec{w}d < \vec{d} \cdot bc\vec{w}d \end{cases}$$

La projecció de \vec{d} sobre el vector $f\vec{w}d$:

$$proj = \sum_l^{3n} d(l) f\vec{w}d(l) \quad (11)$$

Aquesta aproximació recordaria a la primera estratègia que vem provar calculant la projecció d'un vector instantani sobre un vector fix. En aquell cas però el vector que aquí en diem $f\vec{w}d$ era el mateix al llarg de la simulació. Les conclusions que vaig acabar treient eren que la reacció, degut a un centre actiu flexible, no tenia sempre la mateixa direcció (en \mathbb{R}^3) d'avanç de la reacció. En aquest cas, podríem pensar que a cada estructura trobem quin és aquest vector d'abans $f\vec{w}d$ tal que anem en el sentit $s_0 + \partial s$.

Cal tenir en compte que donat que les internes que fem servir no especifiquen un punt concret de l'espai (en tenim menys que les 3N-6), no podem construir una superfície ortogonal al camí de reacció especificada per aquestes internes. D'aquí que una exploració en el plà perpendicular al camí de reacció no té gaire sentit (veure "Reaction paths and free energy profiles for conformational transitions: An internal coordinate approach. J. Chem. Phys. 95, 7612").

3 Aplicació a l'Umbrella Sampling

Un cop definida una coordenada de reacció podem pensar en aplicar-ho a un càlcul d'energia lliure pel mètode de l'umbrella sampling. Totes dues formulacions de coordenades s'han implementat i testejat.

3.1 La funció s

Quan la coordenada és la mateixa funció s tal i com s'ha definit en l'equació 6 el potencial umbrellà tindrà la següent forma

$$U = \frac{1}{2}K(s - s_o)^2 \quad (12)$$

I la derivada respecte a una coordenada cartesiana x_l , que són els tipus de coordenades en les que s'integren les equacions de moviment, és

$$\frac{\delta U}{\delta x_l} = \frac{\delta U}{\delta s} \frac{\delta s}{\delta r} \frac{\delta r}{\delta x_l} \quad \text{per } x_l \text{ implicada en l'internal } r \quad (13)$$

si x_l està implicada en més d'una interna

$$\frac{\delta U}{\delta x_l} = \frac{\delta U}{\delta s} \left(\frac{\delta s}{\delta r} \frac{\delta r}{\delta x_l} + \frac{\delta s}{\delta p} \frac{\delta p}{\delta x_l} \right) \quad (14)$$

on

$$\frac{\delta U}{\delta s} = K(s - s_o) \quad (15)$$

$$\frac{\delta s}{\delta r} = \alpha_{r_{12}} \left(\sum_{k=r,p,q,t} \alpha_{k_{12}}^2 \right)^{-1} \quad (16)$$

$$\frac{\delta r}{\delta x_l} = \frac{\Delta x_l}{r} \quad (17)$$

$$\left\| \begin{array}{l} \Delta x_l = x_l - x_j \quad \text{on els àtoms } l \text{ i } j \text{ estan a distància } r \end{array} \right.$$

En el càlcul de derivades he suposat dues coses.

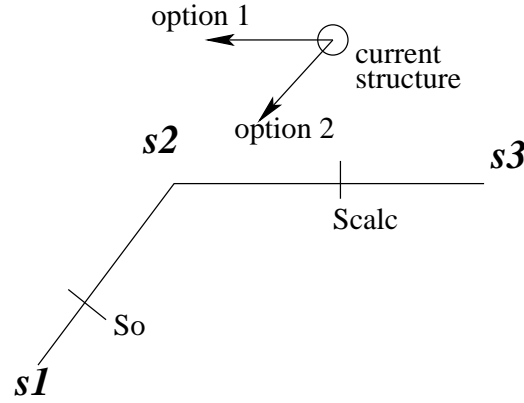
- Que en l'equació 6 només hi ha un segment en què la s calculada cau dins l'interval per la qual és vàlida. Sinó cal escollir-ne un.
- Que aquest interval és el mateix en el que es troba el centre del potencial umbrellà s_o

Si el segon punt no es compleix, per exemple que la s_{calc} està entre s_2 i s_3 mentre que la s_o està entre s_1 i s_2 . En aquest cas tenim dues opcions pel què fa a l'aplicació de la derivada

Opció 1 La derivada es calcula pel segment en el qual l'estructura actual es troba (s_{calc}) i per tant en l'equació 16 cal utilitzar $\alpha_{k_{23}}$ i $\alpha_{r_{23}}$

Opció 2 La derivada es calcula pel segment en el qual el potencial umbrellà es troba (s_o) i per tant en l'equació 16 cal utilitzar $\alpha_{k_{12}}$ i $\alpha_{r_{12}}$

El sentit de la derivada es pot veure esquemàticament en la següent figura



L'opció que em sembla més correcta és la 1.

3.2 Distància a un punt (r,p,q,t)

Per l'aplicació a l'umbrella sampling de la coordenada de la secció 2.2 apliquem un potencial del tipus

$$U = \frac{1}{2}K(norm)^2 \quad (18)$$

o bé

$$U = \frac{1}{2}K(proj)^2 \quad (19)$$

Totes dues aproximacions estan centrades al s_o i per tant el valor de la coordenada serà $s_o + norm$ o $s_o + proj$. En aquest cas les derivades tindran una expressió més senzilla

$$\frac{\partial U}{\partial norm} = K(norm) \quad ; \quad \frac{\partial U}{\partial proj} = K(proj) \quad (20)$$

$$\frac{\partial norm}{\partial x} = \frac{d_x}{norm} \quad ; \quad \frac{\partial proj}{\partial x} = f w d_x \quad (21)$$

4 Resultats

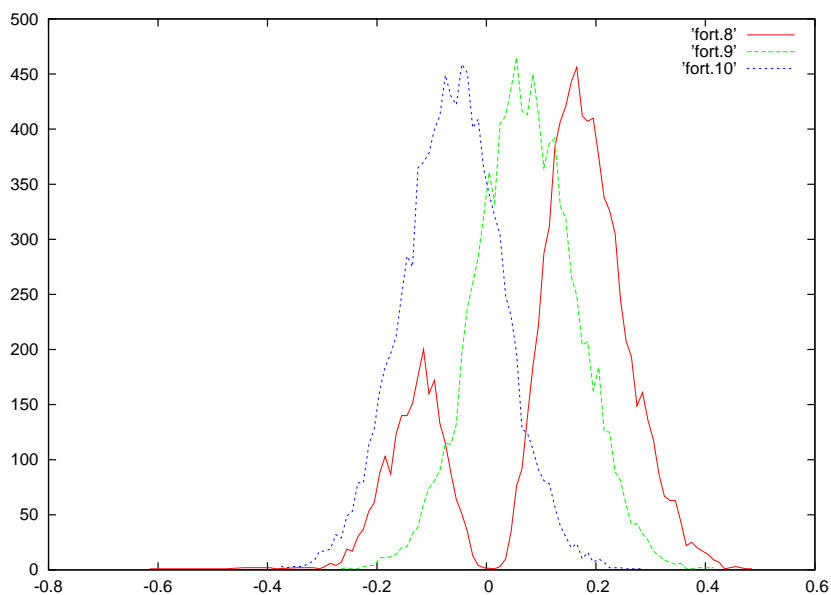
4.1 La funció s

Veig canvis sobtats del valor d' s que no corresponen a canvis significatius de la geometria, a més he observat estructures en les que es dona el cas patològic 2 de la secció. Això fa que al llarg d'una dinàmica hi hagi estructures que no tenen una s associada, arribats a aquest punt el càlcul el paro.

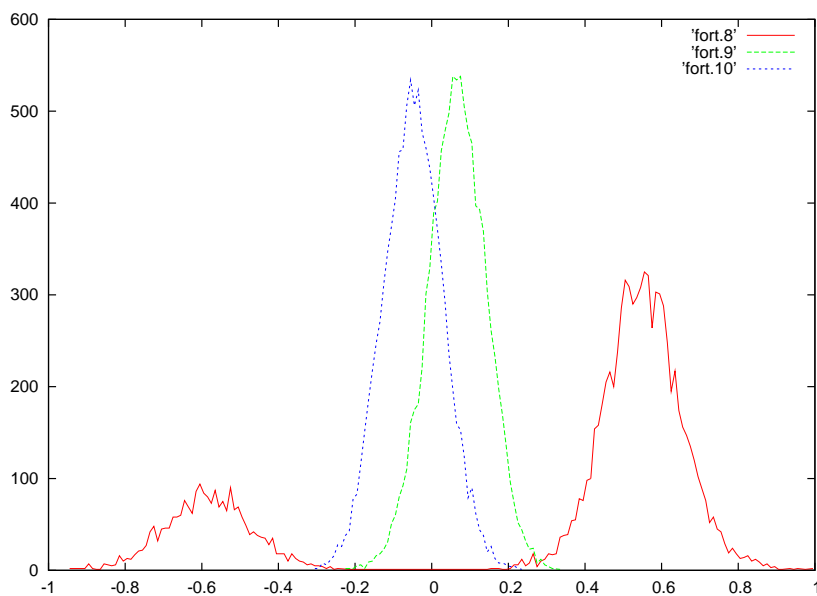
4.2 Distància a un punt (r,p,q,t)

Si faig servir com a coordenada la $norm$ aconsegueixo fer una simulació on la geometria està al voltant d'allà on centro l'umbrella. És a dir, per primera vegada aconsegueixo arrossegar la geometria d'un punt a un altre del camí. El problema està en què la norma té un mínim en el zero que és precisament allà on se centra l'umbrella. Això és normal ja que el fet de fer-la positiva o

negativa és artificial. En el següent gràfic es mostra els histogrames de la *norm*(vermell), i de la projecció sobre el vector fwd(verd) i bcwd(blau) que afortunadament són bastant simètrics



Si faig servir la projecció com a coordenada de reacció, si bé obtinc histogrames semblant als d'aquí dalt no aconsegueixo arrossegar la geometria.



Com és d'esperar el vector norma dona valors més extrems ja que geomètricament estem més lluny d'allà on centrem la simulació