

## ALGORISME D'OPTIMITZACIÓ GLOBAL DE LES MESURES DE SEMBLANÇA QUÀNTICA MOLECULAR

L. Amat, R. Carbó i P. Constans

Institut de Química Computacional, Universitat de Girona  
Albereda, 3-5. 17071 Girona.

---

### RESUM

Les Mesures de Semblança Quàntica Molecular (MSQM) requereixen la maximització del solapament de les densitats electròniques de les molècules que es comparen. En aquest treball es presenta un algorisme de maximització de les MSQM, que és global en el límit de densitats electròniques deformades a funcions deltes de Dirac. A partir d'aquest algorisme se'n deriva l'equivalent per a densitats no deformades.

### RESUMEN

Las Medidas de Semejanza Cuántica Molecular precisan de la maximización del solapamiento de las densidades electrónicas de las moléculas que se comparan. En este trabajo se presenta un algoritmo de maximización de estas medidas de semejanza, que es global en el límite de la densidad electrónica deformada a funciones deltas de Dirac. A partir de este algoritmo se deriva el equivalente para densidades no deformadas.

### ABSTRACT

Quantum Molecular Similarity Measures (QMSM) implicitly require the maximization of the electron density overlap of the molecules involved in the comparison process. This contribution introduces an optimization algorithm for QMSM, which is global if electron densities are deformed to Dirac density functions. A derivation of the algorithm to non-deformed molecular electron densities is proposed and its efficiency tested.

---

## 1. INTRODUCCIÓ

La finalitat de la Semblança Quàntica (SQ) és la identificació de relacions entre sistemes quàntics i la definició de la seva quantificació o mesura. La SQ és una disciplina relativament moderna que té el seu inici al principi de la dècada dels vuitanta amb la introducció de les MSQM [1]. Al nostre laboratori se n'ha fet el desenvolupament teòric [1-14], com també s'ha iniciat la seva aplicació en estudis de relacions quantitatives estructura-activitat i estructura-propietat (QSAR-QSPR). Un aspecte clau en aquesta teoria, intrínsecament lligat a la definició, és el de l'optimització de les MSQM. En aquest treball es presenta un algorisme de maximització, global en el límit de densitats electròniques compactades a deltes de Dirac, determinístic, i eficient quan es compara amb altres estratègies de maximització emprades.

Quan els sistemes quàntics que s'han de comparar són molècules, les seves fun-

cions de densitat electrònica ens les descriuen com a funcions de tres variables. Així, una mesura simple i intuïtiva de la similitud entre dues molècules  $A$  i  $B$  serà la que maximitzarà l'encobriment de les seves densitats

$$Z_{AB}(\Theta) = \int \rho_A(\mathbf{r}) \rho_B(\mathbf{r}; \Theta) d\mathbf{r} \quad (1)$$

on el vector  $\Theta$  representa la dependència de la MSQM respecte a la posició relativa de les molècules en l'espai. En l'estudi de sistemes moleculars s'utilitza normalment l'aproximació de Born-Oppenheimer [15], que desacobla els moviments electrònic i nuclear. Llavors les molècules es poden descriure per les coordenades espacials dels nuclis, normalment fixes en alguna conformació, i les densitats de probabilitat electrònica. Consegüentment, les MSQM seran dependents de la posició relativa de les dues molècules que s'han de comparar i la seva definició inequívoca es pren en el màxim absolut. El procés d'optimització és la part més lenta de tot l'estudi ja que requereix avaluar moltes vegades la integral de semblança. Per això s'ha desenvolupat el càlcul de funcions de densitat aproximades, d'integració ràpida, i que donen mesures de semblança de suficient qualitat, quan es comparen amb valors *ab initio*.

## 2. MSQM APROXIMADES

Una simplificació important de les funcions de densitat electrònica consisteix en la seva representació com a combinació lineal de funcions gaussianes  $1s$  centrades en els àtoms. És l'anomenada aproximació de capes atòmiques (*Atomic Shell Approximation, ASA*) [10] i respon a l'equació general

$$\rho_A^{ASA}(\mathbf{r}) = \sum_{a \in A} \sum_{i \in a} n_i^a S_i^a(\mathbf{R}_a - \mathbf{r}), \quad (2)$$

on  $S_i^a(\mathbf{R}_a - \mathbf{r})$  és una funció gaussiana  $1s$  centrada en la coordenada nuclear  $\mathbf{R}_a$  del nucli de l'àtom  $a$

$$S_i^a(\mathbf{R}_a - \mathbf{r}) = \left( \frac{\xi_i}{\pi} \right)^{3/2} e^{-\xi_i(\mathbf{R}_a - \mathbf{r})^2}, \quad (3)$$

i  $n_i^a$  és el nombre d'ocupació electrònica d'aquesta funció o capa atòmica. Aquests coeficients  $n_i^a$  s'obtenen variacionalment, ajustant la funció densitat aproximada a l'*ab initio*. En un procés iteratiu se seleccionen les funcions  $S_i^a$  compatibles amb la descripció estadística de les densitats de probabilitat, és a dir, que els coeficients  $n_i^a$  que minimitzin l'error quadràtic integral siguin més grans de zero.

Substituint les funcions densitat de l'equació (1) per les funcions ajustades definides en l'equació (2), s'obté una expressió de la MSQM de tipus encobriment on únicament s'han de calcular integrals de dos centres entre funcions  $1s$  gaussianes:

$$Z_{AB}(\Theta) = \sum_{a \in A} \sum_{i \in a} \sum_{b \in B} \sum_{j \in b} n_i^a n_j^b \int S_i^a(\mathbf{R}_a - \mathbf{r}) S_j^b(\mathbf{R}_b(\Theta) - \mathbf{r}) d\mathbf{r}. \quad (4)$$

La reducció en el temps de càlcul que suposa la utilització de les densitats ASA en lloc de les *ab initio* és, com es pot preveure per la seva simplicitat, de diversos ordres de magnitud. Malgrat tot, l'obtenció de densitats ASA, via càlculs *ab initio* previs, pot ser prohibitiva per sistemes moleculars grans. És per això que s'han definit unes funcions densitat empíriques, consistents en la construcció de la densitat electrònica molecular com a simples sumes de les contribucions atòmiques,

$$\rho_A^{EASA}(\mathbf{r}) = \sum_{a \in A} \rho_a(\mathbf{r}). \quad (5)$$

La contribució de cada àtom  $a$  s'expressa com a

$$\rho_a(\mathbf{r}) = Q_a^A \left| S_a^{n_a}(\mathbf{R}_a - \mathbf{r}) \right|^2, \quad (6)$$

on  $Q_a^A$  és la càrrega de població de l'àtom  $a$  en la molècula  $A$  i  $S_a^{n_a}(\mathbf{R}_a - \mathbf{r})$  és una funció esfèrica, *ns* Slater o *Is* gaussiana, centrada en el nucli. S'han definit també funcions de densitat descrivint l'essència de l'estructura atòmica en capes, on s'assigna a cada àtom tantes funcions esfèriques com el nombre de capes característic

$$\rho_a(\mathbf{r}) = \sum_{i_a}^{n_a} m_{i_a} \left| S_a^{i_a}(\mathbf{R}_a - \mathbf{r}) \right|^2, \quad (7)$$

essent el nombre d'electrons usualment associats a les configuracions electròniques de l'àtom  $m_{i_a}$ . Llavors la MSQM de tipus encobriment queda reduïda a l'expressió:

$$Z_{AB}(\Theta) = \sum_{a \in A} \sum_{i_a}^{n_a} \sum_{b \in B} \sum_{j_b}^{n_b} m_{i_a} m_{j_b} \int \left| S_a^{i_a}(\mathbf{R}_a - \mathbf{r}) \right|^2 \left| S_b^{j_b}(\mathbf{R}_b(\Theta) - \mathbf{r}) \right|^2 d\mathbf{r}. \quad (8)$$

Aquestes densitats anàlogues en la forma a les ASA s'han anomenat EASA, abreviació de les sigles angleses *Empirical Atomic Shell Approximation* [13], per la seva construcció aproximada que evita el càlcul *ab initio* de la densitat electrònica.

### 3. SOBREPOSICIONS MOLECULARS

Els mètodes de comparació i quantificació de la similitud estructural de molècules s'han fonamentat tradicionalment en la minimització de la suma de les distàncies entre els àtoms de les molècules que s'han de comparar [16]. Donada a priori una correspondència entre àtoms d'una de les molècules amb els de l'altra, s'optimitzen les posicions relatives de les molècules fins a aconseguir una sobreposició òptima de les dues estructures. Correspondències entre àtoms de molècules diferents porten implícita una certa arbitriietat en la seva elecció que es deixa als criteris dels usuaris dels programes. No definir-les a priori portaria un problema de combinatòria, solament tractable amb algorismes aproximats de naturalesa estocàstica, com per exemple la tècnica del *Simulated Annealing* [17].

També, la maximització de la MSQM  $Z_{AB}(\Theta)$  evita les arbitrarietats en les correspondències atòmiques i el problema combinatori d'establir-les. Aquesta maximització fins ara es duia a terme a partir d'una tècnica mixta d'optimització [12]. A partir d'un mètode de recerca exhaustiva en tot el domini de la funció de semblança es van generant *punts inicials* a partir dels quals, i mitjançant un mètode de gradient o un simplex [18], es fa un refinament per arribar al màxim de semblança més proper. L'eficàcia d'aquesta tècnica mixta depèn en bona part dels *punts inicials* generats, d'acord amb dos aspectes diferents. Primer, com més propers al màxim es trobin, més facilitat tindrà el mètode de gradient o el simplex per arribar al cim. Segon, l'exploració de la superfície ha de ser molt completa, ja que si no es genera un *punt inicial* del màxim absolut, serà molt difícil d'obtenir-lo. Quan aquest *punt inicial* es troba en la conca d'un màxim se l'anomena *punt maximitzador*.

L'algorisme que ara es presenta és un mètode específic associat a les MSQM, útil per a localitzar *punts maximitzadors* de la funció de semblança. Aquesta nova tècnica es deriva de la solució global coneguda per a densitats infinitament compactades. El seu ús a densitats reals, no compactades, està justificat per la seva eficàcia i pels valors màxims trobats.

#### 4. DEDUCCIÓ DE L'ALGORISME: DEFORMACIÓ DE LES DENSITATS

Per a la deducció de l'algorisme de maximització global de la funció  $Z_{AB}(\Theta)$  es proposa la deformació de les densitats electròniques moleculars fins a fer-les col·lapsar a funcions deltes de Dirac. Aleshores, el conjunt dels valors no nuls de  $Z_{AB}(\Theta)$  és finit i la identificació del màxim absolut factible. Introduint un factor de deformació  $t$  a l'equació (2) s'obté:

$$\rho_A(\mathbf{r}; t) = \sum_{a \in A} \sum_{i \in \alpha} n_i^a S_i^a(\mathbf{R}_a - \mathbf{r}; t) \quad (9)$$

que afecta les funcions esfèriques,

$$S_i^a(\mathbf{R}_a - \mathbf{r}; t) = \left( \frac{\zeta_i}{t\pi} \right)^{3/2} e^{-\zeta_i(\mathbf{R}_a - \mathbf{r})^2/t}, \quad (10)$$

de manera que quan  $t$  tendeix a zero la funció esdevé una funció delta de Dirac, mentre que quan és la unitat correspon a la capa no deformada. Així doncs, en el cas límit de  $t$  tendint a zero, la nova funció densitat se simplifica a

$$\tilde{\rho}_A(\mathbf{r}) \equiv \lim_{t \rightarrow 0} \rho_A(\mathbf{r}; t) = \sum_a q_a \delta(\mathbf{R}_a - \mathbf{r}), \quad (11)$$

on  $q_a$  és la suma dels nombres d'ocupació de les capes de cada àtom,

$$q_a = \sum_{i \in \alpha} n_i. \quad (12)$$

A l'equació (11), la funció associada al límit de la funció densitat s'ha marcat

amb una titlla. La MSQM d'encobriment entre aquestes funcions densitat deformades queda reduïda a l'expressió

$$\tilde{Z}_{AB}(\Theta) = \sum_a \sum_b q_a q_b \delta(\mathbf{R}_a - \mathbf{R}_b(\Theta)), \quad (13)$$

que únicament és diferent de zero quan hi ha com a mínim un àtom de la molècula  $A$  que se superposa amb un àtom de la molècula  $B$ . En total són finits els valors no nuls de  $\tilde{Z}_{AB}(\Theta)$  i es poden classificar en els tres conjunts següents:

a) El conjunt de valors de  $\tilde{Z}_{AB}(\Theta)$  corresponents a les posicions de la molècula  $B$  tal que un àtom  $b$  se sobreposa a un àtom  $a$  de la molècula  $A$ , i s'escriu

$$Z_{ab} \equiv \{q_a q_b \delta(0)\} \quad (14)$$

De tot aquest conjunt de solucions el valor màxim  $\tilde{Z}_{ab}^*$  és

$$\tilde{Z}_{ab}^* \equiv \text{Max}(Z_{ab}) \quad (15)$$

b) El conjunt de possibles valors de  $\tilde{Z}_{AB}(\Theta)$  donats per la coincidència de les parelles d'àtoms  $(ab)$  i  $(a'b')$

$$Z_{aba'b'} \equiv \{ (q_a q_b + q_{a'} q_{b'}) \delta(0) \mid d_{aa'} = d_{bb'} \} \quad (16)$$

L'existència d'aquest conjunt està limitada per la presència de distàncies iguals entre les parelles d'àtoms, tal com s'indica en la definició anterior. També és possible identificar-ne el valor màxim per aquest grup

$$\tilde{Z}_{aba'b'}^* \equiv \text{Max}(Z_{aba'b'}) \quad (17)$$

c) El conjunt de possibles valors de  $\tilde{Z}_{AB}(\Theta)$  donats per la coincidència simultània de, com a mínim, les tres parelles  $(ab)$ ,  $(a'b')$  i  $(a''b'')$

$$Z_{aba'b'a''b''} \equiv \{ (q_a q_b + q_{a'} q_{b'} + q_{a''} q_{b''}) \delta(0) + Q_{aba'b'a''b''} \mid d_{aa''} = d_{bb''} \wedge d_{aa''} = d_{bb''} \wedge d_{a'a''} = d_{b'b''} \} \quad (18)$$

Com en el cas anterior, l'existència d'aquest conjunt està condicionada a la presència de les igualtats de distàncies indicades. També, a causa del fet que les molècules  $A$  i  $B$  es prenen rígides en la maximització de la semblança i que el conjunt  $\{a, b, a', b', a'', b''\}$  equival al vector  $\Theta$  de les sis variables que indiquen l'orientació relativa d' $A$  i  $B$ , no es pot definir, en general, un conjunt  $Z_{aba'b'a''b''}$  de tres i només tres superposicions simultànies. Cal, doncs, introduir-hi el terme  $Q_{aba'b'a''b''}$  que modificarà els valors de semblança en el cas que la superposició  $\{a, b, a', b', a'', b''\}$  impliqui també la coincidència de més àtoms:

$$Q_{aba'b'a''b''} \equiv \sum_{a'' \neq a, a', a''} \sum_{b'' \neq b, b', b''} q_{a''} q_{b''} \delta(\mathbf{R}_{a''} - \mathbf{R}_{b''}(a, b, a', b', a'', b'')). \quad (19)$$

Es vol remarcar aquí que el conjunt d'àtoms  $\{a, b, a', b', a'', b''\}$  pot expressar completament, igual que el vector  $\Theta$  que conté els angles d'Euler i les translacions, l'arranjament relatiu de les dues molècules, per bé que, i a diferència de  $\Theta$ , solament definirà un conjunt discret i finit d'arranjaments, que corresponen als que fan  $Z_{AB}(\Theta)$  diferent de zero. La nomenclatura  $\{a, b, a', b', a'', b''\}$  s'interpreta com l'acció de sobreposar l'àtom  $b$  sobre l' $a$ , que és la translació de la molècula, orientar l'àtom  $b'$  de manera que els eixos descrits pels segments  $\overline{aa'}$  i  $\overline{bb'}$  coincideixin, i finalment rotar  $b''$  al voltant de l'eix  $\overline{bb'}$  per fer que el pla descrit pels àtoms  $\{b, b', b''\}$  sigui el mateix que el descrit per  $\{a, a', a''\}$ . Amb aquestes operacions queda perfectament descrit el moviment de la molècula  $B$ .

El valor màxim pel subconjunt  $Z_{ab\ a'b'\ a''b''}$  és:

$$\tilde{Z}_{ab\ a'b'\ a''b''}^* \equiv \text{Max}(Z_{ab\ a'b'\ a''b''}). \quad (20)$$

En aquests tres conjunts estan inclosos tots els valors de  $\tilde{Z}_{AB}(\Theta)$  diferents de zero. El màxim absolut d'aquesta funció límit correspon al valor més gran dels valors dels màxims dels subconjunts (a), (b) i (c):

$$\tilde{Z}_{AB}^* = \text{Max}(\tilde{Z}_{ab}^*, \tilde{Z}_{ab\ a'b'}^*, \tilde{Z}_{ab\ a'b'\ a''b''}^*). \quad (21)$$

Un possible algorisme de recerca del màxim absolut és el que es presenta a la **figura 1**. Aquest algorisme, amb algunes modificacions, és el fonament de l'esquema de maximització de  $Z_{AB}(\Theta)$ , que es presenta a la secció següent.

## 5. MAXIMITZACIÓ DE LES MSQM

S'ha demostrat que quan s'utilitzen funcions densitat deformades a deltes de Dirac el mètode de maximització proposat proporciona el màxim global de la funció de semblança. Quan s'aplica aquest algorisme a sistemes reals com són les densitats ASA i EASA, es fa la hipòtesi que la posició de la molècula  $B$  que s'obté al final del procés correspon a un *punt maximitzador* del màxim absolut de la funció de semblança. A partir d'aquest punt i mitjançant un mètode de gradient és fàcil d'arribar al màxim absolut.

En la **figura 2** es mostra l'esquema general de l'algorisme de maximització, que presenta algunes modificacions respecte al cas de densitats límit. Les més importants es troben en els dues comandaments condicionals de l'algorisme. Mentre que en el cas límit cal que es compleixi estrictament que les distàncies entre les parelles d'àtoms siguin iguals, en el cas real de la funció  $Z_{AB}(\Theta)$  continua aquestes restriccions són expressades per les desigualtats

$$Z_{a'b'}(\text{Min}(d_{a'b'})) > \varepsilon_1. \quad (22)$$

$$Z_{a''b''}(\text{Min}(d_{a''b''})) > \varepsilon_2, \quad (23)$$

que depenen dels paràmetres líndars  $\varepsilon_1$  i  $\varepsilon_2$ . La determinació d'aquests líndars és empírica, i a la pràctica evita un gran nombre de sobreposicions. Aquests condicionals es poden avaluar simplement a través dels valors de les distàncies interatòmiques intermoleculares.

INICIALITZAR  $\tilde{Z}_{AB}^* = 0$

PER TOT  $a \in A$

PER TOT  $b \in B$

$$\tilde{Z}_{AB}^* = \text{Max}(\tilde{Z}_{AB}^*, q_a q_b \delta(0))$$

PER TOT  $a' \in A / a' \neq a$

PER TOT  $b' \in B / b' \neq b$

SI  $d_{aa'} = d_{bb}$  LLAVORS

$$\tilde{Z}_{AB}^* = \text{Max}(\tilde{Z}_{AB}^*, q_a q_b \delta(0) + q_{a'} q_{b'} \delta(0))$$

PER TOT  $a'' \in A / a'' \neq a' \wedge a'' \neq a$

PER TOT  $b'' \in B / b'' \neq b' \wedge b'' \neq b$

SI  $d_{aa''} = d_{bb''} \wedge d_{a'a''} = d_{b'b''}$  LLAVORS

TRASLLADAR  $b$  SOBRE  $a$

ORIENTAR  $bb'$  AMB  $aa'$   $\rightarrow$

ROTAR AL VOLTANT DE  $bb'$  FINS  $d_{a''b''} = 0$

$$\tilde{Z}_{AB}^* = \text{Max}(\tilde{Z}_{AB}^*, \tilde{Z}_{AB}^*(a, b, a', b', a'', b''))$$

FI CONDICIONAL

FI PER TOT  $b''$

FI PER TOT  $a''$

FI CONDICIONAL

FI PER TOT  $b'$

FI PER TOT  $a'$

FI PER TOT  $b$

FI PER TOT  $a$

MÀXIM GLOBAL  $\tilde{Z}_{AB}^*$

Figura 1. Algorisme de maximització global de les MSQM en el límit de densitats compactades.

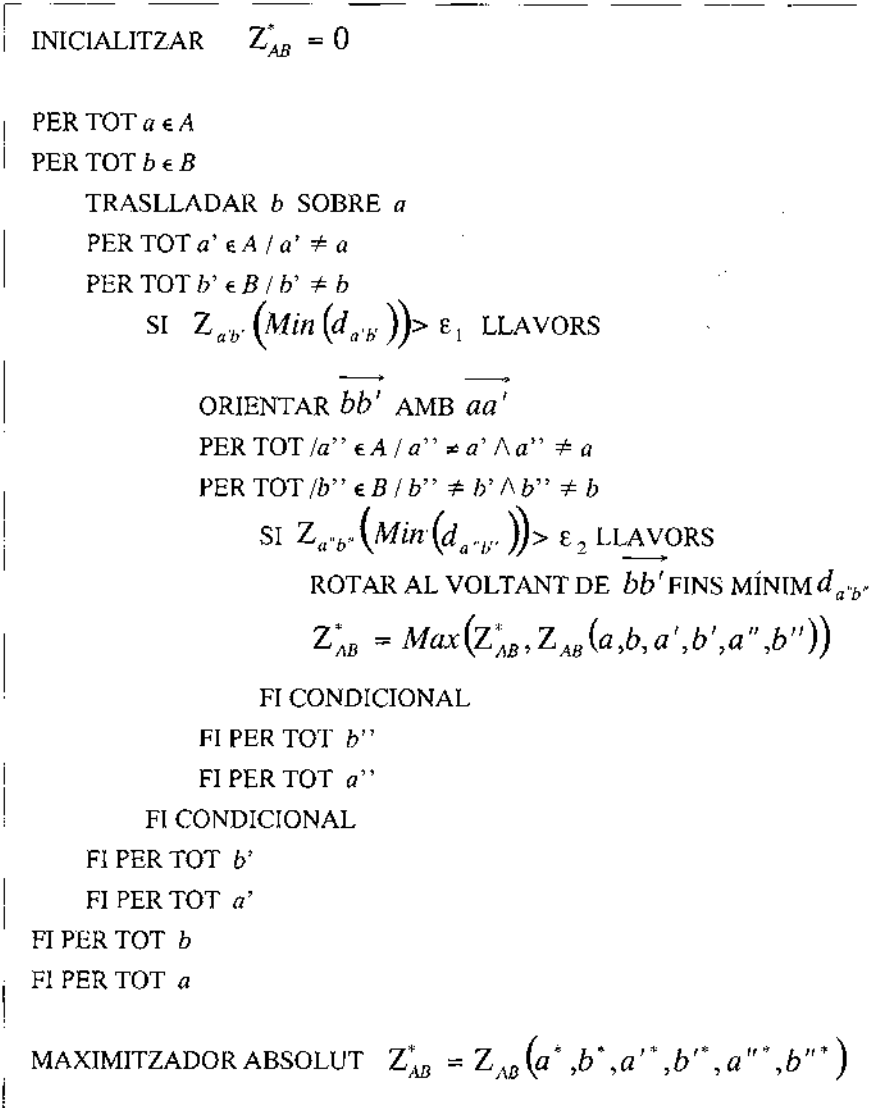


Figura 2. Algorisme d'optimització quasiglobal de les MSQM.

El procés descrit a la figura 2 és d'ordre  $n_A^3 n_B^3$ . Ha estat possible desacoblar els cicles anuats de la figura 2 i obtenir algorismes simplificats però alhora també altament robustos [19]. Així s'han desenvolupat dos algorismes paral·lels al general però d'ordre inferior,  $n_A^2 n_B^2$  i  $n_A n_B$ . L'eliminació dels cicles es fa a partir d'introduir criteris de semblances entre àtoms per escollir les parelles òptimes. Aquestes simplificacions fan que el procés sigui uns quants ordres de magnitud més ràpid.



## 6. EXEMPLE D'APLICACIÓ

Per comprovar la validesa de l'algorisme descrit per la maximització de les MSQM s'ha fet el càlcul de les sobreposicions moleculars per una selecció de sis molècules, les estructures químiques de les quals es donen en la **figura 3**. Són molècules orgàniques dissimilars, on la identificació de correspondències atòmiques seria, si més no, difícilosa. L'estructura molecular és coneguda experimentalment per difracció de raigs X i disponible en la Cambridge Structural Database (CSD) [20]. Això facilitarà, doncs, la reproductibilitat dels valors dels màxims obtinguts.

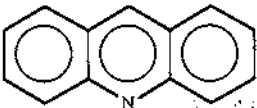
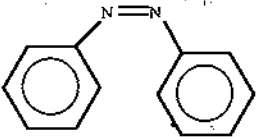
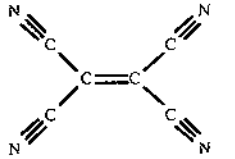
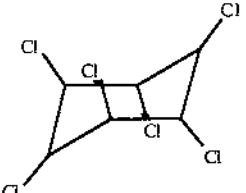
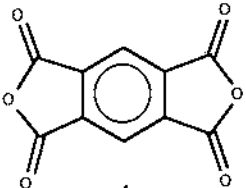

Ref. Cambridge Structural Database	Fórmula	Estructura	Nombre d'àtoms
ACRDIN01	$C_{13}NH_9$		23
AZBENC01	$C_{12}N_2H_{10}$		24
TCYETY	$C_6N_4$		10
HCCYHB	$C_6Cl_6H_6$		18
PYMDAN	$C_{10}O_6H_2$		18
TETDAM03	$C_6N_2H_{12}$		20

Figura 3. Fórmules i estructures químiques de les molècules emprades, com també codis de referència a la CSD.

Per fer les mesures ASA s'ha calculat primer la densitat electrònica *ab initio* de les molècules, al nivell de teoria HF/3-21G, amb el programa *Gaussian 92* [21]. Després s'ha fet l'ajust de les funcions *1s* gaussianes per obtenir la densitat electrònica aproximada i a partir d'aquestes densitats el procés de maximització s'ha portat a terme amb el programa ASAC [22]. Pel que fa a les mesures EASA, s'ha utilitzat el programa *Mol-Simil-96* [23] amb l'ajust dels exponents de les funcions atòmiques a les autosemblances atòmiques en la base 3-21G. Els valors presentats en aquest exemple corresponen a utilitzar una funció *ns* de Slater per capa atòmica, és a dir, s'ha fet servir l'equació (7).

Per a l'optimització del total de quinze parells de molècules s'ha utilitzat l'algorisme general de maximització descrit a la **figura 2**, i amb la posició final que s'obté es fa un refinament amb un mètode de gradient per acabar d'arribar al màxim absolut. Els resultats emprant les funcions densitat ASA es mostren a la **taula 1**, mentre que els resultats amb funcions EASA es poden veure a la **taula 2**. Agafant com a referència els valors ASA, evidentment molt més exactes que els EASA, s'observa que les mesures empíriques tenen sempre un error per excés. El que dona igual per les dues mesures és la sobreposició final de les molècules, és a dir, totes dues van a parar als mateixos màxims de semblança. Fent una anàlisi global, i per tant tenint en compte la diferència de temps de càlcul entre les dues mesures, els resultats presentats justifiquen l'ús de les funcions EASA.

	ACRDIN01	AZBENC01	TCYETY	HCCYHB	PYMDAN	TETDAM03
ACRDIN01	457.623	204.525	170.201	236.453	198.619	117.911
AZBENC01		478.546	144.316	292.388	165.121	113.749
TCYETY			395.053	238.674	144.474	100.980
HCCYHB				6098.180	337.996	211.802
PYMDAN					793.084	108.745
TETDAM03						291.450

**Taula 1.** MSQM per l'aproximació ASA pel conjunt de sis molècules seleccionades.

	ACRDIN01	AZBENC01	TCYETY	HCCYHB	PYMDAN	TETDAM03
ACRDIN01	463.765	206.953	172.511	236.250	200.779	118.996
AZBENC01		484.403	146.159	293.127	167.228	115.103
TCYETY			398.227	238.728	145.048	102.309
HCCYHB				6102.373	337.330	212.092
PYMDAN					799.179	109.513
TETDAM03						294.729

**Taula 2.** MSQM per l'aproximació EASA pel conjunt de sis molècules seleccionades.

A la **taula 3** es mostra el nombre d'avaluacions de la MSQM realitzades per cada parell en el procés de recerca global, que no inclouen les avaluacions fetes en el refinament de la solució i que solen ser de l'ordre de cinc. La **taula 3** també inclou el nombre  $n_A^3 n_B^3$  que dona formalment la dependència de l'algorisme respecte a la mida de les molècules. La reducció en els valors realment calculats és expressada, d'una banda, per les exclusions dels condicionals, on s'han pres els valors d' $\epsilon_1$  i  $\epsilon_2$  com 5.5 i 1.5 ua, respectivament. D'altra banda tampoc s'han considerat els hidrògens en la recerca global degut a que les contribucions d'aquests són negligibles. Una justificació visual i intuïtiva d'aquesta restricció estaria en les representacions dels mapes de mesures de semblança quàntica que s'obtenen entre un àtom i una molècula [14].

	AZBENC01	TCYETY	HCCYHB	PYMDAN	TETDAM03
ACRDIN01	168 196 608	12 167 000	70 957 944	70 957 944	97 336 000
	<b>21 416</b>	<b>6 846</b>	<b>23 304</b>	<b>25 934</b>	<b>11 876</b>
AZBENC01		13 824 000	80 621 568	80 621 568	110 592 000
		<b>5 068</b>	<b>16 716</b>	<b>21 652</b>	<b>9 120</b>
TCYETY			5 832 000	5 832 000	8 000 000
			<b>5 556</b>	<b>8 928</b>	<b>2 040</b>
HCCYHB				34 012 224	46 656 000
				<b>20 256</b>	<b>7 200</b>
PYMDAN					46 656 000
					<b>8 716</b>

*Taula 3. En cursiva valor total de mesures  $n_A^3 n_B^3$  i en negreta el nombre real d'avaluacions de la MSQM, descomptant-ne les excloses en algun dels dos condicionals i els àtoms d'hidrogen.*

Finalment, i per tenir una referència del comportament del nou mètode de maximització presentat en aquest treball, s'ha fet un càlcul pel conjunt de sis molècules utilitzant un mètode d'optimització Monte Carlo acoblat amb un *simplex* [12]. S'ha emprat el programa *MolSimil-96*, amb les mateixes densitats empíriques utilitzades pels resultats de la **taula 2**. Per tenir uns resultats comparatius entre els dos mètodes de maximització, s'han realitzat tantes avaluacions de la MSQM com les que havien estat necessàries per obtenir els resultats de la **taula 2**. Com mostra la **taula 4**, amb el mètode de Monte Carlo més *simplex* sempre s'obtenen valors inferiors. En un dels parells comparats, el ACRDIN01-PYMDAN, es produeix una petita diferència entre els dos valors calculats que és deguda a la falta de refinament del mètode *simplex* per acabar d'arribar al màxim. En aquest cas es pot dir que s'ha obtingut el mateix màxim, però en la resta de parells estudiats s'ha arribat a màxims de valor inferior als trobats amb el nou mètode.

	ACRDIN01	AZBENC01	TCYETY	HCCYHB	PYMDAN	TETDAM03
ACRDIN01	463.765	202.485	150.567	216.680	200.281	96.492
AZBENC01		484.403	76.946	287.199	133.948	85.682
TCYETY			398.227	206.598	136.720	72.220
HCCYHB				6102.373	277.285	194.911
PYMDAN					799.179	79.817
TETDAM03						294.729

Taula 4. MSQM per l'aproximació EASA utilitzant el mètode Monte Carlo més simplex.

## 7. CONCLUSIONS

L'algorisme presentat resol de manera sistemàtica i determinística el problema del càlcul pràctic de les MSQM, que per la seva definició requereixen el coneixement del valor del màxim encobriment de les densitats electròniques de les molècules que s'han de comparar. Aquest tipus de mesura de semblança basada en les funcions de densitat de primer ordre connecta el concepte de molècula amb el d'un objecte tridimensional, i permet fer la recerca de la sobreposició molecular òptima sense necessitat d'haver d'utilitzar mètodes de combinatòria. És un mètode que resol un dels inconvenients de les tècniques de correspondències com és el de considerar les molècules com un conjunt de boles, i s'adapta a la distribució de probabilitats de la descripció quàntica. Aquest algorisme possiblement es pot estendre a altres tipus de mesures de semblança.

## Bibliografia

1. R.CARBÓ, L.LEYDA, M.ARNAU. *Int. J. Quantum Chem* 1980, 17, 1185.
2. R.CARBÓ I C.ARNAU a F.G. DE LAS HERAS, S.VEGA (ed.), *Molecular engineering: a general approach to QSAR*, in de las Heras, F.G., and Vega, S. (ed.), *Medicinal Chemistry Advances*. Pergamon Press, Oxford, 1981, Vol. 85.
3. R.CARBÓ, L.DOMINGO. *Int. J. Quantum Chem.* 1987, 23, 517.
4. R.CARBÓ, B.CALABUIG. *Comp. Phys. Commun.* 1989, 55, 117.
5. R.CARBÓ, B.CALABUIG. *J. Mol. Struct. (Theochem)* 1992, 254, 517.
6. R.CARBÓ, B.CALABUIG. *J. Chem. Inf. Comput. Sci.* 1992, 32, 600.
7. R. CARBÓ, B. CALABUIG, E. BESALÚ, A. MARTÍNEZ. *Molecular Engineering* 1992, 2, 43
8. R.CARBÓ, E.BESALÚ, B.CALABUIG, L.VERA. *Adv.Quant.Chem.* 1994, 25, 253.
9. *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, R. CARBÓ (ed.). Kluwer Academic Publishers, 1995
10. P. CONSTANS, R. CARBÓ, *J. Chem. Inf. Comput. Sci.*, 1995, 35, 1046.
11. R. CARBÓ, E. BESALÚ, L. AMAT, X. FRADERA, *J. Math. Chem.* 1995, 18, 237

12. LL. AMAT, E. BESALÚ, R. CARBÓ, X. FRADERA. *Scientia Gerundensis*, 1995, 21, 127.
13. P. CONSTANS, LL. AMAT, X. FRADERA i R. CARBÓ. *Quantum Molecular Similarity Measures and the Atomic Shell Approximation*, a *Proceedings of the 2nd. Girona Seminar on Molecular Similarity*, juliol 1995, R. Carbó and P.G. Mezey (ed.), *Advances in Molecular Similarity*. Vol 1, JAI PRESS INC. Greenwich (Conn.), en premsa.
14. R. CARBÓ, E. BESALÚ, LL. AMAT i X. FRADERA, *Quantum Molecular Similarity Measures: Concepts, Definitions and Applications*, a *Proceedings of the 2nd. Girona Seminar on Molecular Similarity*, juliol 1995, R. Carbó and P.G. Mezey (ed.), *Advances in Molecular Similarity*. Vol 1, JAI PRESS INC. Greenwich (Conn.), en premsa.
15. a) M. BORN i J.R. OPPENHEIMER, *Annln. Phys.* 1927, 84, 457.  
b) M. BORN i K.HUANG, *Dynamical theory of crystal lattices*, Clarendon Press, Oxford 1954  
c) H.C. LONGUET-HIGGINS. *Adv. in Spectrosc.* 1961, 2, 429.
16. a) A.D. MCLACHLAN. *Acta Cryst.* 1982, A38, 871.  
b) D.J. DANZIGER, P.M. DEAN, J. THEOR. *Biol.* 1985, 116, 215.  
c) M.T. BARAKAT i P.M.DEAN, *J. Computer-Aided Mol. Design* 1990, 4, 295.
17. a) S. KIRKPATRICK, C.D. GELATT i M.P. VECCHI, *Science* 1983, 220, 671.  
b) S. KIRKPATRICK, *J. Statistical Physics*, 1984, 34, 975.
18. W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING i B.P.FLANNERY. *Numerical recipes the art of scientific computing*, Cambridge University Press, 1992 (387-449).
19. P. CONSTANS, LL. AMAT i R. CARBÓ, *Towards a Global Maximization of the Similarity Function: The Superposition of two Molecules*, *J. Comp. Chem.* (en premsa).
20. *3D Search and Research using the Cambridge Structural Database*. ALLEN, F.H., KENNARD, O. *Chemical Desing Automation News* (1993), 8, 1 31-37.
21. M.J. FRISCH, G.W. TRUCKS, M. HEAD-GORDON, P.M.W. GILL, M.W. WONG, J.B. FORESMAN, B.G. JOHNSON, H.B. SCHLEGEL, M.A. ROBB, E.S. REPLOGEL, R. GOMPERTS, J.L. ANDRES, K. RAGHAVACHARI, J.S. BINKLEY, C. GONZALEZ, R.L. MARTIN, D.J. FOX, D.J. DEFREES, J. BAKER, J.J.P. STEWART, and J.A. POPL, *GAUSSIAN 92-DFT, Revision G.1*, Gaussian Inc., Pittsburgh, PA, 1992
22. P. CONSTANS i R. CARBÓ, *ASAC Calculations* v. 2.0 (Institut de Química Computacional, Universitat de Girona, CAT 1995)
23. LL. AMAT, E. BESALÚ i R. CARBÓ. *MolSimil-96* (Institut de Química Computacional, Universitat de Girona, CAT 1996)