

SEMBLANÇA MOLECULAR: REPRESENTACIÓ n-DIMENSIONAL D'UN CONJUNT DE MOLÈCULES

Ramon Carbó, Blanca Calabuig & Antoni Martínez

Laboratori de Química Computacional. Estudi de Ciències. Estudi General de Girona.
Plaça de l'Hospital, 6. 17071 Girona.

RESUM

Es presenta una sèrie de conceptes de semblança molecular quàntica i uns procediments associats de càlcul i representació gràfica dels resultats. Donada una sèrie de molècules es poden obtenir diversos tipus de gràfics que mostren les relacions entre elles. Com a exemple d'aplicació d'aquest procés s'estudia una família de drogues antitumorals.

RESUMEN

Se presentan conceptos de semejanza molecular cuántica y una serie de procedimientos de cálculo y representación gráfica de los resultados. A partir de una serie de moléculas se pueden obtener varios tipos de gráficos que muestran las relaciones existentes entre ellas. A modo de ejemplo de todo el proceso se estudia una familia de drogas antitumorales.

ABSTRACT

This work presents some concepts and computer techniques developed in our laboratory. The quantum molecular similarity can be a powerful tool to study the relationships between molecules. The first part is an introduction to quantum molecular similarity. It shows the way to compute quantum similarity measures and plot some types of graphics. The graphics are useful to represent simultaneously a set of quantum objects and their relationships. The second part is an example showing all the practical aspects discussed in the theoretical section. The molecular set used in this study is a family of antitumoral agents.

Key words: Density Functions; LCAO-MO Theory n-dimensional Graphs Quantum Molecular Similarity; Quantum Molecular Taxonomy; Representation of n-dimensional Clouds of Points; Structure Activity Relationships.

INTRODUCCIÓ

D'acord amb els principis de la química quàntica, tota la informació sobre un sistema està continguda en les funcions densitat que té associades, calculades a partir de la funció d'ona del sistema. Aquestes funcions es poden fer servir per calcular mesures de semblança molecular. En aquest treball es fan servir les funcions densitat de primer ordre però es poden utilitzar funcions d'ordre superior i, fins i tot, altres funcions moleculars derivades, com el potencial electrostàtic. Aquest treball està dividit en dues parts: la primera és la base matemàtica i teòrica de la semblança molecular quàntica, i la segona estudia, com a exemple, una família de molècules amb propietats antitumorals: les 2-fenilbenzimidazole-4-carboxamides substituïdes (1).

BASE MATEMÀTICA

Mesures de semblança

Es pot simbolitzar la distància euclídia entre dues funcions F_a i F_b , suposades aquí reals, com $D(F_a, F_b)$ i el cosinus de l'angle entre dues funcions com $C(F_a, F_b)$. A partir d'això es pot definir la següent mesura, que té caràcter de distància:

$$D_c(F_a, F_b) = 2 \arccos [C(F_a, F_b)] \quad (1)$$

Es pot veure que, quan les funcions són iguals, ($F_a = F_b$), $D = 0$, $C = 1$ i $D_c = 0$. Quan les funcions són completament diferents, $D \rightarrow +\infty$, $C \rightarrow 0$, $D \rightarrow \pi$.

El problema és definir $D(F_a, F_b)$ o $C(F_a, F_b)$ coneixent la natura de les funcions. Si l'espai funcional al qual pertanyen les funcions F_a i F_b té definit un producte escalar, el problema ja estarà resolt en teoria. El producte escalar es pot construir en general fent servir un operador hermitic definit positiu Z mitjançant la integral de volum

$$(F_a | Z | F_b) = \int F_a Z F_b dV \quad (2)$$

Si Z és l'operador unitat, la fórmula anterior es transforma en la integral

$$(F_a | I | F_b) = (F_a | F_b) = \int F_a F_b dV \quad (3)$$

Quan el producte escalar està definit amb la forma de l'equació 2, la distància euclídia entre dues funcions es pot calcular com

$$D(F_a, F_b) = [(\Omega | Z | \Omega)]^{1/2} \quad (4)$$

on $\Omega = |F_a - F_b|$. Així mateix, el cosinus de l'angle entre les funcions pren la forma

$$C(F_a, F_b) = |(F_a | Z | F_b)| / [(F_a | Z | F_a)(F_b | Z | F_b)]^{1/2} \quad (5)$$

D'aquesta manera queden definides les dues operacions bàsiques en mesures de semblança molecular.

Relacions funció-objecte

Es pot relacionar l'ordre generat en un conjunt de funcions amb l'ordre en els elements d'algun conjunt d'objectes connectats a les funcions.

Suposem que en un conjunt d'objectes cada element pot connectar-se a un altre d'un conjunt de funcions. Les mesures de semblança fetes sobre aquestes funcions poden, potencialment, ordenar-les d'alguna manera. Un cop ordenades les funcions, es pot induir ordre en el conjunt dels objectes mitjançant la connexió objecte-funció.

En un conjunt ordenat d'objectes-funcions es poden fer agrupacions de les propietats relacionades amb els objectes. Per tant, es poden obtenir relacions objecte-propietat analitzant els resultats de les consideracions anteriors.

Conjunts difusos i mesures de semblança

La teoria dels conjunts difusos dóna la base necessària per consolidar tot el procés explicat a l'apartat anterior.

Un conjunt difús és senzillament un conjunt clàssic associat amb una funció de pertinença $M(x)$, la qual, aplicada sobre un element x del conjunt, dóna un valor en l'interval $[0,1]$. Els conjunts clàssics, sota aquesta definició, són un cas particular de conjunts difusos: quan un element pertany a un conjunt clàssic C , la funció de pertinença dóna 1, i si no hi pertany, la funció dóna 0. La difuminació apareix quan els elements del conjunt C estan definits de tal manera que $0 < M(x) < 1$.

Es pot construir, seguint aquesta definició, el producte cartesià d'un conjunt difús, amb funcions de pertinença aplicades sobre parelles dels elements del conjunt. Es pot considerar el valor del cosinus de l'angle entre dues funcions, tal com s'ha definit abans, com una possible funció de pertinença a aquest conjunt: una parella $\{F_a, F_b\}$ és membre de ple dret del conjunt si $C(F_a, F_b) \approx 1$; en canvi un element és marginal quan $C(F_a, F_b) \approx 0$.

Ordenació d'un conjunt d'objectes

Com que a cada funció se li pot associar, un a un, un objecte d'un conjunt ben definit, com molècules, estats o orbitals moleculars, la mesura C i l'ordre que ha induït sobre el conjunt de funcions es pot estendre al conjunt d'objectes. Per tant, ordenant el conjunt de funcions també s'ordena el conjunt d'objectes associats.

Donats dos conjunts, un de funcions

$$F = \{ F_1, F_2 \dots \}$$

i un altre d'objectes

$$O = \{ o_1, o_2 \dots \}$$

es pot establir una correspondència un a un

$$K = [F \leftrightarrow O \mid \{ F_1 \leftrightarrow o_1, F_2 \leftrightarrow o_2, \dots \}]$$

Si s'estableix una regla d'ordre R al conjunt F tal que

$$R[F] = \{ F_1 < F_2 < \dots \},$$

on $\{ < \}$ representa alguna convenció d'ordenació difusa, l'ordre a F es pot estendre a O mitjançant la mateixa regla

$$R[O] = \{ o_1 < o_2 < \dots \}$$

Es pot unir qualsevol nova funció i l'objecte relacionat amb ella als conjunts ja existents. Sols s'hauran de fer les mesures de semblança entre la nova funció i les antigues del conjunt. Si afegim la funció f al conjunt de funcions F seguint la regla R, haurem de trobar un índex k tal que

$$R[F \cup \{ f \}] = \{ \dots < F_k < f < F_{k+1} < \dots \}$$

Com que la nova funció f es correspon amb un nou objecte w , $f \leftrightarrow w$,

$$R[O \cup \{ w \}] = \{ \dots < o_k < w < o_{k+1} < \dots \}$$

MESURES DE SEMBLANÇA SOBRE FUNCIONS DE LA MECÀNICA QUÀNTICA

Mesures de semblança entre funcions densitat

Tractant les molècules com objectes, es pot associar una funció a cada molècula mitjançant la teoria quàntica. Donada una estructura molecular es pot calcular la funció d'ona d'algun estat de sistema i a partir d'aquí, la funció densitat de primer ordre. Aquestes funcions permeten la construcció de les mesures definides anteriorment (distància [equació 4] i cosinus de l'angle [equació 5]) mitjançant el producte escalar (equacions 2 i 3).

Per aplicar les definicions de semblança molecular sobre funcions densitat no-més caldrà considerar les parelles $\{F_a, F_b\}$ de les definicions anteriors com parelles $\{P_a, P_b\}$ de funcions densitat. Un operador hermític definit positiu que pot servir en la definició general (equació 2) és l'operador de Coulomb, r^{-1} .

Mesures de semblança intermolecular

Si s'agafa el problema des d'un punt de vista *ab initio* en el context LCAO-MO, les funcions densitat es poden escriure com

$$P_a = \sum_{\alpha} \sum_{\beta} D_{a,\alpha,\beta} \Phi_{\alpha} \Phi_{\beta} \quad (6)$$

on $D_{a,\alpha,\beta}$ són els elements de la matriu densitat de la molècula a en l'estat que s'estudia i $\{\Phi_{\alpha}\}$ és el conjunt de les funcions de base dels orbitals atòmics suposats reals. Es pot dir el mateix per a la molècula b.

A partir de l'equació 6 i de P_b es pot construir un producte escalar semblant al definit anteriorment a l'equació 2

$$(P_a | Z | P_b) = \sum_{\alpha} \sum_{\beta} \sum_{\mu} \sum_{\sigma} D_{a,\alpha,\beta} D_{b;\mu,\sigma} (\alpha\beta | Z | \mu\sigma) \quad (7)$$

on es fa necessari calcular la integral

$$(\alpha\beta | Z | \mu\sigma) = \int \Phi_{\alpha} \Phi_{\beta} Z \Phi_{\mu} \Phi_{\sigma} dV \quad (8)$$

Si utilitzem l'operador Unitat $Z = I$, l'equació 7 queda com una integral de solapament entre quatre orbitals atòmics en lloc dels dos que són habituals (2).

Per a la majoria d'aplicacions l'expressió 7 és massa complicada. Podem utilitzar les aproximacions CNDO (3, 4, 5) per arribar a una forma més tractable computacionalment

$$(P_a | Z | P_b) = \sum_{i \in a} \sum_{j \in b} Q_{a,i} Q_{b,j} (|S_{a,i}|^2 |Z| |S_{b,j}|^2) \quad (9)$$

on $Q_{a,i}$ és la població atòmica de Mulliken (6) a l'àtom i de la molècula a, i $\{|S_{a,i}|^2\}$ és el quadrat de la funció de l'orbital de tipus Slater de simetria s amb el nombre quàntic principal de l'última capa electrònica de l'àtom i de la molècula a. Es pot dir el mateix per a la molècula b i l'àtom j. L'equació 9 ha estat deduïda de l'equació 7 tot aplicant les aproximacions CNDO (3, 4, 5). L'equació 9 té l'avantatge que es poden substituir $Q_{a,i}$ i $Q_{b,j}$ per paràmetres de càrrega de qualsevol origen, sempre que se segueixi el mateix mètode de càlcul per a totes les molècules. En aquest treball s'ha fet servir l'expressió 9 amb Z sent l'operador unitat I. Per calcular les poblacions atòmiques s'ha utilitzat un procediment EHT modificat. El programa EHT-90 ha estat desenvolupat precisament per a aquest propòsit.

Detalls de programació

Totes les mesures de semblança d'aquest treball han estat realitzades en ordinadors compatibles IBM-PC sota el sistema operatiu MS-DOS mitjançant el procediment MOLSIMIL-90 (7), escrit en Fortran-77 i transportable a altres màquines i sistemes operatius. En aquest programa s'ha implementat l'equació 9, i pot treballar amb molècules construïdes amb àtoms de tota la taula periòdica. Un altre programa auxiliar interactiu crea un fitxer de control que condueix tot el procés.

Les mesures quàntiques de semblança molecular depenen de les posicions relatives de les dues molècules que s'han de comparar, suposant que les coordenades moleculars es mantinguin fixes. En aquest cas, el producte escalar de l'equació 9 depèn de sis paràmetres: les tres components de la posició relativa del centre de coordenades d'una molècula respecte del de l'altra i dels tres angles de rotació al voltant dels tres eixos cartesianes de la posició d'una molècula respecte de l'altra. Es pot resoldre l'arbitrarietat del centre de coordenades agafant com a origen comú el centre de càrrega de les dues molècules comparades. Els tres paràmetres restants es poden optimitzar a canvi d'un temps molt gran de càlcul. Els resultats d'aquest treball s'han obtingut d'aquesta manera, però.

DETALLS PRÀCTICS

Tractament dels resultats

Donada una mesura de semblança entre funcions densitat, es pot definir una matriu simètrica Z , els elements de la qual siguin valors del cosinus de l'angle entre dues funcions densitat tal com es descriu a l'equació 5. Per construir aquesta matriu, sols cal conèixer la matriu dels productes escalars calculada mitjançant l'equació 2 o la 3 aplicada sobre funcions densitat

$$Z_{ij} = (i | j) / [(i | i) (j | j)]^{1/2} \quad (10)$$

on

$$(i | j) = \int P_i Z P_j dV \quad (11)$$

La matriu Z tal com està definida a l'equació 10 es pot considerar com la matriu de la mètrica del subespai generat pel conjunt ben definit de funcions de base de l'equació 6.

Un cop coneguda la matriu Z , es pot obtenir força informació dels arbres i grafs que es poden construir. En aquest treball s'utilitza la tècnica de l'Arbre de Kruskal (arbre d'expansió mínima) (2,8,9), i una tècnica de representació de grafs en tres dimensions que hem anomenat Grafs de Veïns Més Propers, desenvolupada pel nostre grup de treball (3).

Representació dels resultats

La representació gràfica de relacions de semblança quàntica entre funcions densitat presenta problemes especials a causa de la natura infinit-dimensional dels espais als quals pertanyen aquestes funcions. Les tècniques de dibuix que es presenten es poden aplicar a relacions de semblança derivades de les representacions tant finit com infinit-dimensional de la informació molecular. Aquestes representacions inclouen Arbres de Kruskal, representacions n-dimensional basades en components principals i gràfics en n-dimensions, les quals poden programar-se amb l'ajut de les nombroses llibreries de funcions gràfiques disponibles al mercat (10).

Arbres de Kruskal i Grafs de Veïns Més Propers

Els Arbres de Kruskal uneixen les molècules més semblants a una de donada de manera que com a mínim cada una tingui un lligam. No hi ha límit superior del nombre de connexions per cada molècula. Les connexions cícliques no es permeten, de manera que es conservi l'estructura d'arbre. Procedint d'aquesta manera és possible representar les molècules com vèrtexs d'un hipercub, i apareix una estructura en grups que reuneix totes les molècules amb propietats semblants. Els Grafs de Veïns Més Propers també recullen i uneixen les molècules més semblants a una de donada, de manera que cada una estigui unida a un nombre mínim de molècules semblants, sense estar connectada a més d'un nombre màxim. El nombre màxim i mínim de connexions es poden escollir a voluntat o bé fer-ho de manera automàtica. A diferència dels Arbres de Kruskal, es permeten connexions cícliques (d'aquí el nom de graf) i es pot fixar el nombre màxim i mínim de lligams.

Segons la manera d'escollir les distàncies mínimes per tal d'establir un lligam, es poden construir dos tipus de grafs de veïns més propers: els convencionals i els mínims. El graf mínim de veïns més propers es construeix ordenant totes les distàncies possibles entre parelles de molècules i escollint les mínimes per establir lligams. Els convencionals, en canvi, es construeixen buscant per a una molècula donada la distància mínima amb una altra.

Un programa anomenat ND-GRAPH construeix aquests arbres i grafs interactivament, i permet obtenir fàcilment diagrames de l'estructura, altrament molt complicada, dels grafs i dels arbres. Utilitza la matriu **Z** calculada pel programa MOLSIMIL-90 com a dades d'entrada.

Representació n-Dimensional

Juntament amb la tècnica anterior, s'ha desenvolupat un altre sistema de representació del conjunt de molècules. Aquesta tècnica es basa en el fet de considerar cada molècula com un objecte n-dimensional gràcies als vectors propis de la matriu **Z**, els quals poden servir com coordenades moleculars.

Si es fa que la i -èssima columna o filera de la matriu dels vectors propis de Z representi la i -èssima molècula, es poden representar les n molècules en l'espai n -dimensional. Quan n és molt gran, l'agrupament del núvol de punts resultant és difícil de veure. Si s'escull adequadament un subconjunt de vectors propis, es pot reduir considerablement el nombre de dimensions sense perdre gaire informació, i guanyar claredat en la representació. El nostre grup de treball ha descrit i posat a punt un mètode per fer aquesta selecció (9).

Hom pot escollir una mesura de distància entre els punts n -dimensionals. Per fer l'algorisme el més ràpid possible s'ha escollit la mesura de distància de Minkowski, definida per a qualsevol parella de vectors n -dimensionals $\{v(n)_i, v(n)_j\}$

$$M [v(n)_i, v(n)_j] = \sum_{\alpha=1}^n |v(n)_{\alpha i} - v(n)_{\alpha j}| \quad (12)$$

Es pot definir la dispersió total d'un conjunt de n punts d'un núvol de n -dimensions com

$$S(n, n) = \sum_{i, j \in \{n\}} M [v(n)_i, v(n)_j] \quad (13)$$

Per obtenir una projecció del núvol en un subespai de dimensió menor, s'escull un espai m -dimensional amb $m < n$. La mesura de la dispersió per a n punts en aquest subespai es transforma en

$$S(n, m) = \sum_{i, j \in \{m\}} M [v(m)_i, v(m)_j] \quad (14)$$

Si escollim les m direccions tals que

$$|S(n, m) - S(n, n)| = \text{mínim} \quad (15)$$

la pèrdua d'informació deguda a la reducció dimensional també serà mínima.

El núvol resultant es pot fer girar i transformar un altre cop per poder-lo visualitzar sobre un pla de manera que la nova dispersió $S(n, 2)$

$$|S(n, 2) - S(n, m)| = \text{mínim} \quad (16)$$

D'aquesta manera la informació representada serà màxima. Un programa anomenat ND-CLOUD realitza totes aquestes transformacions manualment o automàtica, utilitzant com a entrada la matriu Z calculada amb el programa MOLSIMIL-90.

Taula 1. N-(N,N-dimetil-2-aminoetil) 2-fenilbenzimidazole-4-carboxamides substituïdes (1): propietats estudiades.

Núm.	Substituent	Pf (°C)	Activitat	
			P388/W	P388/A
1	2' -CH ₃	230-234	22,9	32,7
2	2' -OCH ₃	220-223	8,0	12,5
3	2' -Cl	203-205	16,0	22,5
4	2' -aza	237-240	19,0	30,2
5	3' -CH ₃	161-164	9,5	14,4
6	3' -OCH ₃	184-185	11,7	24,3
7	3' -Cl	155-156	5,9	8,0
8	3' -aza	235-237	11,7	35,2
9	4' -CH ₃	177-181	7,8	12,5
10	4' -OCH ₃	150-154	5,8	12,7
11	4' -Cl	154-155	4,0	7,7
12	4' -aza	266-270	6,6	15,9
13	4' -fenil	253-259	1,3	2,0
14	4' -NHAc	275-280	32,5	> 40
15	4' -N(CH ₃) ₂	213-217	2,6	7,0
16	5 -CH ₃	255-259	18,0	> 40
17	6 -CH ₃	188-189	21,0	22,4
18	7 -CH ₃	261-264	18,6	29,3
19	2', 3' -benzil	160-170	8,9	8,0
20	3', 4' -benzil	181-183	2,9	4,4
21	2', 3' -(-OCH ₂ -)	225-230	2,7	4,6
22	2', 3' -(OCH ₃) ₂	215-220	12,5	7,7
23	2', 3', 4' -(OCH ₃) ₂	184-187	7,8	23,3

Superfícies tridimensionals

Per tenir una idea general de les coordenades de tots els n punts del núvol es pot dibuixar una superfície tridimensional agafant la matriu de coordenades ortogonals com els valors d'una funció de dues variables. Els valors positius es veuran en el gràfic com muntanyes, i les valls seran valors negatius. El programa ND-SURF, que té unes característiques semblants als programes ND-GRAPH i ND-CLOUD, és l'encarregat d'aquesta feina.

ESTUDI D'UNA FAMÍLIA DE MOLÈCULES

En aquest apartat es farà l'estudi d'una família de drogues antitumorals: les N-(N,N-dimetil-2-aminoetil) 2-fenilbenzimidazole-4-carboxamides (1).

El primer pas ha estat definir la geometria de les molècules a estudiar mitjançant un programa de disseny molecular (11). Dins el mateix programa s'ha op-

timitzats l'estructura de les molècules amb un procediment tipus mecànica molecular. Un cop totes les molècules han estat definides, els fitxers de sortida del programa anterior es transformen, mitjançant un simple interface, als fitxers d'entrada del programa EHT que s'ha utilitzat per calcular les funcions densitat. Aquest programa, EHT-90, l'ha desenvolupat en Fortran-77 el nostre grup de treball precisament per a aquesta tasca.

Els fitxers de sortida del programa EHT serveixen directament de fitxers d'entrada del programa MOLSIMIL-90 (7). Aquest programa fa les mesures de semblança entre les funcions densitat de les diferents molècules.

La sortida del programa MOLSIMIL-90 s'usa d'entrada per el conjunt de programes de representació gràfica ND-CLOUD, ND-GRAPH i ND-SURF, juntament amb les dades de les propietats, recollides a la taula 1. En aquesta taula hi figura el punt de fusió (Pf) i l'activitat biològica enfront dos tipus de cèl·lules. L'activitat es refereix a la concentració de droga en μM necessària per inhibir el creixement in vitro de les cèl·lules en un 50%. Les cèl·lules estudiades són les Wild-Type P388 murine leukemia (P388/W) i les Amsacrine resistant P388 (P388/A).

Representacions n-dimensionals

El programa ND-CLOUD permet representar gràficament un núvol de punts n-dimensionals. El nombre de dimensions de l'espai inicial és igual al nombre de molècules (23 en el nostre cas). Aquest nombre de dimensions el podem reduir a voluntat escollint les components principals, es a dir, conservant les dimensions en les quals la dispersió calculada mitjançant la fórmula 14 és màxima. D'aquesta manera el subespai creat conservarà el màxim d'informació. Posteriorment es torna a projectar el núvol en un pla per poder-lo representar a la pantalla o sobre paper.

La figura 1 representa dos gràfics de la mateixa projecció del núvol de punts, amb els símbols representant els punts de fusió: a) les etiquetes són els punts de fusió, i b) el número de molècula. La projecció escollida és la que dona el màxim valor per la dispersió. Els nombres dels eixos són les components principals escollides.

En la figura 2 es comparen les propietats estudiades entre elles. a) mateixa vista que la figura 1, però comparant el punt de fusió (símbols) i l'activitat enfront les cèl·lules P388/W (etiquetes). b) compara les activitats enfront les cèl·lules P388/A (símbols) i P388/W (etiquetes).

La figura 3 es una altra projecció del núvol de punts respecte les components principals 3 i 2. Els símbols representen l'activitat enfront les cèl·lules P388/W, i les etiquetes el número de molècula.

En les figures 1, 2 i 3 la dimensió del subespai és 5.

La figura 4 compara l'activitat de la droga respecte les cèl·lules P388/A (símbols) i el número de molècula (etiquetes) de les projeccions en tres subespais: a) espai de dimensió 5; b) espai de dimensió 4, i c) espai de dimensió 3.

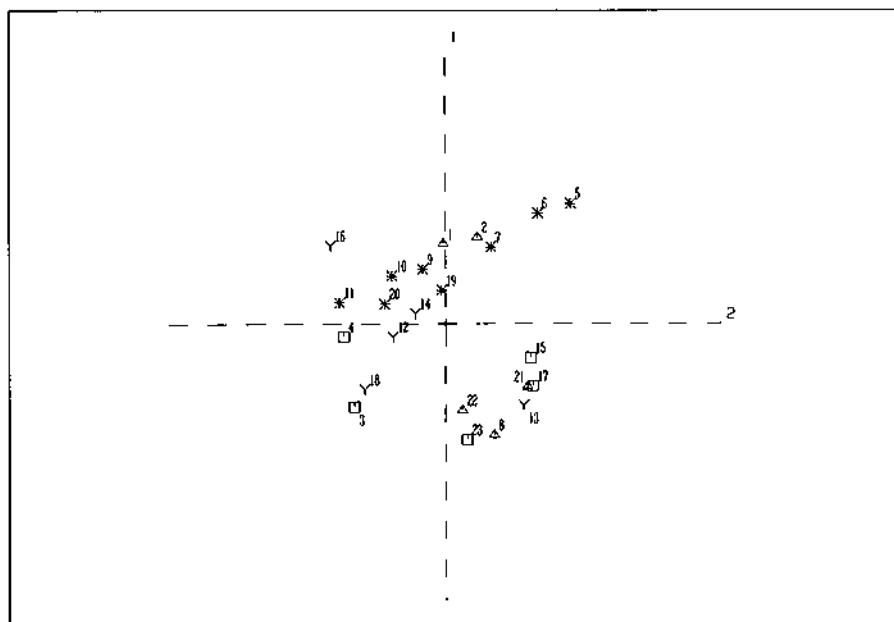
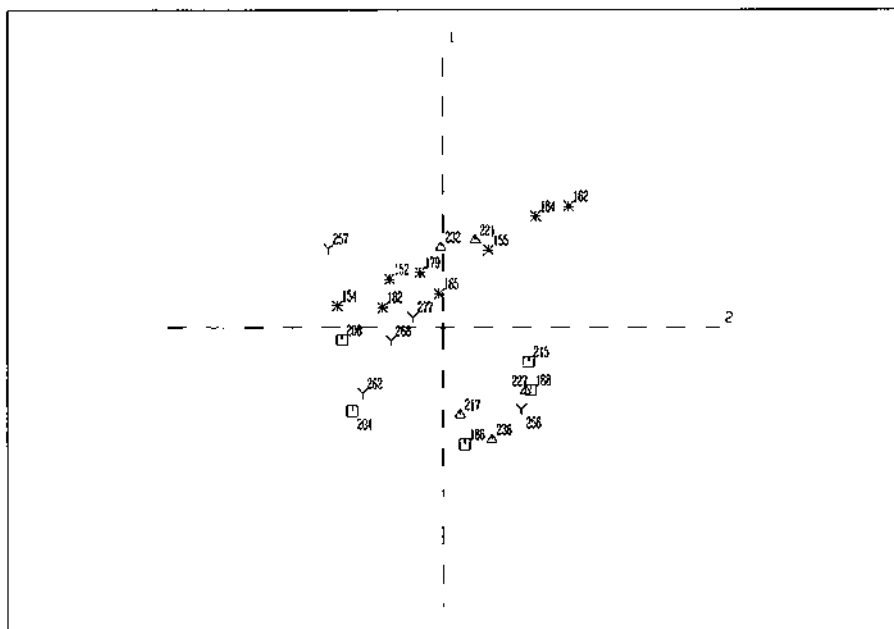


Figura 1. Projectió del núvol on els símbols representen el punt de fusió. Les etiquetes representen a) el punt de fusió i b) el número de molècula. Punt de vista amb dispersió màxima.

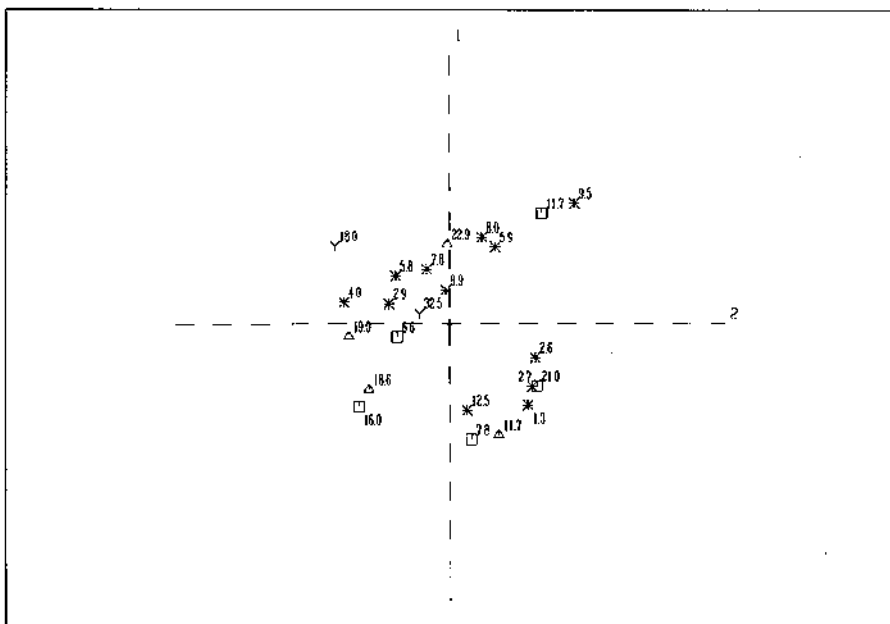
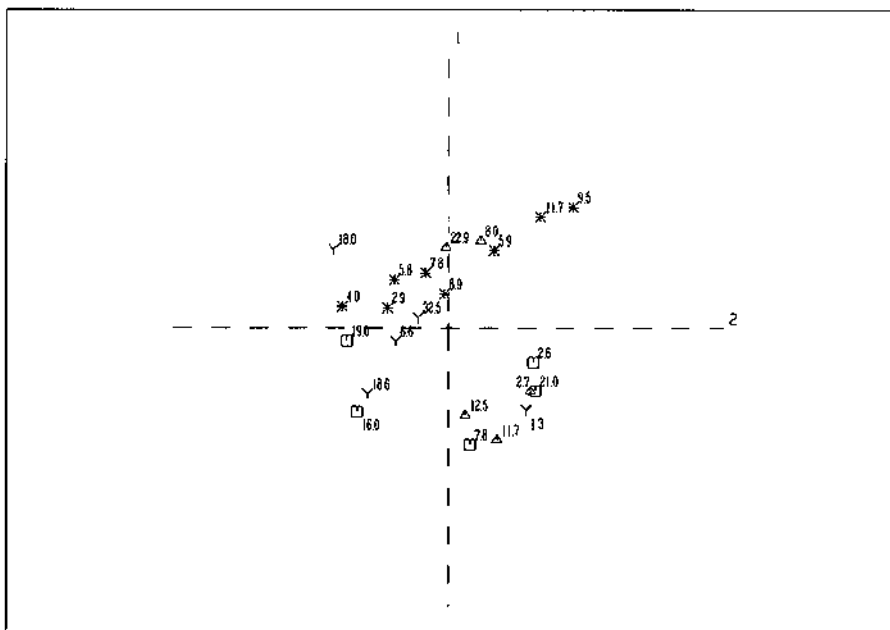


Figura 2. Projectió que mostra les relacions entre propietats: a) els símbols són el punt de fusió i les etiquetes són l'activitat enfront de les cèl·lules P388/W; b) activitats enfront les cèl·lules P388/A (símbols) i P388/W (etiquetes).

Grafs i arbres

El programa ND-GRAPH representa els grafs i arbres posant les molècules als vèrtexs d'un hiper-cub n-dimensional, amb prou vèrtexs per encabir-les totes. Les dimensions del subespai es trien per tal que es compleixi aquesta condició. En tots els diagrames que es presenten en aquest apartat, la dimensió del subespai és 5, on l'hipercub té 32 vèrtexs (sols hi ha 23 molècules per representar).

Per representar els diagrames es projecta l'hipercub en el pla, després d'haver-lo fet girar fins que s'assoleix un bon punt de vista. En tots els dibuixos s'ha establert una partició de les propietats de la família: els valors més grans de la propietat es representen per una Y, i per ordre decreixent, per un triangle, un quadrat i un asterisc, el qual correspon als valors més baixos de la propietat representada.

La figura 5 representa dos arbres de Kruskal: *a*) calculat a partir de la distància euclídia donada per la fórmula 4, i *b*) a partir de la distància de correlació derivada de les fórmules 1 i 5. La propietat representada pels signes és el punt de fusió mitjà, i les etiquetes, el número de la molècula.

La figura 6 representa també els punts de fusió mitjans de les molècules en dos grafs de veïns més propers, amb tres lligams com a màxim i un com a mínim; *a*) ha estat calculat amb distàncies euclídiades i *b*) amb distàncies de correlació.

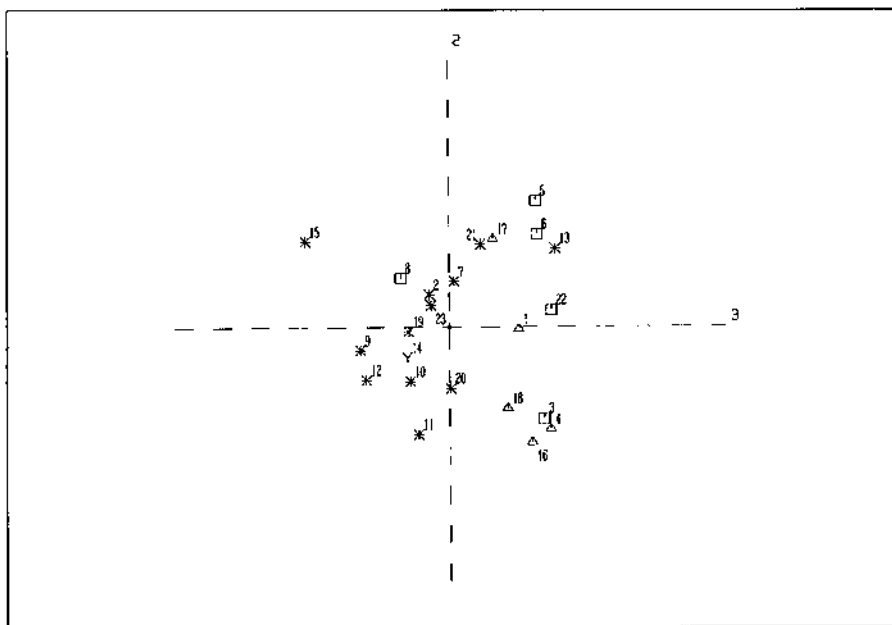


Figura 3. Projectió del núvol: els símbols són l'activitat amb les cèl·lules P388/W i les etiquetes el número de molècula.

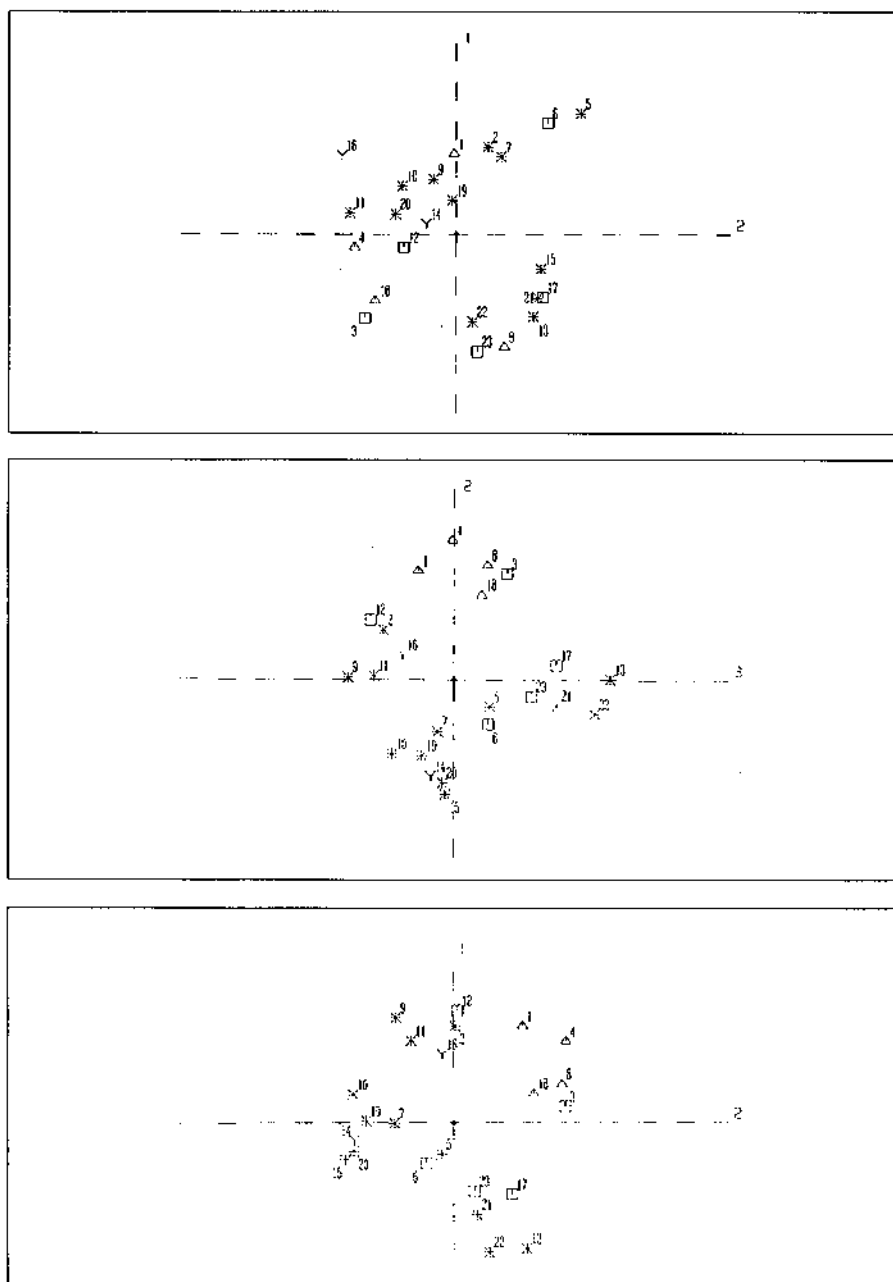


Figura 4. Projeccions del núvol a diversos subespais a) de 5 dimensions, b) de 4 dimensions i c) de 3 dimensions, posteriorment projectats als subespais de 2 dimensions indicats. Els símbols són l'activitat amb les cèl·lules P388/A i les etiquetes el número de molècula.

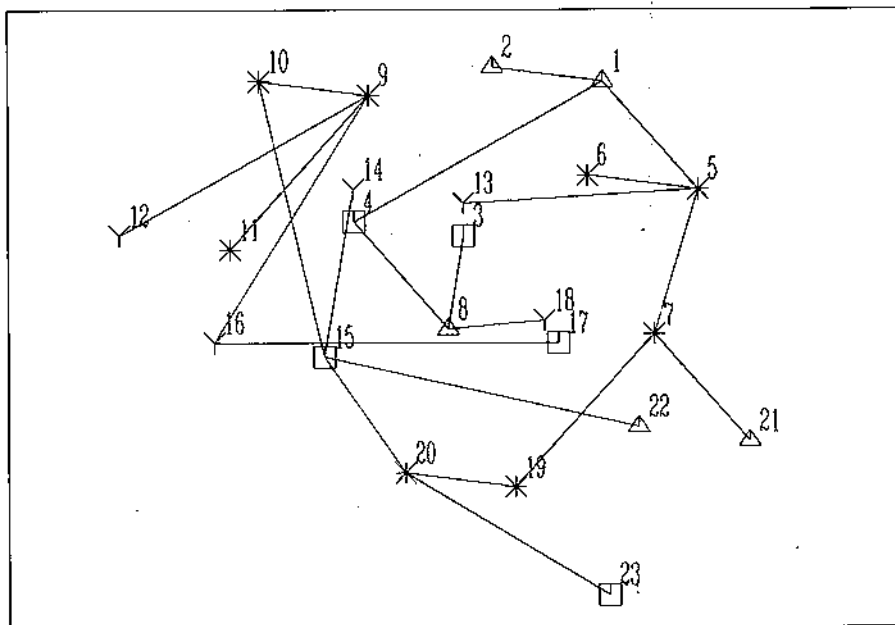
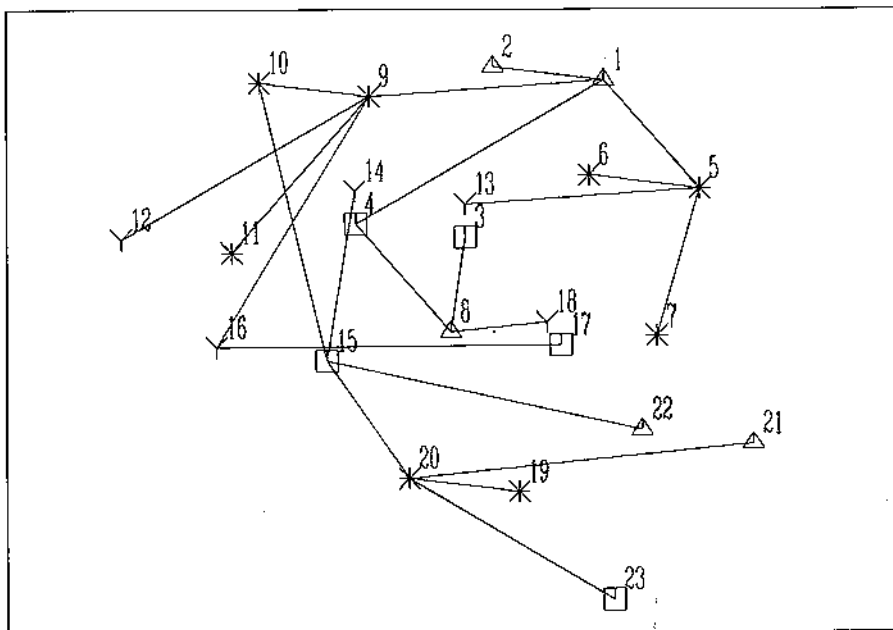


Figura 5. Arbres de Kruskal: a) utilitzant distàncies euclídiades i b) utilitzant distàncies de correlació. Els símbols representen el punt de fusió i les etiquetes, el número de molècula.

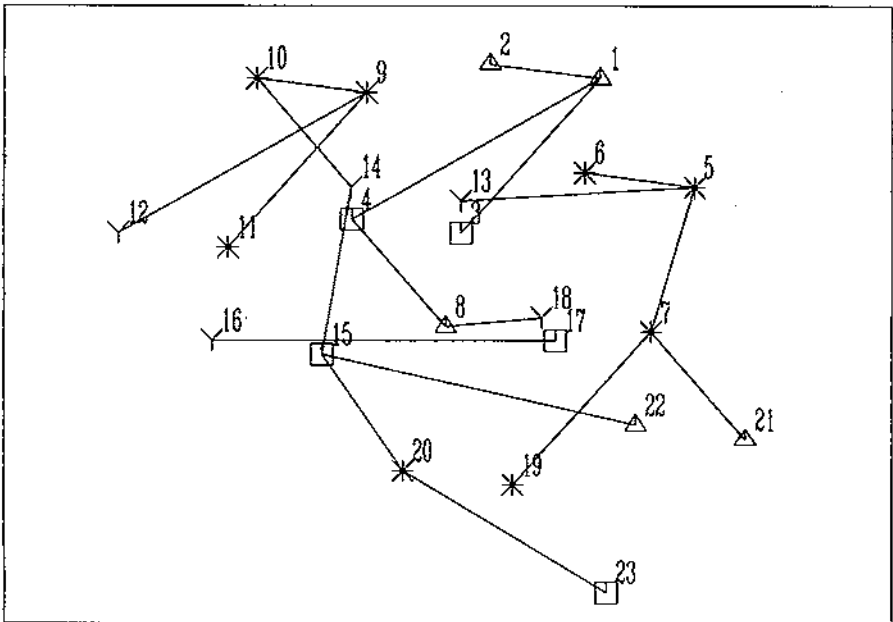
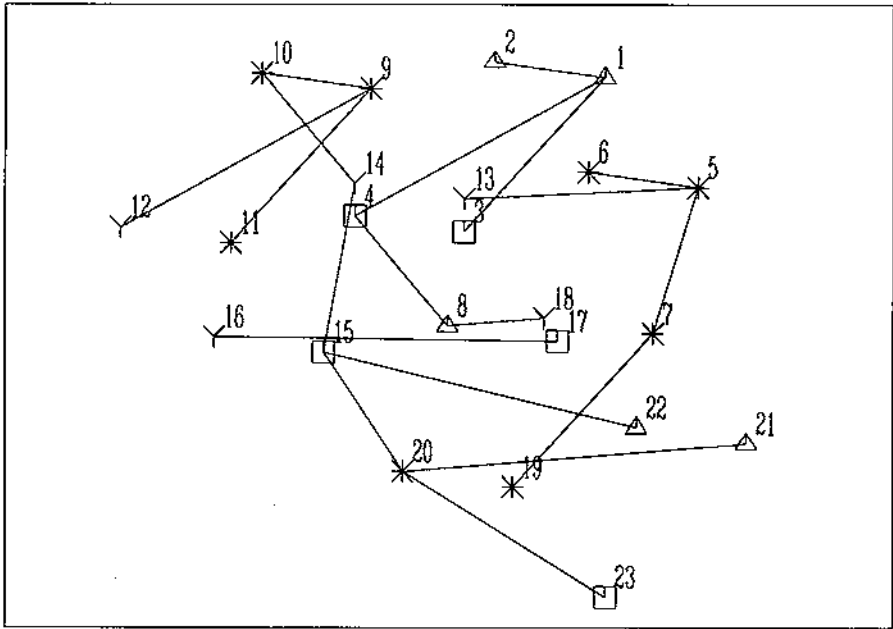


Figura 6. Gracs de veïns més propers: a) amb distàncies euclídiades i b) amb distàncies de correlació.

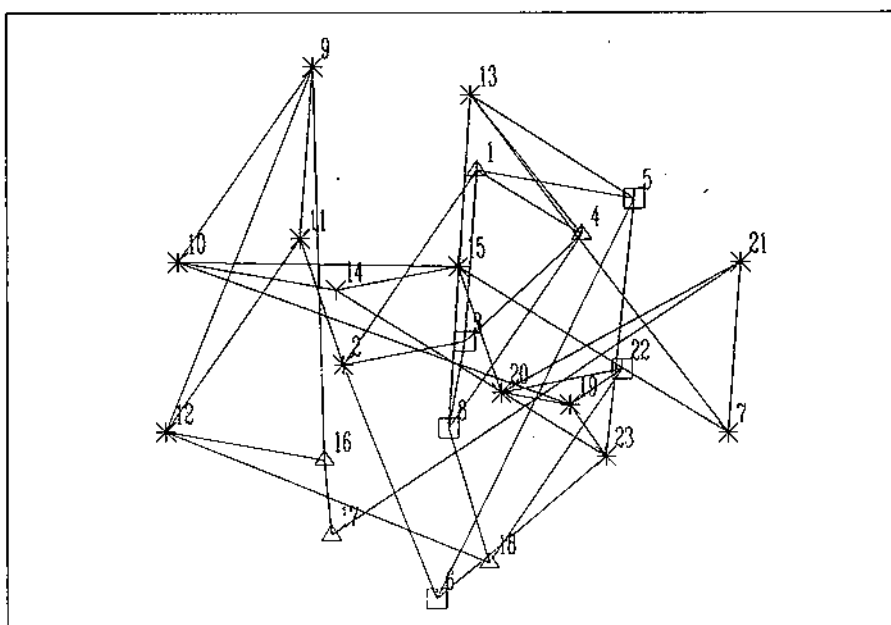
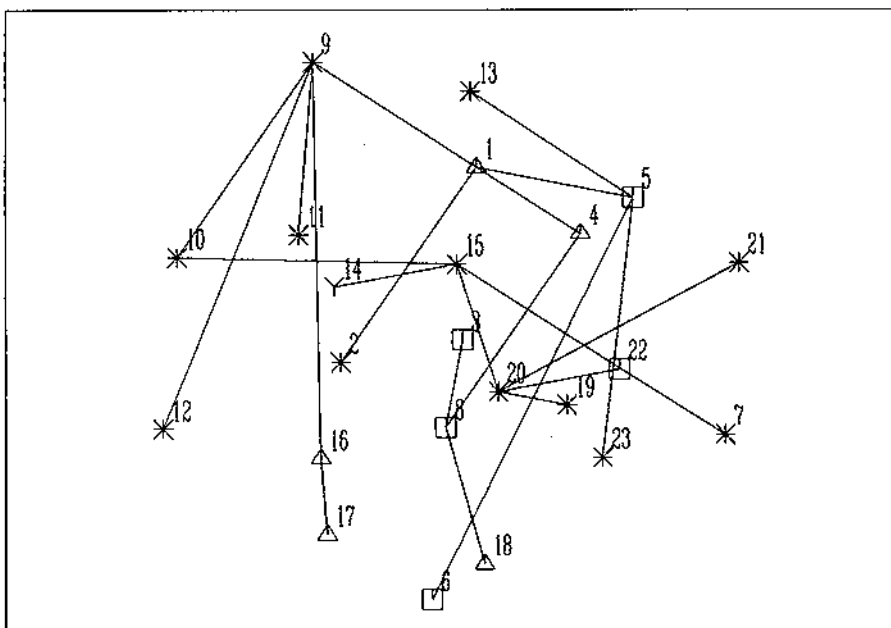


Figura 7. a) Arbre de Kruskal. b) Graf de veïns més propers. Els símbols són l'activitat en front de les cèl·lules P388/W. S'han utilitzat distàncies euclídiades.

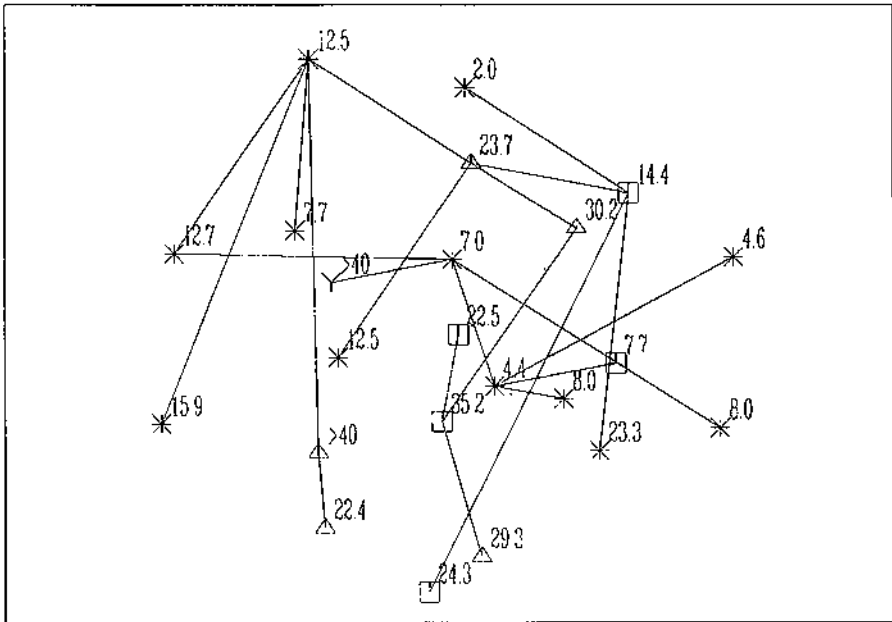
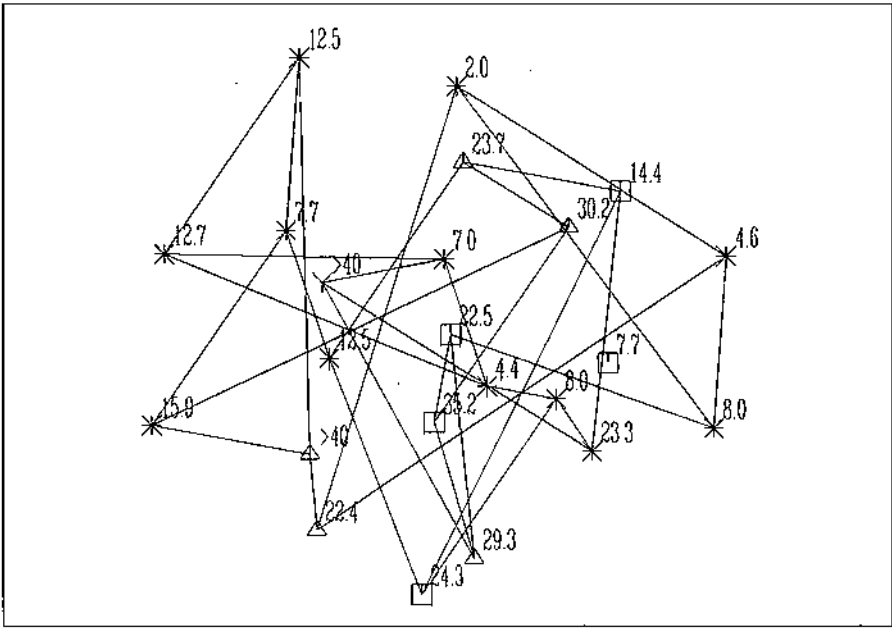


Figura 8. a) Graf mínim de veïns més propers. b) Arbre de Kruskal. Els símbols són l'activitat amb les cèl·lules P388/W i les etiquetes amb les P388/A.

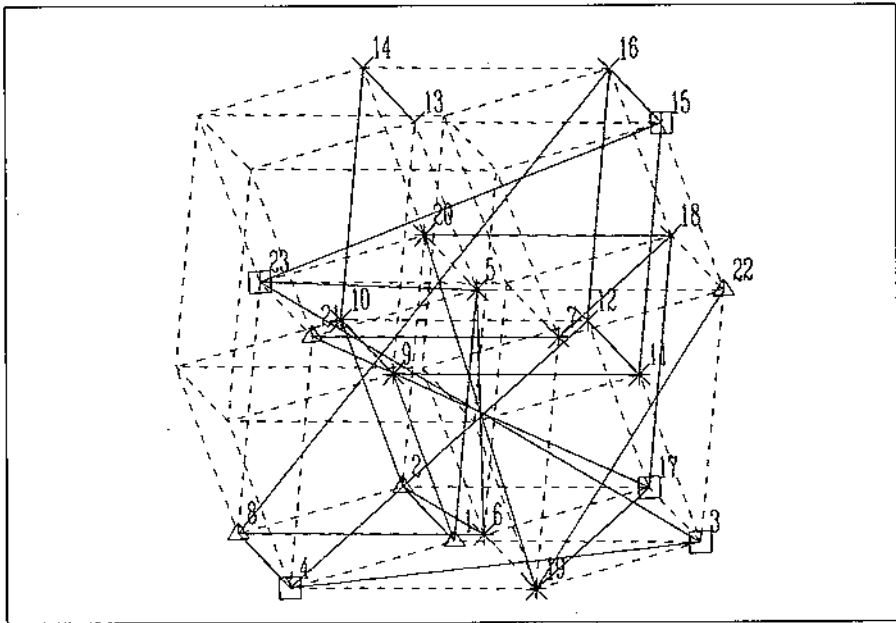
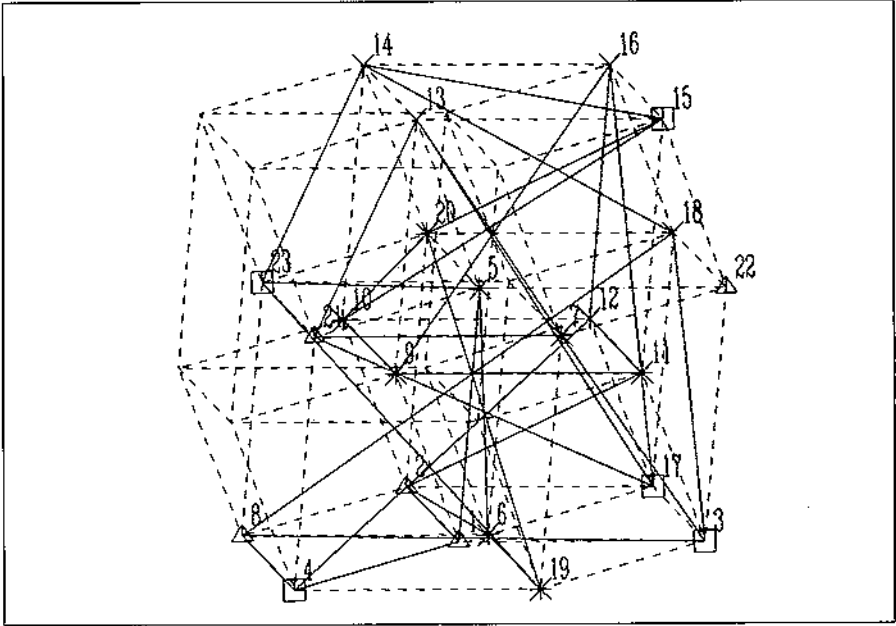


Figura 9. Gracs mínims de veïns més propers i l'hipercub on són inscrits.
a) Sense restriccions. *b)* Els lligams han de ser arestes de l'hipercub.
 Els símbols són el punt de fusió.

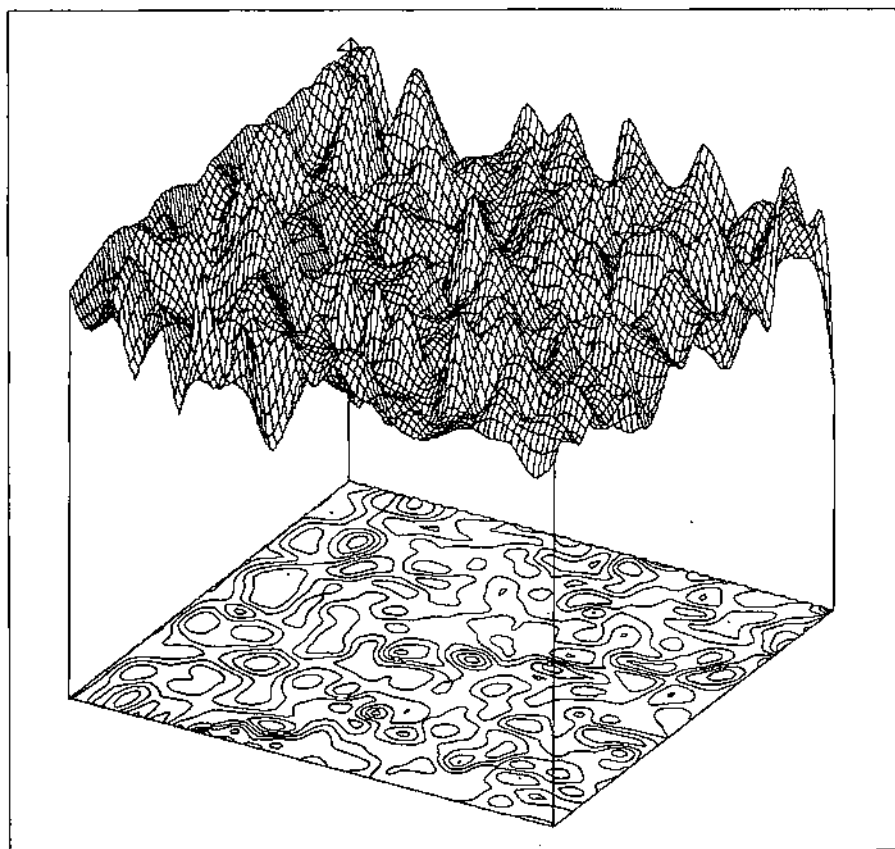


Figura 10. Superfície de les components principals.

La figura 7 compara *a*) un arbre de Kruskal i *b*) un graf de veïns més propers de dos connexions com a mínim i quatre com a màxim, on els signes representen l'activitat enfront de les cèl·lules P388/W i les etiquetes, el número de la molècula. Com que les activitats es donen en forma de concentració efectiva, cal tenir en compte que activitats altes corresponen a concentracions petites. La distància utilitzada ha estat l'euclídia en tots dos casos.

La figura 8 compara *a*) un graf mínim de veïns més propers de 3 connexions com a màxim i una com a mínim amb *b*) un arbre de Kruskal. Els símbols representen l'activitat enfront de les cèl·lules P388/W i les etiquetes l'activitat enfront de les P388/A.

La figura 9 representa dos grafs de veïns més propers mínims i l'hipercub on estan inscrits; *b*) té la restricció que les unions entre molècules sols poden estar situades a l'hipercub, la qual cosa no compleix el primer *a*). En tots dos casos la propietat representada pels símbols és el punt de fusió mitjà, i per les etiquetes, el número de la molècula.

Superfícies tridimensionals

Les components principals de la matriu Z es poden representar gràficament agafant els coeficients com si fossin els valors d'una funció de dues variables. D'aquesta manera s'aconsegueixen gràfics que permeten veure d'un cop d'ull les estructures dels vectors propis procedents de càlculs quàntics de semblança molecular.

La figura 10 és la superfície de les components principals de la matriu Z procedents d'un càlcul de semblança tipus recobriment.

CONCLUSIONS

La informació infinit-dimensional com la donada per funcions densitat quàntiques en molècules es pot transformar en un subespai de dimensió finita mitjançant càlculs de semblança quàntica. La manipulació geomètrica del núvol de punts resultant permet, mitjançant rotacions i projeccions òptimes i conservant el màxim possible la informació, d'obtenir uns gràfics on poden analitzar-se les relacions simultànies entre les molècules.

Agraïments

Aquest treball ha tingut l'ajut número FAR 88-0617 del Ministerio de Industria y Energía sota el Programa Nacional de Investigación y Desarrollo Farmacéuticos.

Bibliografia

1. DENNY, W.A.; REWCASTLE, G.W. & BAGULEY, B. C. (1990). *J. Med. Chem.*, 33, 814.
2. CARBÓ, R. & DOMINGO, LI. (1987). *Int. J. Quantum Chem.*, 23, 517.
3. CARBÓ, R.; LEYDA, L. & ARNAU, M. (1980). *Int. J. Quantum Chem.*, 17, 1185.
- 4a. POPLE, J.A. (1965). *J. Chem. Phys.*, 43, 5129.
- 4b. POPLE, J.A. & BEVERIDGE, D.L. (1970). *Approximate Molecular orbital Theory*, McGraw-Hill, New York.
5. CARBÓ, R. & ARNAU, C. (1981). «Medicinal Chemistry Advances» F. G. de las Heras i S. Vega (eds.), Pergamon Press, Oxford.
6. MULLIKEN, R. S. (1945). *J. Chem. Phys.*, 23, 1833, 1841, 2338, 2343.
- 7a. CARBÓ, R. & CALABUIG, B. (1989). *Computer Physics Communications*, 55, 117.
- 7b. MOLSIMIL-88 es pot demanar a: Ref: ABJG. CPC Program Library. Queen's University of Belfast. Belfast BT71NN. Northern Ireland.
8. FU, K.S. (1970). «Digital Pattern Recognition», *Communications and Cybernetics*, Springer Verlag, Berlin, Vol. 10.
9. CARBÓ, R. & CALABUIG, B. (1990). «Concepts and Applications of Molecular Similarity», Mark A. Johnson i Gerald M. Maggiora (eds.), John Wiley & Sons Inc., New York, pàg. 147.
10. YOUNG, T.L. & WOERT, M. L. Van (1988). «PLOT88 Software Library Reference Manual», versió Microsoft Fortran V4.1, Plotworks, Ramona (California).
11. ALCHEMY II (1988). Tripos Associates Inc., St. Louis, Missouri.