SCIENTIA gerundensis, 21: 127-143 (1995)

PRACTICAL APPLICATIONS OF QUANTUM MOLECULAR SIMILARITY MEASURES (QMSM): PROGRAMS AND EXAMPLES

Lluís Amat, Emili Besalú, Ramon Carbó & Xavier Fradera

Institut de Química Computacional
Universitat de Girona
17071 Girona.

RESUM

Es descriu l'aproximació de Capes Atòmiques dins de la teoria de la Semblança Molecular Quàntica. Partint només de dades teòriques, s'ha trobat una relació entre estructura molecular i activitat biològica per a diversos conjunts de molècules. Es descriuen els aspectes teòrics de la Semblança Molecular Quàntica i alguns exemples d'aplicació.

RESUMEN

Se describe la aproximación de Capas Atómicas dentro de la teoría de la Semejanza Molecular Cuántica. A partir sólo de datos teóricos, se ha encontrado una relación entre la estructura molecular y la actividad biológica para varios conjuntos de moléculas. Se describen tanto los aspectos teóricos de la Semejanza Molecular Cuántica como algunos ejemplos de aplicación.

ABSTRACT

The Atomic Shell (AS) approach of QMSM theory is described and employed over some molecular sets. Molecular structure-biological activity relationships are found using only theoretical data. Both aspects of QMSM framework are described: the theoretical one and some application examples.

Keywords: Molecular Quantum Similarity, Quantum Similarity, Atomic Shell Approximation, Quantum Similarity Indices, Structure-Activity Relationships

INTRODUCTION

QMSM definition and interpretation can be found in several previous papers [1]. As a consequence of the Quantum Mechanical Postulates, all the information which can be extracted from a molecular system is contained, to some degree, in the molecular density function. Thus, if two *n*th order density functions for two systems *A* and *B* are known, $\Gamma_A^{(n)}(r_1)$ and $\Gamma_B^{(n)}(r_2)$, using a positive definite operator, a QMSM between the quantum systems *A* and *B* can be defined as the integral:

$$Z_{AB}^{(n)}=\iint \Gamma_A^{(n)}(r_1)\ \Omega(r_1,r_2)\Gamma_B^{(n)}(r_2)\ dr_1\ dr_2 \qquad (1)$$

And these integrals can be used to extract information from the set of studied molecules.

ATOMIC SHELL (AS) APPROACH

Most molecules with biological activity have a large number of atoms. Moreover, the studied molecular sets can have also a large cardinal. This makes impossible for the moment to make systematic *ab-initio* Similarity Calculations, due to the high computational cost. In the present paper QMSM have been obtained under an Atomic Shell approach [1d,e,f]. The first-order density function, under this kind of AS approach, transforms into the expression:

$$\Gamma^{(1)}(r) \approx \sum_{I} R_I(r) \tag{2}$$

where the following definition holds:

$$R_I(r) = Q_I \mid \phi_I(r) \mid^2 \tag{3}$$

and the sum in the equation (2) is performed over all the atomic shells of the molecule. Within the simplest choice only one shell is assumed per atom: Q_I is the gross atomic population of the Ith atom and $\delta_I(r)$ is a *ns*-STO function centered at the Ith atom. Applying this approximation into equation (1) one can obtain the following first-order QMSM:

$$Z_{AB} \approx \sum_{I \in A} \sum_{J \in B} \iint R_I(r_1)\, \Omega(r_1, r_2)\, R_J(r_2)\, dr_1 dr_2 \tag{4}$$

Four different operators are normally used, giving four kinds of QMSM. These are:

a) *Overlap-like QMSM*. Using a Dirac Delta function $\delta(r_1 - r_2)$ instead of the operator Ω in equation (4), the following expression is reached:

$$\begin{aligned} Z_{AB} &\approx \sum_{I \in A} \sum_{J \in B} \int R_I(r)\, R_J(r)\, dr \\ &\approx \sum_{I \in A} \sum_{J \in B} Q_I Q_J\, (II : JJ), \end{aligned} \tag{5}$$

where

$$(II : JJ) = \int \mid nS^I(r) \mid^2 \mid n'S^J(r) \mid^2 dr \tag{6}$$

b) *Cioslowski-McWeeny-like QMSM*. Applying the AS approach to the QMSM defined by Cioslowski [2] and McWeeny [3], the following expression for the non diagonal term of the density matrix is obtained:

$$\Gamma^{(1)}(r_1, r_2) \approx \sum_{I} R_I(r_1, r_2), \tag{7}$$

where

$$R_I(r_1, r_2) = Q_I \phi_I(r_1)\, \phi_I(r_2) \tag{8}$$

one can define the following approximate Similarity Measure:

$$\begin{aligned} Z_{AB} &= \int\int \Gamma_I^{(1)}(r_1,r_2)\, \Gamma_J^{(1)}(r_1,r_2)\, dr_1 dr_2 \\ &= \sum_{I\in A}\sum_{J\in B} \int\int R_I(r_1,r_2)\, R_J(r_1,r_2)\, dr_1 dr_2 \\ &= \sum_{I\in A}\sum_{J\in B} Q_I Q_J S_{IJ}^2, \end{aligned} \tag{9}$$

where S_{IJ} is the overlap matrix element related to the functions I and J of the molecules A and B, respectively:

$$S_{IJ} = \int [nS^I(r)]\, [n'S^J(r)]\, dr \tag{10}$$

c) *Coulomb-like QMSM*. Using equation (4) with the Coulomb operator, then:

$$\begin{aligned} Z_{AB} &= \sum_{I\in A}\sum_{J\in B} \int\int R_I(r_1)\, |r_1 - r_2|^{-1} R_J(r_2)\, dr_1 dr_2 \\ &= \sum_{I\in A}\sum_{J\in B} Q_I Q_J (II \mid JJ), \end{aligned} \tag{11}$$

where

$$(II \mid JJ) = \int\int |nS^I(r_1)|^2\, |r_1 - r_2|^{-1} |n'S^J(r_2)|^2 dr_1 dr_2 \tag{12}$$

d) *Triple Density QMSM*. In equation (4) the positive definite operator Ω is substituted by another first order density function $\Gamma_C^{(1)}(r)$, giving:

$$\begin{aligned} Z_{AB,C} &= \sum_{I\in A}\sum_{J\in B} \int R_I(r)\, \Gamma_C^{(1)}(r)\, R_J(r)\, dr \\ &= \sum\sum\sum Q_I Q_K Q_J (I \mid K \mid J), \end{aligned} \tag{13}$$

where

$$(I \mid K \mid J) = \int |nS^I(r)|^2 |n'S^K(r)|^2 |n''S^J(r)|^2 dr \tag{14}$$

A *Cioslowski-McWeeny-like* Similarity Measure can be also defined into the Triple Density framework in this way:

$$\begin{aligned} Z_{AB,C} &= \sum_{I\in A}\sum_{K\in L}\sum_{J\in B} \int\int\int R_I(r_1,r_2)\, R_K(r_2,r_3)\, R_J(r_3,r_1)\, dr_1 dr_2 dr_3 \\ &= \sum\sum\sum Q_I Q_K Q_J S_{IK}^2 S_{KJ}^2 S_{JI}^2, \end{aligned} \tag{15}$$

SIMILARITY MEASURES OPTIMIZATION

When a QMSM between two or three molecules has to be computed, it is not sufficient a single calculation: a search of the relative position of the molecules with maximal QMSM is needed. The strategy used to get the optimal similarity value is to have a molecule fixed and rotate and translate the other(s) until a maxim QMSM

value is obtained. When the measure is sought between two molecules, there are six variables to search: three rotations and three translations, but when Triple Density Measures are computed there are twelve variables. Due to the high number of maximums appearing in the QMSM surface, optimization is the costly part of the computation.

The best optimization strategy has been found to be made using two search techniques at the same time. All the computations in the present paper have been performed within a united Monte Carlo - Simplex methodology [4]. The Monte Carlo method explores the Similarity surface and every promising zone on this surface can then be scanned via the Simplex method.

SIMILARITY INDICES

There are various possible manipulations of QMSM. Among all these possibilities [1f,1g], one can choose two classical ones. From the matrix elements of the optimal similarity Z_{AB}, the Carbó index R_{AB} [1a] is given by:

$$R_{AB} = \frac{Z_{AB}}{\sqrt{Z_{AA} Z_{BB}}} \tag{16}$$

and the euclidean distance D_{AB} can be easily computed as:

$$D_{AB} = \sqrt{Z_{AA} + Z_{BB} - 2Z_{AB}} \tag{17}$$

VISUALIZATION AND INTERPRETATION OF SIMILARITY DATA

From QMSM analysis, some trends can be put into evidence in order to classify the molecular set. All the numerical manipulations within this step are totally independent on the way the QMSM has been computed.

The method used [1] is based in the consideration that, for any set of N molecules, every one of them can be represented by a N-dimensional vector: a column vector of the QMSM matrix. The vector components are the similarities of a given molecule with the N molecules in the set. This means that every molecule can be represented as a point in the N-dimensional space, a *Point-Molecule*. The closer two Point-Molecules are, more similar will be the molecules they represent. A set of Point-Molecules is a *Molecular Cloud*.

There are many ways to extract information from the QMSM. One of the possibilities is the utilization of graph-algorithms, which use the distances between the molecules to make connections between them. The distances can be those of the Distance Index Matrix or the distances computed between the Point-Molecules. The distance normally used is an euclidian one, but any kind of distance could be used. Two kinds of graphs are currently considered [5]:

- *Kruskal Tree:* A minimal set of connections is made between the Point-Molecules, with the only restriction that all the points must be connected through the same

graph without cycles, trying to minimize the total length of the tree formed. This is also known as a Minimal Spanning Tree [1c].

- *Nearest Neighbour Graph:* Every point is connected with those nearest to it, keeping a minimum and maximum of connections for each point. This method allows the formations of cycles, and even clusters, that is, graphs not connected between them. The algorithm has been designed by us.

The next possibility is to make a projection of the Molecular Cloud from the original N-dimensional space to any subspace having dimension $M < N$. One can apply the *Mendeleev Principle* [6], that is, performing different projections, an ordering of the molecules can be achieved for any property. Then, if a projection is found such that it can classify a set of molecules with known property values into groups, one can expect that molecules with unknown property values will be spatially located in such a way that they will be associated to any of the groups of known molecules. In that way, a prediction of the property for these unknown molecules could be done, if not quantitatively, at least qualitatively.

MOLSIMIL AND ND-CLOUD PROGRAMS

A set of programs was developed in order to test the previous concepts. Both of them are coded in Fortran and can be used in an interactive way through a pull-down menu system. Both programs were compiled with NDP-486 Fortran Compiler [7] and the S-GKS graphical Library [8], and run in a PC environment, although they are being adapted to run on IBM/6000 RISC machines and Silicon Graphics Systems.

Previous versions of MOLSIMIL [9] and TRIDENT [10] programs were updated and merged into a single code: MOLSIMIL '95. This program is used to compute QMSM for any set of molecules. In order to perform fast QMSM calculations, specially in the optimization steps, *overlap-like* measures can be computed using a *1s*-GTO functions centered in each atom. *ns*-STO functions can be used in order to get more accurate results for any of the operators previously described (*overlap*, *Cioslowski-McWeeny* or *Coulomb*). For the Coulomb operator, the Ohno [12] or Mataga [11] approximations can be used to perform fast optimizations, and then exact Coulomb integrals employed.

The program ND-CLOUD can read QMSM obtained by means of MOLSIMIL '95 code. Then, the Point-Molecules are defined and the Molecular Cloud can be projected in any plane and rotated through any axis to find the best ordering for the studied set of molecules and properties. This process has been made user transparent through a subroutine that tries to group the molecules with similar property values [13]. Kruskal Tree and various kinds of Nearest Neighbour Graph can be generated and graphically visualized.

Some examples of successful classifications achieved using ND-CLOUD code are described below.

APPLICATION EXAMPLES

a. Similarity between the atoms of a molecule

In this example, QMSM were computed not for sets of molecules but for atoms

belonging to a molecule. The geometry of all the molecules was fully optimized with AMPAC [14] program, using AM1 methodology [15]. STO *overlap-like* QMSM were computed between pairs of atoms, keeping them fixed in their positions according to the AM1 geometry. Thus, the QMSM will depend not only on the kind of atoms involved, but also on the distance between them. In this way, the proximity between Point-Molecules will be related to the interatomic distances, which is, in fact, the property studied in this case. QMSM were computed in this way for all the isomers of heptane and for a set of benzene derivatives. We tried, using the graphs described before, to recover information about the atomic connectivity. In particular, a Kruskal Tree has reproduced all the bonds of heptane, and Nearest Neighbour algorithms with 1 to 3 links were able to reproduce nearly all the atomic bonds for the benzene derivatives. Figures 1, 2 and 3 show the obtained links.

b. Benzodiazepines

This example was made of a set of 35 imidazo[1,4]benzodiazepine (I), imidazo[1,5-a]pyrrolo[2,1-c][1,4]benzodiazepinone (II) and imidazo[1,4]thienodiezopinone (III) esters [16]. The geometry of all these molecules was fully optimized with the AMPAC program, using AM1. From AM1 geometries, a first *overlap-like* optimized QMSM was computed using *1s*-GTO functions. The optimization was then completed using *ns*-STO, and *overlap*, *Ciosłowski-McWeeny* and *Coulomb* QMSM were finally obtained and used as ND-CLOUD input. The results presented here correspond to the *overlap* QMSM.

The first proposed goal was the ordering of the molecules in a structure-dependent way. Structures and activities for all the molecules are given in tables 1 to 3 and figures 4 to 6. Figures 7, 8 and 9 show the graphs obtained with ND-CLOUD, which presents an ordering of the molecules based mainly on their R_8 substituent. Most of the molecules with Cl as R_8 substituent are connected and very close in the obtained graphics. Moreover, they appear near to the Br, I, and in some cases F, R_8 substituted molecules. The H R_8 substituted molecules also form a clear group. The same happens for the molecules with N_3 and NO_2 R_8 groups. A grouping following the base structure does also appear, being the molecules 12 and 13 (structure III) clearly isolated from the rest, in some cases connected only between themselves. Molecules 17, 19 and 25 (structure II) appear also very close and interconnected. On the other hand, the R group, being bigger, in general, than R_8, seems to have much less influence on the ordering of the Molecular Cloud. This feature can be explained in terms of electronic density: differences between R_8 substituents, which range from H to I, are greater than those between R substituents

A biological property has been also studied for the same set of molecules: the selectivity of this compounds in front of two subtypes of the benzodiapine receptor: the *diazepam insensitive* (DI) and the *diazepam sensitive* (DS) ones, being this selectivity expressed as the quotient between the affinities to DI and DS. The logarithm of this selectivity was used to group the molecules, and a projection was sought in order to reflect this appropriate grouping. This is shown in figure 10.

c. Triazines

A set of seven Baker triazines with Dihydrofolate Reductase Inhibition (DHFR) Activity [17] was studied. Structures and activities for these compounds are given in figure 11 and in table 4. The molecules selected have a relatively small number of conformational degrees of freedom, being the most important feature, in conformational terms, the dihedral angle Θas defined in figure 11. According to Hopfinger [17], active conformers are those having $\Theta = 270^{\circ}$. Thus, the geometry of all the molecules was optimized with AMPAC program system using AM1 methodology, freezing Θ at 270º and letting a full optimization for the rest of coordinates. As in the preceding example, after a preliminary optimization with *1s*-GTO functions, an *Overlap ns*-STO QMSM was computed with molecular pairs in optimal superposition.

Using this QMSM as ND-CLOUD program input, a projection is readily achieved where the triazines are clearly ordered in the plane according to their DHFR activity. Figure 12 shows the triazines connected by a Kruskal Tree, orders them from lowest to highest activity. In figure 13, there are two different Nearest Neighbour Graphs. Both of them show high and low activity clusters. Compounds 3 and 4, having intermediate activity values (6.92 and 6.52), are not clearly assigned to the same cluster in both graphs.

CONCLUSIONS

The concept of Quantum Similarity, applied to any molecular set, is a valuable tool to get information about the relationships between the members of the set, using only quantum mechanical principles and geometric reasoning.

An application is readily found in the field of structure-activity relationships and activity predictions. An Atomic Shell approach has been shown to be fast and reliable in order to work with relatively big compounds with attached pharmacological activity. Within this framework, good classifications have been found for various sets of atoms and molecules.

ACKNOWLEDGEMENTS

One of us (Ll. A.) benefits from a grant from the "Ministerio de Educación y Ciencia".

Bibliografia

1. a) CARBÓ, R., ARNAU, M., and LEYDA, L., Int.J.Quantum Chem., *1980*, 17, 1185.
b) CARBÓ, R., and ARNAU, C., *Molecular engineering: a general approach to QSAR*, in de las Heras, F.G., and Vega, S., (eds.), *Medicinal chemistry advances*, Pergamon Press, Oxford, Vol. 85, 1981.
c) CARBÓ, R., and DOMINGO, L., Int.J.Quantum Chem., *1987*, 23, 517.
d) CARBÓ, R., and CALABUIG, B., *Molecular Similarity and Quantum Chemistry*, chapter 6, p. 147, in Johnson, M. A., and Maggiora, G. M., Eds., "Concepts and applications of Molecular Similarity", Wiley, New York, 1990.

e) CARBÓ, R., and CALABUIG, B., Int.J.Quantum Chem., *1992*, 42, 1681, 1695.
f) CARBÓ, R., BESALÚ, E., CALABUIG, B., and VERA, L., Adv. Quantum Chem., *1994*, 25, 253-313.
g) BESALÚ, E., CARBÓ, R., MESTRES, J., and SOLÀ, M., in: Sen, K., Ed., "Topics in Current Chemistry", Springer-Verlag, Berlin, 1995

2. CIOSLOWSKI, J., and FLEISCHMANN, E.D., J.Am.Chem.Soc., *1991*, 113, 64.
3. MCWEENY, R., and AMOVILLI, C., J.Mol.Struct.(Theochem), *1991*, 227, 1.
4. a) PIERRE, D.A., "Optimization Theory with Applications", John Wiley & Sons, Inc., 1969.
   b) PRESS, W.H., STEUKOLSKY, S.A., VETTERLING, W.T., and FIANNERY, B.P., "Numerical Recipes", Cambridge University Press, 1986.
5. a) TOU, J.T., and GONZÁLEZ, R.C., "Pattern Recognition Principles", Addison-Wesley Reading, MA, 1974.
   b) HARDIGAN, J.A., "Clustering Algorithms", Wiley, New York, 1975.
   c) SNEATH, P.H.A., and SOKAL, R.R., "Numerical Taxonomy", W.H.Freeman, San Francisco, 1973.
   d) JARDINE, N., and SIBSON, R., "Mathematical Taxonomy", Wiley, New York, 1977.
6. CARBÓ, R., and BESALÚ, E., "Theoretical Foundation of Quantum Molecular Similarity", Proceedings of 1st Girona Seminar on Molecular Similarity, Kluwer Amsterdam, in press.
7. NDP Fortran 486 v4.0.2 "User's manual and Library". Microway Inc., Kingston, 1992.
8. S-GKS 3.0. "System & User's Guide", Scientific Software Limited, 1993.
9. CARBÓ, R., and CALABUIG, B., Comp.Phys.Commun., *1989*, 55, 117.
10. CARBÓ, R., CALABUIG, B., BESALÚ, E., and MARTÍNEZ, A., Molecular Engineering, *1992*, 2, 43.
11. OHNO, K., Theoret.Chim.Acta, *1964*, 2, 219.
12. a) MATAGA, N., and NISHIMOTO, K., J.Phys.Chem., *1954*, 13, 140.
    b) Mataga, N., Bull. Chem. Soc. Japan, *1958*, 31, 453.
13. BESALÚ, E., Scientia Gerundensis, *1994*, 20, 87-93.
14. LIOTARD, D.A., HEALY, E.F., RUIZ, J.M., and DEWAR, M.S.J., AMPAC-version 2.1. Quantum Chemistry Program Exchange, program 506, QCPE Bull., 9, 1989.
15. a) DEWAR, M.J.S., ZOEBISCH, E.G., HEALY, E.F., and STEWART, J.J.P., J. Am. Chem. Soc., *1985*, 107, 3902.
    b) Hehre, W.J., Stewart, R.F., and Pople, J.A., J. Chem. Phys., *1969*, 51, 2657.
16. ZI-QIANG GU, WONG, G., DOMÍNGUEZ, C., DE COSTA, B. R., RICE, K. C. and SKOLNICK, P., J. Med. Chem., *1993*, 36, 1001.
17. HOPFINGER, A. J., J. Am. Chem. Soc., *1980*, 102, 7196.

compound	R	R_8	DI/DS ratio
1	$C(CH_3)_3$	$N_3(CH_3)_2$	0.6
2	$CH_2CH_2C(CH_3)_3$	Cl	>1000
3	$CH_2C(CH_3)_3$	Cl	0.67
4	$CH(CH_2CH_3)_2$	Cl	4.6
5	$C(CH_3)_3$	N_3	0.20
6	$C(CH_3)_3$	NCS	0.73
7	$C(CH_3)_3$	NO_2	0.77
8	$C(CH_3)_3$	I	0.21
9	$C(CH_3)_3$	Br	0.17
10	$C(CH_3)_3$	Cl	0.42
11	$C(CH_3)_3$	H	19.2
14	CPM	Cl	4.1
15	$CH(CH_3)_2$	Cl	0.8
16	$CH_2CH_2CH_3$	Cl	1.3
18	CH_2CH_3	N_3	0.6
20	CH_2CH_3	F	73
21	CH_2CH_3	Cl	3.1
22	CH_2CH_3	H	164.6
23	CH_3	Cl	4.1
24	CH_3	H	>1000

Table 1. Molecules within structure (I).

compound	R	DI/DS ratio
12	Cl	0.73
13	H	50

Table 2. Molecules within structure (II).

compound	R	DI/DS ratio
12	Cl	0.73
17	$C(CH_3)_3$	13
19	$CH_2CH_2CH_3$	27
25	CH_3	57

Table 3. Molecules within structure (III).

Compound	R_1	R_2	R_3	R_4	DHFR Activity
1	H	Cl	Cl	H	8.54
2	H	Cl	H	H	7.76
3	H	H	H	H	6.92
4	Cl	Cl	H	H	6.52
5	F	H	H	H	4.74
6	Cl	H	H	H	4.15
7	Cl	H	H	Cl	3.43

Table 4. Compounds within structure (IV).

	1	2	3	4	5	6	7
1	45.7624						
2	37.7882	37.4417					
3	29.8627	29.4814	29.0934				
4	37.0783	36.8687	29.3409	45.7528			
5	30.2525	29.8977	29.5324	34.9841	37.8753		
6	29.7611	29.2751	28.9632	37.7895	34.5874	37.4339	
7	33.5872	29.5828	29.1608	38.1958	34.7415	37.7430	45.7717

Table 5. QMSM matrix for the Baker triazines set.

	1	2	3	4	5	6	7
1	1.0000						
2	0.9129	1.0000					
3	0.8184	0.8933	1.0000				
4	0.8103	0.8908	0.8042	1.0000			
5	0.7267	0.7939	0.8897	0.8404	1.0000		
6	0.7191	0.7820	0.8776	0.9131	0.9186	1.0000	
7	0.7339	0.7146	0.7991	0.8347	0.8344	0.9118	1.0000

Table 6. Carbó Index matrix for the Baker triazines set

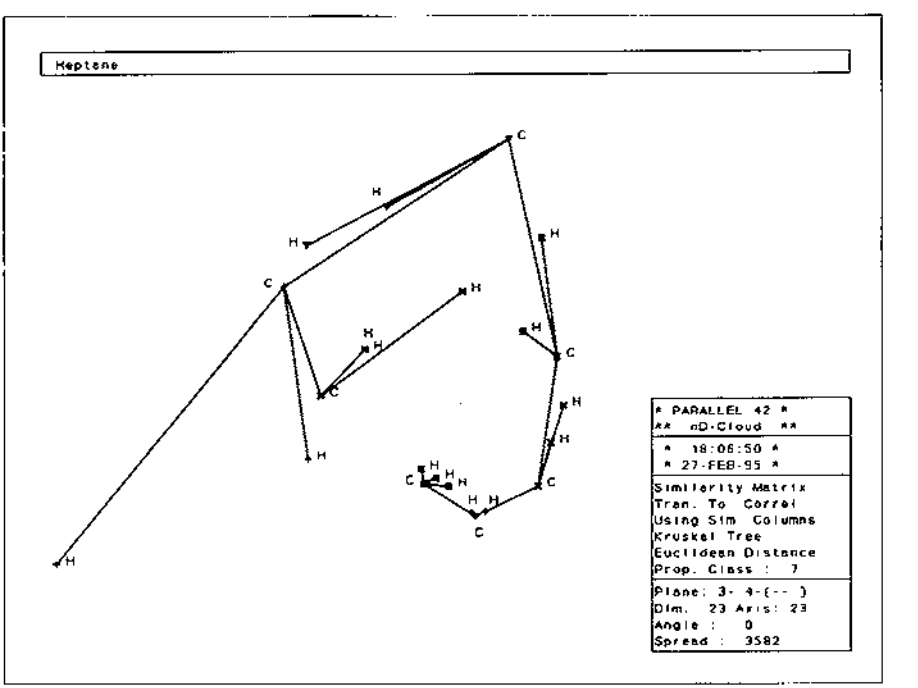

Figure 1 : Heptane atoms represented into a Kruskal Tree, computed over the Similarity Distance Matrix.

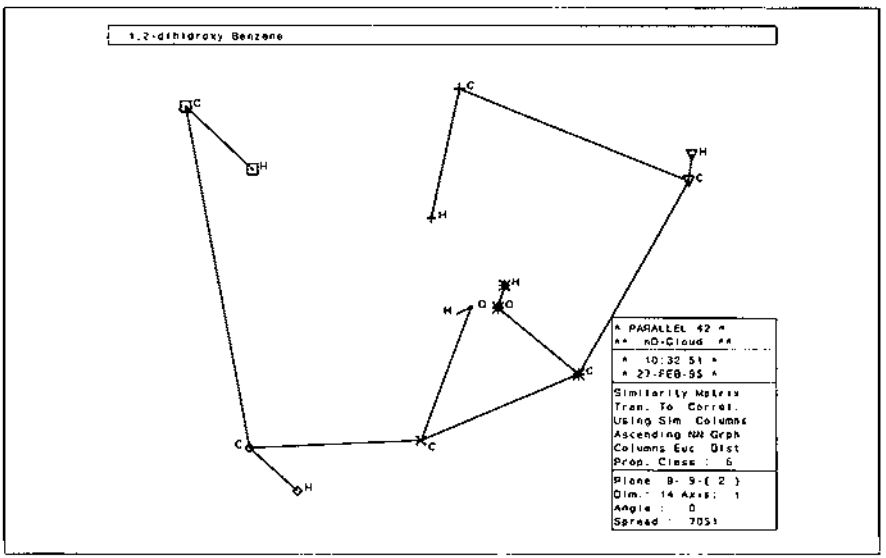

Figure 2 : 1,2-dihidroxibenzene is represented here. A Nearest Neighbour Graph, with 1 to 3 links, reproduces the atomic connectivity but for a C-C bond of the benzene cycle open. Each H atom has been grouped into its corresponding C, being this groupping reflected in the projection. The size of the vertices represents here a third dimension.

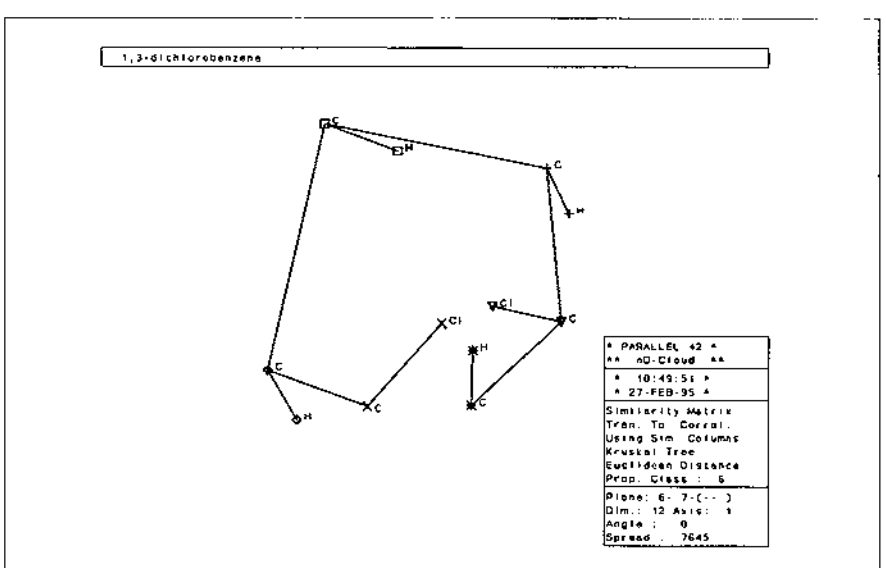

Figure 3 : 1,3-diclorbenzene, with its atoms connected by a Kruskal Tree, computed over the Similarity Distance Matrix. The tree coincides exactly with the connectivity, but the benzene cycle is not closed. The result is the best possible within a Kruskal tree, which cannot form cycles.

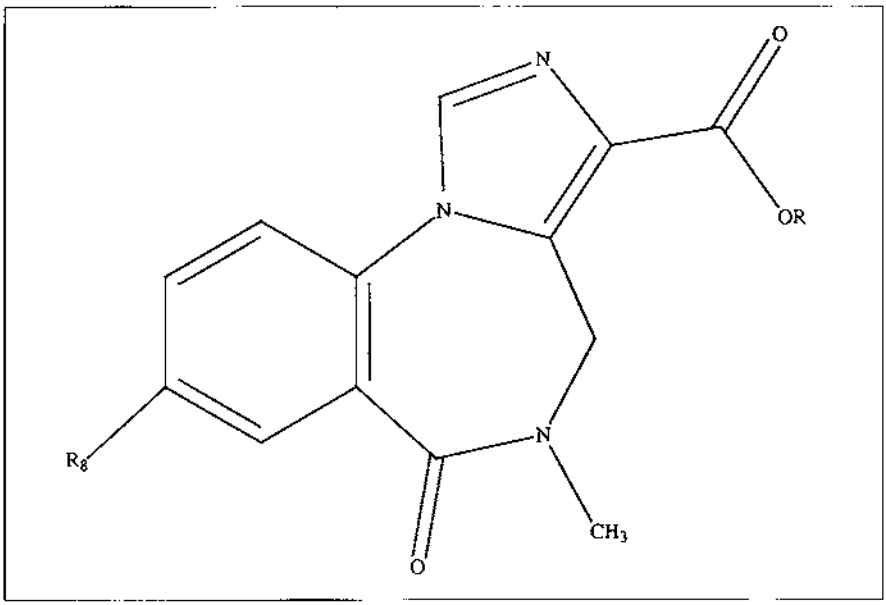

Figure 4. Structure (I) for the imidazo[1,4]benzodiazepine esters.

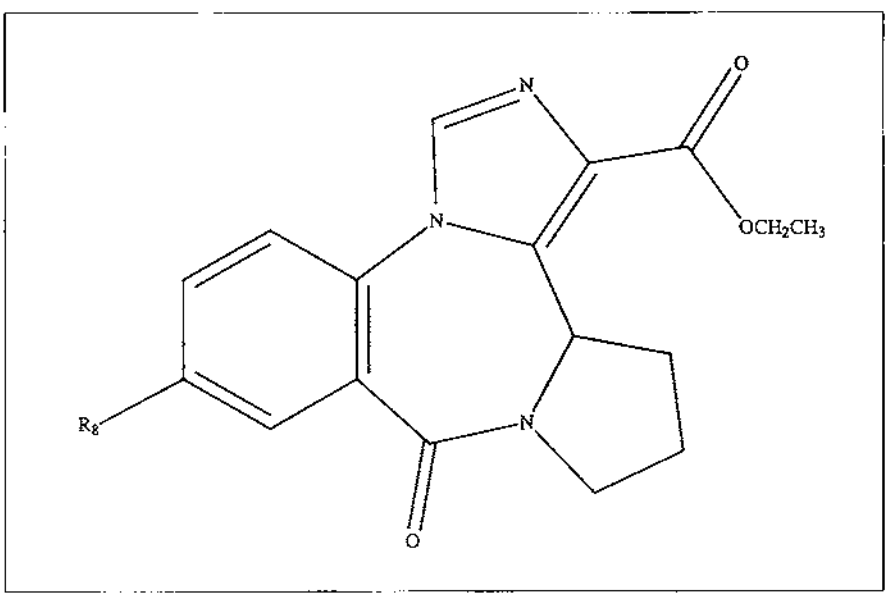

Figure 5. Structure (II) for the imidazo[1,5-a]pyrrolo[2,1-c][1,4]benzodiazepinone esters.

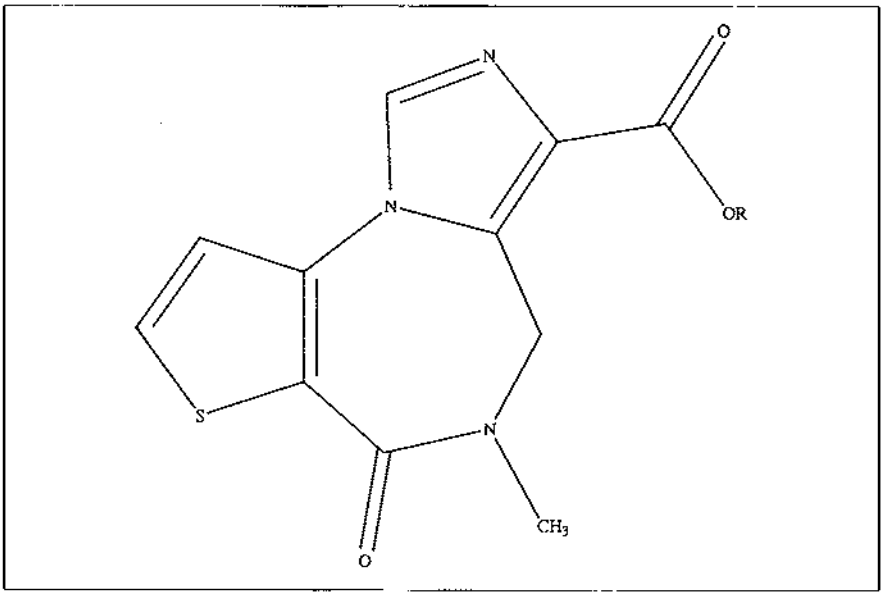

Figure 6. Structure (III) for the imidazo[1,4]thienodiezopinone esters.

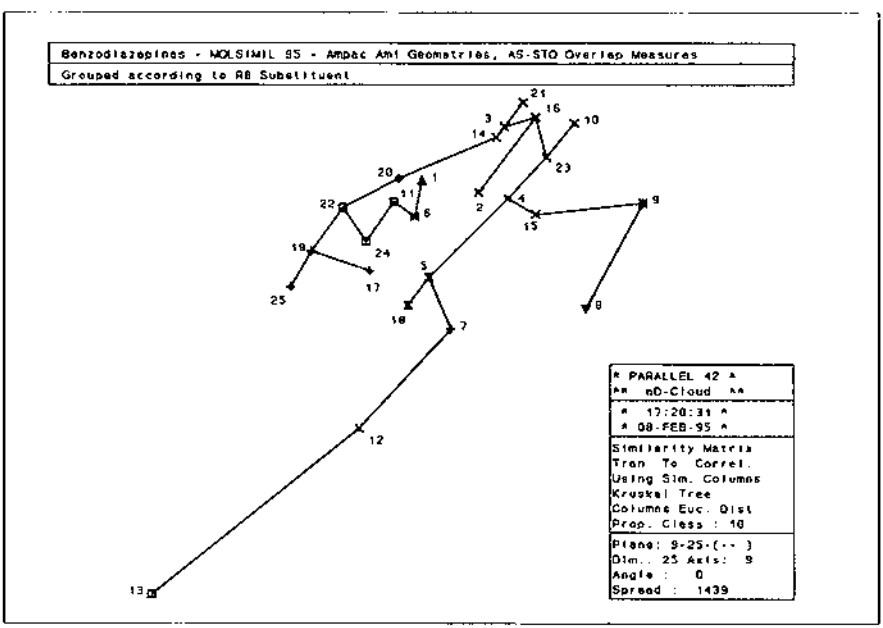

Figure 7 : Kruskal Tree of benzodiazepines, using the Point-Molecule distances.

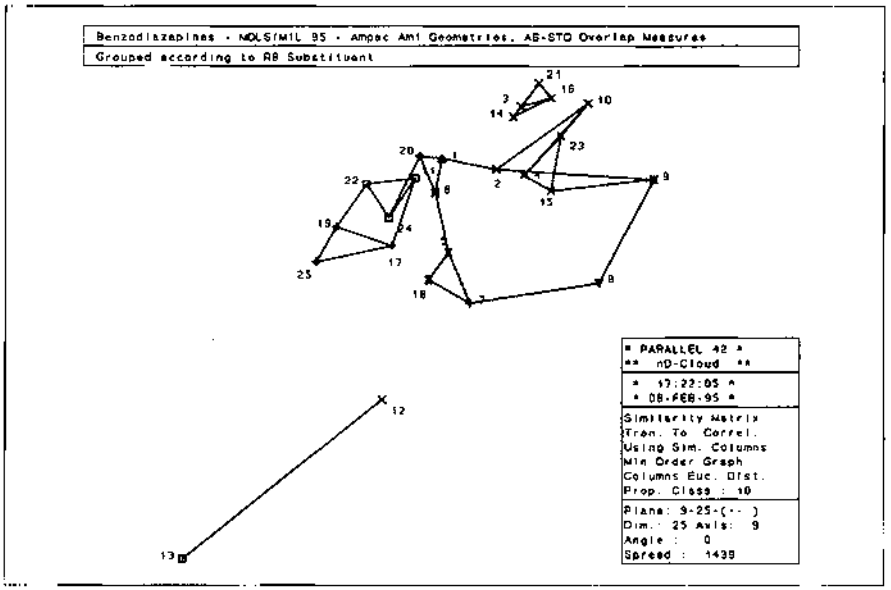

Figure 8 : Nearest Neighbour Graph of benzodiazepines, using the Point-Molecule distances.

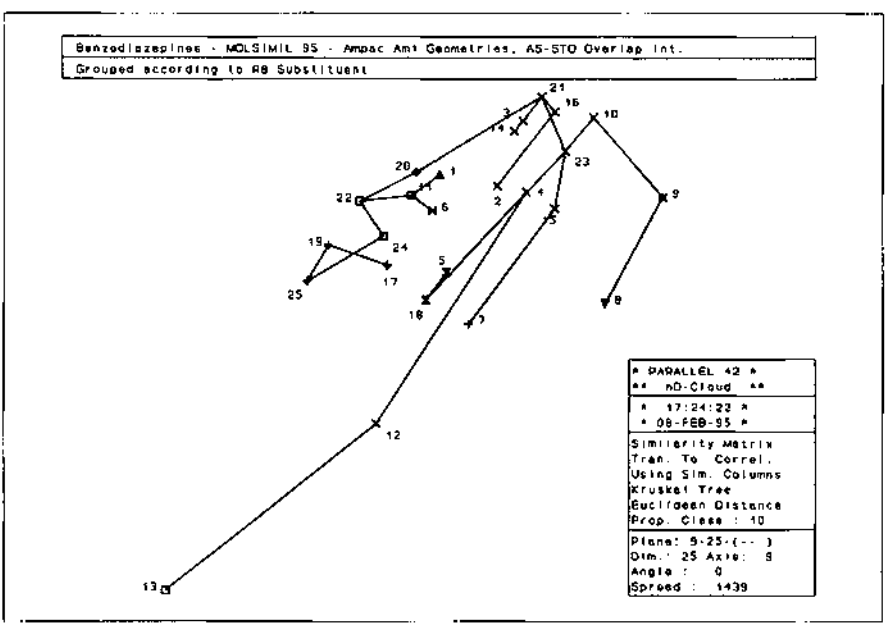

Figure 9 : Benzodiazepines Kruskal Tree, computed using the Similarity Euclidean Distance Matrix.

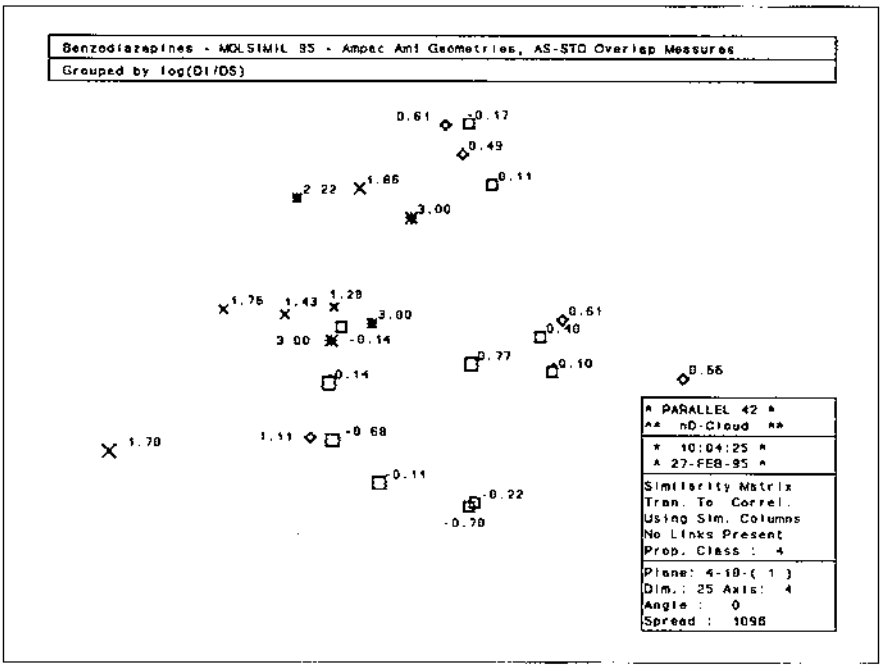

Figure 10: Projection of the benzodiazepines Molecular-Cloud into the (4,10,1) subspace.

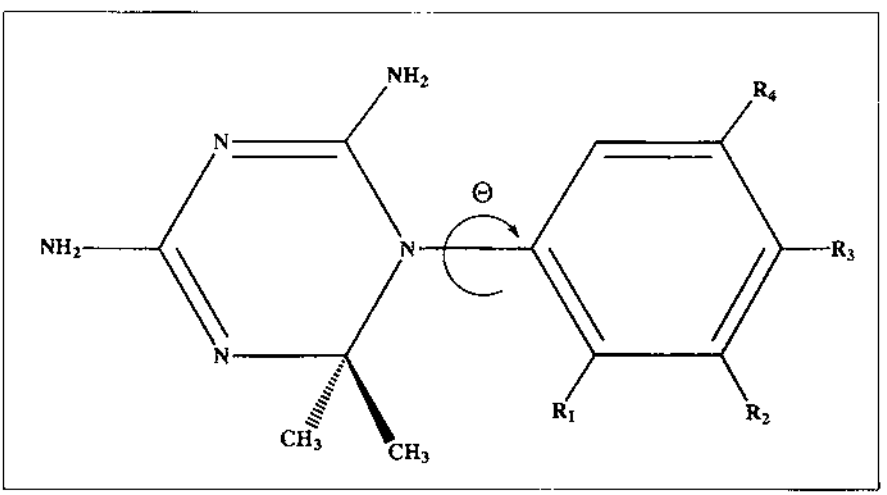

Figure 11: Structure (IV) for the Baker triazines.

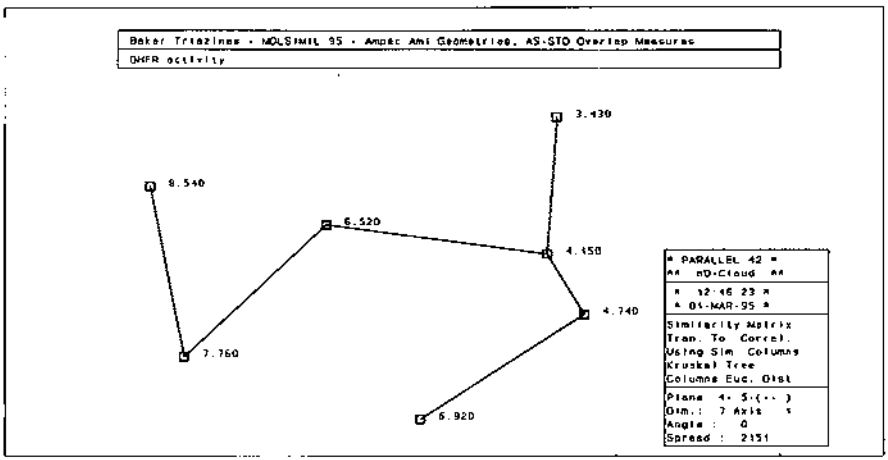

Figure 12: Baker triazines connected using a Kruskal Tree

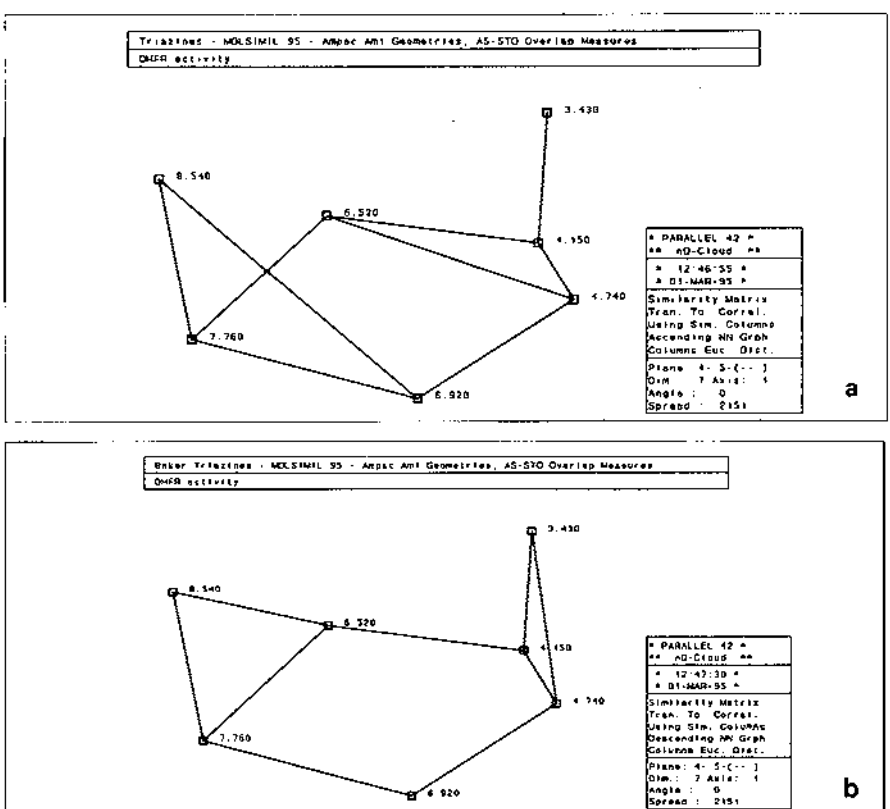

Figure 13: Baker triazines connected using two different Nearest Neighbour Graphs.