Structural Identification of Local Maxima in Low-Resolution Promolecular Electron Density Distributions

Laurence Leherte,* Laurent Dury,† and Daniel P. Vercauteren

Facultés Universitaires Notre-Dame de la Paix, Laboratoire de Physico-Chimie Informatique, Rue de Bruxelles 61, B-5000 Namur, Belgium

Received: December 21, 2002; In Final Form: September 10, 2003

In this paper, we present a theoretical method to describe molecular structures in terms of hierarchically related substructures. The approach is based on the location of local maxima (peaks) in promolecular electron density distributions (EDD) established at continuously varying resolution levels. For each of the so-calculated EDD, the local maxima are determined by using a hierarchical clustering algorithm wherein peaks obtained at a given resolution are used as starting points for discovering peaks at the next lower resolution level through gradient trajectories of the EDD. The use of such an approach allows assignment of molecular fragments or chemical groups to peaks, at any resolution level. Results, obtained for a set of four benzodiazepine-related molecules and three thrombin inhibitors, are presented in terms of dendrograms wherein each node corresponds to a well-defined molecular substructure.

Introduction

In the frame of a promolecular representation of a molecule or a crystal, a chemical structure is considered to be the superposition of independent spherical atomic contributions. Such a representation is often used in crystallography to compute electron density (ED) difference maps and thus gain information about atomic charges, chemical bonds, etc.[1]

In molecular modeling, it is extremely common to access the shape of a molecule by fitting hard spheres at the atom locations, such as the well-known van der Waals (vdW) spheres,[2] or by using derived methods as presented, for example, by Good and Richards[3] and Grant and Pickup.[4,5] In their paper, Good and Richards used atomic ED functions rather than vdW radius data to describe the shape of the atoms. Gaussian functions were fitted to atomic ED derived from ab initio quantum mechanical STO-3G calculations. To allow some hardness of the atoms when bound in a molecule, the authors used modified ED distributions which are set to zero beyond the vdW radius of the atoms. Molecular similarity calculations were then carried out for molecules composed of C, H, N, O, and S using the well-known Carbó index formula. Grant and Pickup presented a Gaussian-based description of molecular shapes, which allowed an analytical description of molecular volumes and surfaces. Comparisons were made with the conventional hard sphere representation, and very similar molecular volume values were found. The authors also proposed applications in the field of shape comparison using the so-called shape multipoles or moments. In a similar fashion, Maggiora et al.[6] used spherically symmetric Gaussian functions located on selected atoms and characterized by adjustable magnitudes and widths to modulate the degree of details needed for their protein alignment procedure. Within such a representation, each amino acid of a protein is described by a linear combination of a limited number of Gaussians, e.g., one, three, or four functions for the backbone atoms. Duncan and Olson[7] also defined molecular surfaces as a sum over atomic Gaussian functions. In their work, emphasis was given on the resolution of these surfaces, which can be modified by convolving the density with a three-dimensional (3D) Gaussian function of selected variance. Applications were proposed in the fields of molecular complementarity and molecular comparison, as well as in visual interpretation of molecular surfaces.

Tsirelson et al.,[8,9] Gironés et al.,[10,11] Mitchell and Spackman,[12] and Downs et al.[13] showed that, to a good approximation, ED distributions of various types of molecular systems can be considered as just the superposition of spherical atomic densities. Indeed, as specified by Irle et al.,[14] the atomic core densities are actually orders of magnitude higher than the atomic valence shell densities, which are still an order of magnitude larger than the genuine atomic charge deformations caused by the interatomic interactions.

Gironés et al. discussed comparisons between ab initio HF/STO-3G, HF/3-21G, and HF/6-31++G**[10] as well as 6-21G[11] molecular ED and promolecular ED. They showed that differences in molecular isodensity contours (also called MIDCOs) occur at the level of the bonds involving H atoms and at the H locations themselves. However, these differences do not significantly affect molecular similarity measurements. In their works related to the promolecular atom shell approximation (PASA), Amat and Carbó-Dorca used atomic Gaussian ED functions that were fitted on 6-311G atomic basis set results.[15] A molecular or promolecular ED distribution is thus a sum over atomic Gaussian functions wherein expansion coefficients are positive to preserve the statistical meaning of the density function in the fitted structure. Mitchell and Spackman[12] emphasized the advantage of using promolecular surfaces over the usual Corey-Pauling-Koltun (CPK) or Connolly surfaces and showed the very good qualitative and quantitative similarity between ab initio and promolecular ED isosurfaces. Downs et al.[13] compared procrystal ED calculations applied to silicates and oxide materials with results obtained by using the programs CRYSTAL98[16] and TOPOND96.[17] They found out that for

* Corresponding author. E-mail: laurence.leherte@fundp.ac.be.
† FRIA Ph.D. Fellow.

10.1021/jp022641h CCC: $25.00
Published on Web 10/24/2003

every bond critical point (BCP) located in ab initio ED distributions, an equivalent one was found in the procrystal ED model, with similar magnitudes of the ED, and at similar positions along the bonds. The curvatures of the ED obtained from both approaches were also highly correlated. The largest discrepancies between the CP properties were the curvature values of the density perpendicular to the bond paths. Chemical bonding was also discussed by Tsirelson et al.[8,9] who studied the critical points (CP) of experimental, ab initio 6-11G+, 7-311G*, 8-511G*, and 8-411G* Hartree−Fock, and promolecular ED distributions of crystalline MgO and rock salt crystals. The authors observed similar CP characteristics in experimental and procrystal ED maps. They, therefore, asked whether one could consider the existence of BCPs as a sign of chemical bonding, as postulated by Bader,[18] and concluded that the topological analysis of a procrystal could provide an a priori estimate of the structural features and properties of real rock salt crystals.

The usefulness of promolecular ED representations was also supported by an earlier work of Spackman and Maslen,[19] who compared various promolecular properties, such as interaction energies, electronegativity, atomic radii, and charges, to the corresponding quantities obtained from experimental studies for diatomic species. These authors showed that a promolecular ED could generally explain a large part of the experimental observables. In their model, the promolecular ED distribution was simply a sum over spherically averaged atomic ED wherein the interaction energy is reduced to a sum of pairwise electrostatic energies. At the same time, Maksic et al.[20] studied the second moments and diamagnetic susceptibilities of various linear vdW complexes. Their formalism included small corrections arising from the spatial extension of the atomic orbitals. It appeared that those two properties were not very sensitive to the degree of intramolecular charge transfer, unless atoms with highly different electronegativity values were involved.

The use of molecular properties such as the ED reconstructed in the frame of a promolecular representation presents several essential advantages. First of all, computation times are drastically reduced. This is especially true and appealing when studying macromolecular systems.[21] Second, a promolecular representation expressed in terms of Gaussian functions allows the calculations of numerous properties using analytical formulas such as overlap integrals appearing in similarity measurements,[3,6,21] surface curvatures,[7] surface areas and volumes,[4,5] and convolution products with another Gaussian.[7,22] We have exploited this last advantage in recent works on multiresolution CP analyses[23,24] of molecular ED distributions for molecular superposition approaches.[23−26] It was particularly shown how low-resolution graph representations of molecular structures could be used to easily compare different drug molecules. These graph representations, composed of vertexes and edges that were generated using a CP analysis[27] of low resolution ED maps, led to pharmacophore models of biologically active/affine molecules. Working at low-resolution values actually allows a reduction in the number of CPs which are no longer located on atoms but are rather associated with groups of atoms. The molecular descriptors that were considered were the density values and distances between the CPs of low-resolution ED maps. As this approach showed to work well, a further step was to analyze why local low-resolution density values and simple geometrical parameters are good and apparently sufficient molecular descriptors. This led us to the present work; i.e., the implementation of a method to provide a chemical meaning to low-resolution critical points.

Describing molecular structures in terms of reduced or superatom (groups of atoms) representations is not new. But, to our knowledge, this has never been achieved using ED data. However, we mention here a work achieved by Zimpel and Mezey[28] who proposed a tree representation of the evolution of the topological description of ED isocontours as a function of a varying density threshold.

Regarding reduced graph representations, Takahashi et al.[29] proposed an approach to generate reduced graphs wherein each node describes a functional atomic group and each edge is weighted according to the topological path length (the number of bonds) between the two corresponding chemical groups. The obtained reduced representations were used in a similarity search procedure. Later, Rarey and Dixon[30] also used reduced molecular graphs wherein nodes represented fragments of the molecule. Their so-called feature trees were built, first, by identifying and then reducing rings. For the acyclic part of the molecule, atoms with more than one bond form a separate node in the tree, while end atoms are merged with their linked atom. Edges were added between nodes containing atoms that are connected in the molecule. Steric and chemical attributes, such as the vdW volume, the H-donor/acceptor character, etc. could also be assigned to each node. In research on the development of algorithms for structure search in organic reaction databases, original methods were also proposed for fast retrieval of structural information using reduced graph representations.[31,32] It was shown how partitioning the molecules could be achieved on the basis of ring and nonring components or carbon and heteroatom groups, as in research works reported by Gillet et al.[33,34] In all the above-mentioned works, graphs are obtained by reducing chemical groups to points. A different strategy is described in the present work wherein points are related to chemical groups.

The aim of the present work is to establish a hierarchical description of CP graphs based on promolecular ED representations in order to assign peaks (or local maximum of the ED function) observed at a given resolution to a chemical function or group of atoms. In this context, every peak at a given resolution is related to peaks obtained at a higher resolution level and eventually to the constituting atoms (seen as the highest resolution peaks) of the molecular structure. It is thus expected that CPs of a low-resolution graph are characterized not only by their local density values but also by a chemical meaning for further applications in molecular similarity search, for example. To our knowledge, no attempts have been made so far to assign a molecular substructure to CPs at a resolution value lower than the atomic resolution.

In the following sections, a brief mathematical description of the PASA approach and the hierarchical clustering of CPs is first presented. It is followed by results from applications to a set of benzodiazepine-related molecules and thrombin inhibitors. These systems were chosen following previous studies on the same molecules.[23−35]

Theoretical Background

Atomic Shell Approximation. In the atomic shell approximation (ASA) approach, a promolecular electron density (ED) distribution ρ_M is calculated as a weighted summation over atomic ED distributions ρ_a, which are described in terms of series of squared 1s Gaussian functions fitted from atomic basis set representations:[36]

$$\rho_a(\mathbf{r}-\mathbf{R}_a)=\sum_{i=1}^{5} w_{a,i}\left[\left(\frac{2\zeta_{a,i}}{\pi}\right)^{3/4} e^{-\zeta a,i\left|\mathbf{r}-\mathbf{R}_a\right|^2}\right]^2 \qquad (1)$$

TABLE 1: $w_{a,i}$ and $\zeta_{a,i}$ Atomic Shell Approximation Coefficients as Reported in Ref 37 for C Atom ($Z_a = 6$)

i	$w_{a,i}$	$\zeta_{a,I}$, bohr^{-2}
1	0.507 950 102 1	0.206 511 393 9
2	0.188 826 528 4	0.555 342 809 7
3	0.176 510 617 4	7.353 198 307 6
4	0.112 597 344 7	22.087 591 762 5
5	0.014 115 407 4	107.346 385 025 0

where $\mathbf{R}_a$ is the position vector of atom a, and $w_{a,i}$ and $\zeta_{a,i}$ are the fitted parameters, respectively, as reported in ref 37. ρ_M is then calculated as

$$\rho_M = \sum_a Z_a \rho_a \tag{2}$$

where Z_a is the atomic number of atom a. As an example, $w_{a,i}$ and $\zeta_{a,i}$ coefficients fitted from the atomic 3-21G basis set for the carbon atom are reported in Table 1.[37]

In our approach to generate low-resolution three-dimensional (3D) functions, an ED map is a deformed version of ρ_M that is directly expressed as the solution of the diffusion equation according to the formalism presented by Kostrowicki et al.:[22]

$$\rho_{a,t}(\mathbf{r} - \mathbf{R}_a) = \sum_{i=1}^{5} a_{a,i}(1 + 4b_{a,i}t)^{-3/2} e^{-b_{a,i}|\mathbf{r}-\mathbf{R}_a|^2/(1+4b_{a,i}t)} \tag{3}$$

where

$$b_{a,i} = 2\zeta_{a,i} \quad \text{and} \quad a_{a,i} = w_{a,i}\left(\frac{b_{a,i}}{\pi}\right)^{6/4} \tag{4}$$

In this context, t is seen as the product of a diffusion coefficient with time. This expression can also be seen as the result of a Fourier−Poisson integral with dimensionality $d = 3$:[22]

$$F(\mathbf{r},t) = \frac{1}{(4\pi t)^{d/2}} \int f(\mathbf{s}) e^{-|\mathbf{r}-\mathbf{s}|^2/4t}\, d\mathbf{s} \tag{5}$$

On a similar basis, Duncan and Olson[7] generated low-resolution molecular surfaces by convolving the density function with a 3D Gaussian function $g(\mathbf{r},t)$ of appropriate variance:

$$g(\mathbf{r},t) = \frac{1}{(4\pi t)^{3/2}} e^{-|\mathbf{r}|^2/4t} \tag{6}$$

Hierarchical Clustering/Merging Algorithm. To follow the pattern of critical points (CP), and more particularly peaks, as a function of resolution, we implemented the algorithm presented by Leung et al.[38] This algorithm was originally established to cluster data by modeling the blurring effect of lateral retinal interconnections based on scale space theory. In scale space theory, it is considered that an image $f(\mathbf{r})$ is embedded into a continuous family $F(\mathbf{r},t)$ of gradually smoother versions of it. The original image corresponds to the scale $t = 0$ and increasing the scale simplifies the image. It is proven that one can blur the image in a unique and sensible way in which $F(\mathbf{r},t)$ is the convolution of $f(\mathbf{r})$ with the Gaussian kernel $g(\mathbf{r},t)$.[38]

Leung et al.'s work was adapted to our needs; for example, $F(\mathbf{r},t)$ is now considered a molecular electron density (ED) distribution function. The various steps of the resulting clustering/merging algorithm are as follows:

1. At scale $t = 0$, each atom of a molecular structure is considered as a local maximum (peak) of the promolecular ED distribution function. They are consequently considered as the starting points of the merging procedure described below.

2. As t increases continuously from 0.0 to a given maximal value, each peak moves continuously along the gradient path to reach a location in the 3D space where $\nabla\rho = 0$. On a practical point of view, this consists of following the trajectory of the peaks obtained at a resolution t on the ED distribution surface calculated at resolution level $t + \Delta t$:

$$\mathbf{r}_{\text{peak}}(t) = \mathbf{r}_{\text{peak}}(t - \Delta t) + \frac{\Delta}{\rho_{\text{peak}}(t - \Delta t)} \nabla\rho_{\text{peak}}(t) \tag{7}$$

During the testing phase of this approach, it appeared that the density of the peak, $\rho_{\text{peak}}(t - \Delta t)$, had to be included in the algorithm in order to modulate the magnitude of the displacement step, Δ. Indeed, for heavy atoms such as chlorine or sulfur, too large a displacement led to instabilities in the trajectories and final divergence. This phenomenon was also observed when testing a Newton−Raphson type algorithm.

The trajectory search is stopped when $\nabla\rho_{\text{peak}}(t)$ is lower or equal to a limit value, grad_{lim}. Once all peak locations are found, close peaks are merged if their interdistance is lower than the initial value of $\Delta^{1/2}$. The procedure is repeated for each selected value of t.

If the initial Δ value is too small to allow convergence toward a local maximum of the ED within the given number of iterations, its value is doubled (a scaling factor that is arbitrarily selected) and the procedure is repeated until final convergence.

3. The procedure can be carried out until the whole set of maxima becomes one single point. This would be the ultimate stopping criteria of the merging procedure. In that case, the final point that would be obtained will correspond to the whole molecular structure.

This approach can be seen as the inverse of the method proposed by Kostrowicki et al.[22] wherein the global optimization of atom clusters was achieved by progressively increasing the value of t in order to start from a potential hypersurface with one global minimum to reach the real high-resolution potential hypersurface with many energy minima.

The results obtained with our algorithm can be interpreted in terms of dendrograms. Visual results were generated by using the Web version of the program Phylodendron.[39] Input data were written in the adequate format using DENDRO,[40] a homemade program implemented using Delphi,[41] an object-oriented programming language that allows representation and processing of data in terms of classes of objects.

Applications. Our critical point (CP) clustering/merging procedure was applied to biological ligands that have already been studied through multiresolution analysis methods.[23,24,35] Our first set of molecular ligands included benzodiazepine-related compounds which present affinity values to both their central and mitochondrial receptors: Diazepam (7-chloro-1,3-dihydro-1-methyl-5-(phenyl)-2*H*-1,4-benzodiazepin-2-one), Ro5-4864 (7-chloro-1,3-dihydro-1-methyl-5-(*p*-chlorophenyl)-2*H*-1,4-benzodiazepin-2-one), QUIZ (6-chloro-*N*1-methyl-4-phenyl-quinazolin-2-one), and TZQ ([1,2,4]triazolo[1,5-*c*]quinazolin-5(6*H*)-one) (Figure 1). The three-dimensional (3D) atomic coordinates of the considered ligands were derived from crystallographic data. Two structures are reported in the Cambridge Crystallographic Database:[42] Diazepam (DIZPAM10) and Ro5-4864 (FULWUE). Atom coordinates of molecules TZQ and QUIZ were obtained from X-ray diffraction experiments from our institution.[43]

Several pharmacophore models for describing benzodiazepine affinity toward their receptors have already been published (see

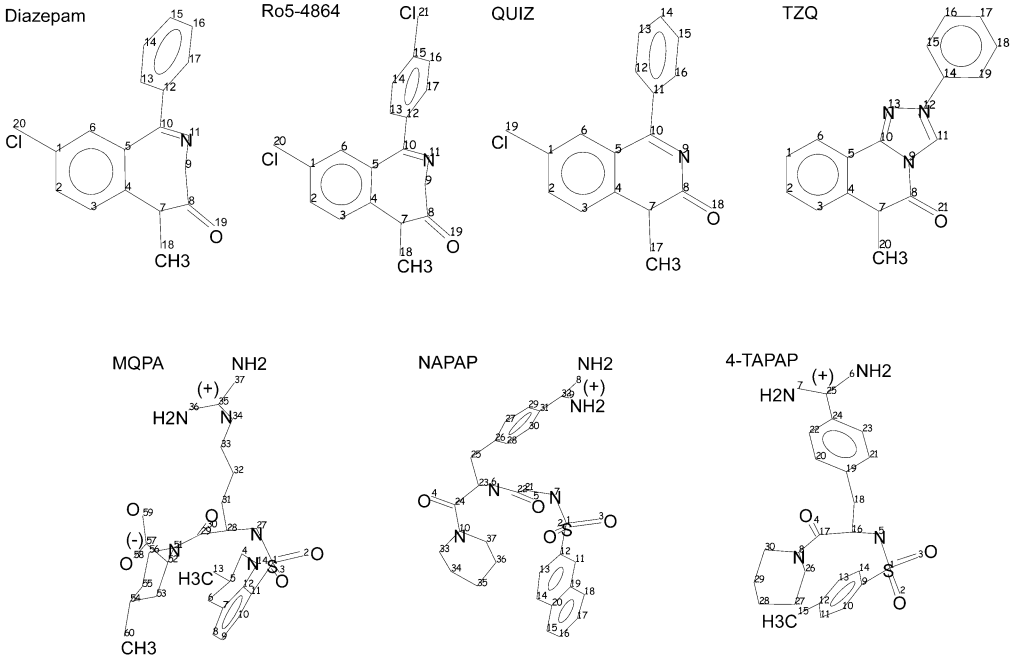

Figure 1. Structure formula and atomic numbering of benzodiazepine-related compounds (Diazepam, Ro5-4864, QUIZ, TZQ) and thrombin inhibitors (MQPA, NAPAP, 4-TAPAP). Hydrogen atoms are not shown for clarity.

ref 35 and references therein). Besides the presence of a rigid planar region constituting of the benzo rings, all models possess each of the following features or a selected combination of them: (a) a proton acceptor region which includes a carbonyl group; (b) a lipophilic or steric group, i.e., the benzo ring and/or its chlorine atom; (c) a free rotating aromatic region, i.e., a phenyl group.

Another set of molecules that belong to the thrombin inhibitor family was selected following the work of Nissink et al.[44] These authors compared three thrombin inhibitors, (2*R*,4*R*)-4-methyl-1[*N*α-((*RS*)-3-methyl-1,2,3,4-tetrahydro-8-quinolenesulfonyl)-L-arginyl]-2-piperidine carboxylic acid (MQPA), *N*α-(2-naphthylsulfonyl-glycyl)-DL-*p*-amidino-phenylalanyl-piperidine (NAPAP), and *N*α-(4-toluenesulfonyl)-DL-*p*-amidinophenylalanylpiperidine (4-TAPAP),[45] at the particular cristallographic resolution value of 2.5 or 3 Å. All three selected molecules adopt a star shape with three branches terminated by either a cyclic substructure linked to a sulfonyl function, a piperidine ring, and a guanidinium (in MQPA) or an amidino group (Figure 1). These three structures present various groups involved in H-bonds and hydrophobic interactions with their receptor, e.g., on one hand, guanidinium and carboxylate (in MQPA only), amidino groups, carbonyl, and amine groups that are present in the inhibitor skeleton, sulfone groups (forming H-bonds with solvent molecules) and, on the other hand, piperidine ring, alkyl chain (MQPA), quinolene (MQPA), β-naphthyl (NAPAP), and toluene (4-TAPAP). The 3D atom coordinates of the considered protein ligands were derived from crystallographic data reported in the Protein Data Bank (PDB).[46] The PDB codes for MQPA, NAPAP, and 4-TAPAP interacting with the thrombin receptor are 1etr, 1ets, and 1ett, respectively. H atoms were added to the crystalline PDB structures using InsightII[47] in order to generate the structures described in ref 45.

Electron Density Map Calculations. The procedure that is described in the theoretical section does not require any calculation of the ED maps. It is based solely on the knowledge of the analytical expression of the promolecular ED function and its first derivative. However, to generate graphical representations of the ED distributions, ED maps were calculated by using eq 3. The map size was set to 128 × 64 × 128 and 128 × 128 × 128 for the benzodiazepine-related molecules and the thrombin inhibitors, respectively, with a grid interval equal to 0.125 Å.

Hierarchical Clustering/Merging Results. To generate a hierarchical description of the seven molecular structures under study, our clustering/merging algorithm was used with the following input parameters for each molecular structure: t varies between 0 and 4.0 bohr^2, with a step Δt set equal to 0.05 bohr^2. The initial value Δ_{init} is equal to 0.000 001 bohr^2, and $grad_{lim}$ is 0.0001 e$^-$/bohr^4. The value of Δ is doubled if convergence is not observed after 2000 iterations. Calculations were run on one node of a bi-PentiumIII 1GHz and took 1h16′ (Diazepam), 1h20′ (Ro5-4864), 1h03′ (QUIZ), 1h15′ (TZQ), 4h24′ (MQPA), 4h46′ (NAPAP), and 3h07′ (4-TAPAP).

The approach described in this paper allows an easy partitioning of a molecular structure through a hierarchical clustering/merging of atoms with resolution. Results of that procedure were expressed in terms of dendrograms (Figures 2 and 3). By that way, molecular fragments can be assigned to peaks observed at any resolution value. In Figures 2 and 3, each leaf of the dendrograms is labeled according to the atom numbers given in Figure 1 and corresponds to resolution $t = 0.0$ bohr^2. All atom numbers which are not present in Figure 1 actually represent H atoms. Reduced representations, i.e., locations of the local maxima of the ED distributions at $t > 0.0$ bohr^2, are displayed in Figures 4−11 for each set of molecular structures at four different resolution levels $t = 1.1$, 1.6, 1.9, and 2.5 bohr^2. Fragments obtained at those four values of t for the benzodiazepine-related compounds and thrombin inhibitors are shown in Figures 12 and 13, respectively. At those resolution values, each molecule is reduced to a limited number of points (local maximum of the ED) as illustrated in Figures 4−7 for the benzodiazepine-related compounds and in Figures 8−11 for the thrombin inhibitors. To show how the information that is present in the dendrograms (Figures 2 and 3) is related to the local maxima of the ED maps (Figures 4−11) and to their corresponding molecular substructure (Figures 12 and 13), we added

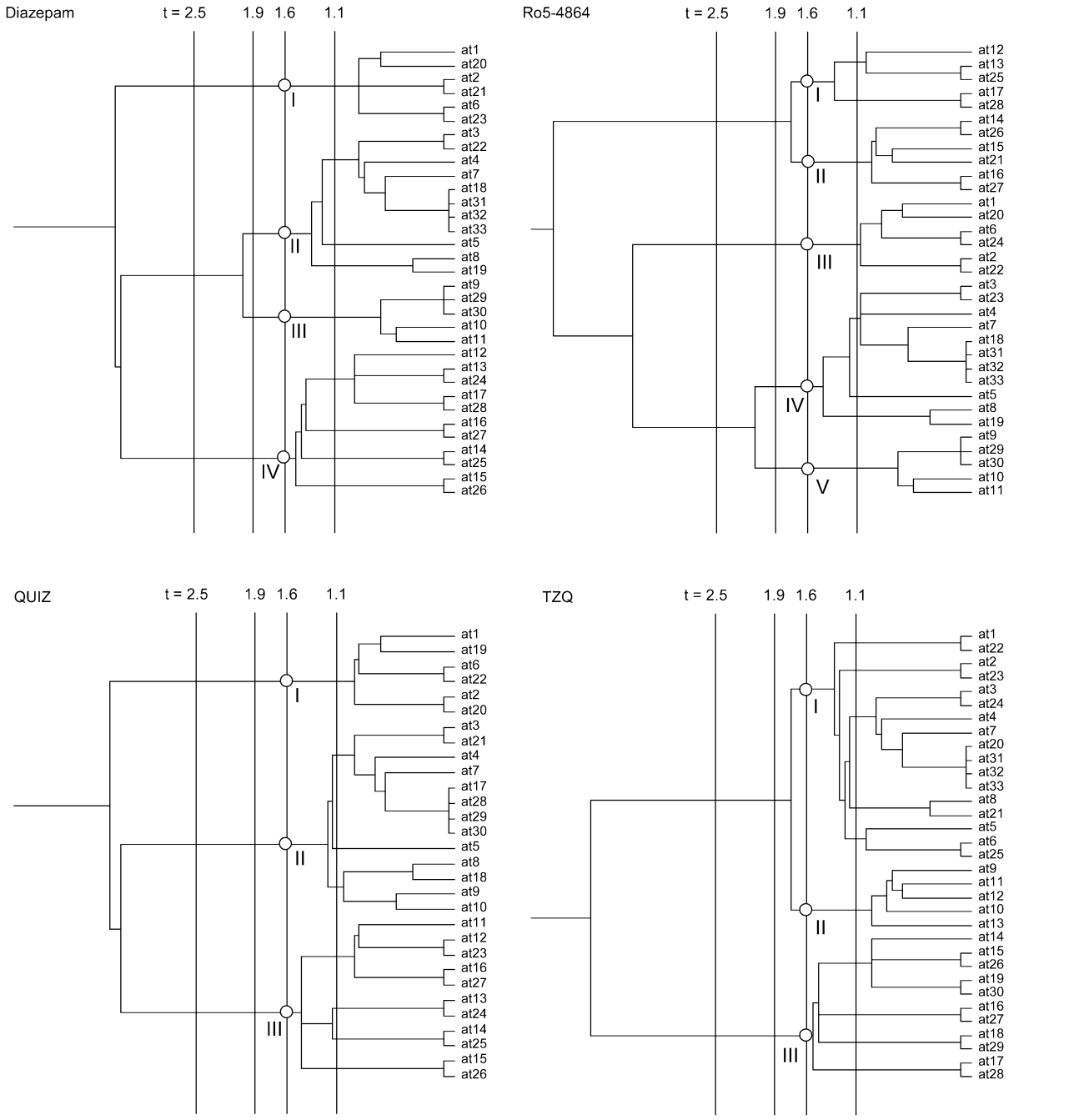

Figure 2. Dendrograms obtained from the application of a hierarchical clustering/merging procedure applied to PASA electron density maps of benzodiazepine-related molecules. Results at $t = 1.1$, 1.6, 1.9, and 2.5 bohr^2 are emphasized by vertical lines. Roman numbers point to the electron density maxima shown in Figure 5. Drawings were generated using Phylodendron.[39]

labels (roman numbers) for the particular case of $t = 1.6$ bohr^2. The same exercise can obviously be done at any resolution level.

In the next section, the obtained dendrograms are analyzed so as to assign molecular substructures (atoms or groups of atoms) to local maxima of the ED distributions as a function of the resolution level.

Dendrograms Analysis. Analysis of the dendrograms (Figures 2 and 3) shows that the resolution-based clustering/merging occurs as follows. First of all, some common features are observed in both the benzodiazepine-related molecules and thrombin inhibitors. All H atoms are merged with their linked C or N atom at $t = 0.05-0.1$ bohr^2. Afterward, come the carbonyl groups at $t = 0.4-0.45$ bohr^2, regardless of the molecular structure. It is also observed that the C atoms that are parts of unsubstituted phenyl or cyclohexyl groups as in Diazepam, QUIZ, TZQ, or NAPAP, 4-TAPAP, respectively, stay isolated up to $t = 1.1$ bohr^2. When the ring is substituted, the C atom bearing the substituent acts as an attractive group with respect to its neighboring atoms. This is observed, for example, in various atom sequences such as (1 2 6 20) and (12 13 17) in Diazepam, (14 15 16 21), (12 13 17), and (1 2 6 20) in Ro5-4864, (1 2 6 19) and (11 12 16) in QUIZ, (14 15 19) in TZQ, (51 52 56) and (53 54 55 60) in MQPA, (10 33 37), (26 27 28), and (29 30 31) in NAPAP, and (19 20 21) and (22 23 24) in 4-TAPAP.

Besides those common features, atoms are merged on a different basis as shown by the analysis presented hereafter for the two sets of molecules under study.

Benzodiazepine-Related Molecules. In the case of the benzodiazepine-related molecules, the merging of the C═N groups

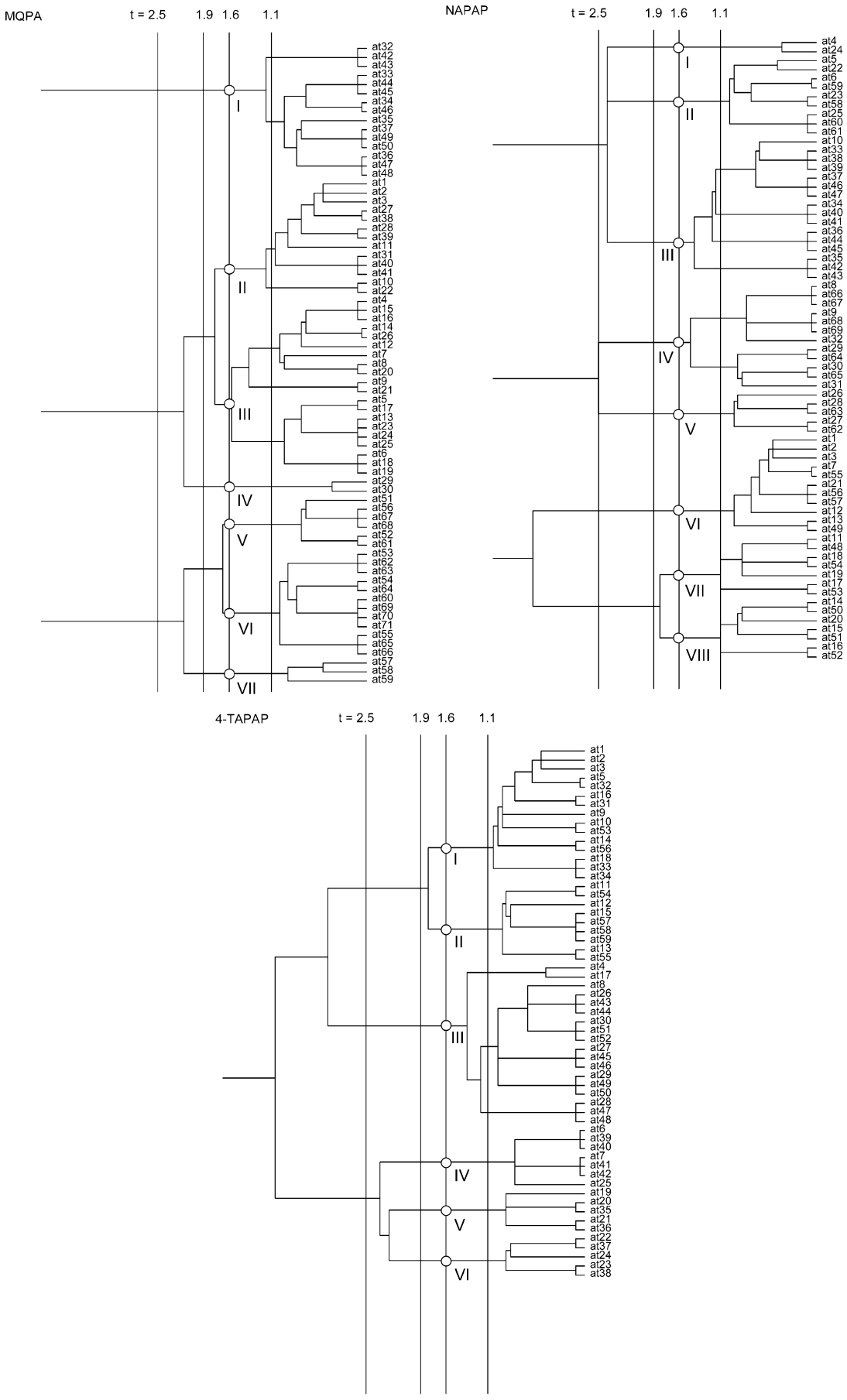

Figure 3. Dendrograms obtained from the application of a hierarchical clustering/merging procedure applied to PASA electron density maps of thrombin inhibitors. Results at $t = 1.1$, 1.6, 1.9, and 2.5 bohr^2 are emphasized by vertical lines. Roman numbers point to the electron density maxima shown in Figure 9. Drawings were generated using Phylodendron.[39]

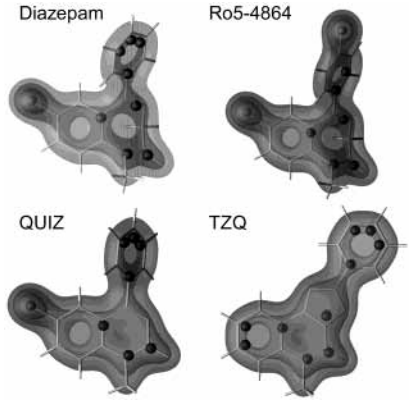

Figure 4. Isocontours of the electron density distribution of benzodiazepine-related compounds calculated with the PASA formalism at $t = 1.1$ bohr^2 (iso = 0.1, 0.15, 0.2 e$^-$/bohr^3). Local maxima of the density are shown as spheres and were obtained by using the hierarchical clustering/merging approach. Figures were generated by using DataExplorer.[48]

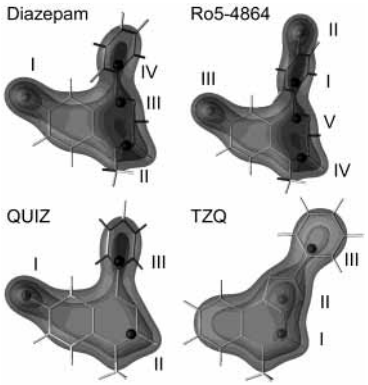

Figure 5. Isocontours of the electron density distribution of benzodiazepine-related compounds calculated with the PASA formalism at $t = 1.6$ bohr^2 (iso = 0.1, 0.125, 0.15 e$^-$/bohr^3). Local maxima of the density are shown as spheres and were obtained by using the hierarchical clustering/merging approach. Roman numbers correspond to the labels shown in Figures 2 and 12. Figures were generated by using DataExplorer.[48]

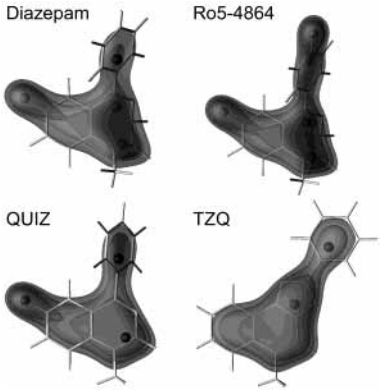

Figure 6. Isocontours of the electron density distribution of benzodiazepine-related compounds calculated with the PASA formalism at $t = 1.9$ bohr^2 (iso = 0.1, 0.11, 0.12 e$^-$/bohr^3). Local maxima of the density are shown as spheres and were obtained by using the hierarchical clustering/merging approach. Figures were generated by using DataExplorer.[48]

occurs at $t = 0.55$ bohr^2 for Diazepam ($d_{at10-at11} = 1.29$ Å), Ro5-4864 ($d_{at10-at11} = 1.29$ Å), and QUIZ ($d_{at9-at10} = 1.30$ Å) and at a larger $t = 0.65$ bohr^2 in TZQ wherein the distance $d_{at11-at12} = 1.32$ Å is also longer. In this very last case, the merging of atoms 11 and 12 of TZQ corresponds to the first

Figure 7. Isocontours of the electron density distribution of benzodiazepine-related compounds calculated with the PASA formalism at $t = 2.5$ bohr^2 (iso = 0.05, 0.075, 0.01 e$^-$/bohr^3). Local maxima of the density are shown as spheres and were obtained by using the hierarchical clustering/merging approach. Figures were generated by using DataExplorer.[48]

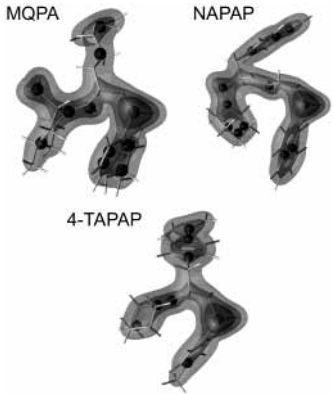

Figure 8. Isocontours of the electron density distribution of thrombin inhibitors calculated with the PASA formalism at $t = 1.1$ bohr^2 (iso = 0.1, 0.15, 0.2 e$^-$/bohr^3). Local maxima of the density are shown as spheres and were obtained by using the hierarchical clustering/merging approach. Figures were generated by using DataExplorer.[48]

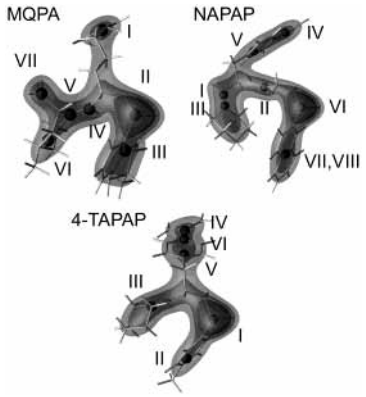

Figure 9. Isocontours of the electron density distribution of thrombin inhibitors calculated with the PASA formalism at $t = 1.6$ bohr^2 (iso = 0.1, 0.125, 0.15 e$^-$/bohr^3). Local maxima of the density are shown using spheres and were obtained using the hierarchical clustering/merging approach. Roman numbers correspond to the labels shown in Figures 3 and 13. Points VII and VIII of NAPAP are extremely close and appear superposed. Figures were generated by using DataExplorer.[48]

step of the reduction of the triazolo ring. Then follows the merging of the N-Me groups at $t = 0.6-0.65$ bohr^2. Finally, motif I in Figure 14 appears at $t = 0.9-1.0$ bohr^2. Then, depending on the structural similarities among the molecules, that last motif leads to motif II at $t = 1.15-1.25$ bohr^2, except

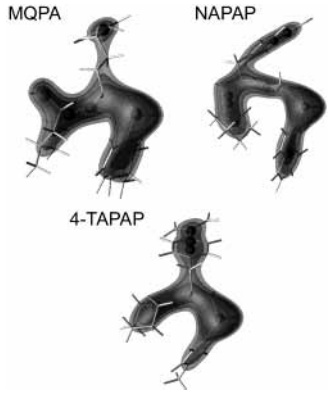

Figure 10. Isocontours of the electron density distribution of thrombin inhibitors calculated with the PASA formalism at $t = 1.9$ bohr^2 (iso = 0.1, 0.11, 0.12 e$^-$/bohr^3). Local maxima of the density are shown as spheres and were obtained by using the hierarchical clustering/merging approach. Figures were generated by using DataExplorer.[48]

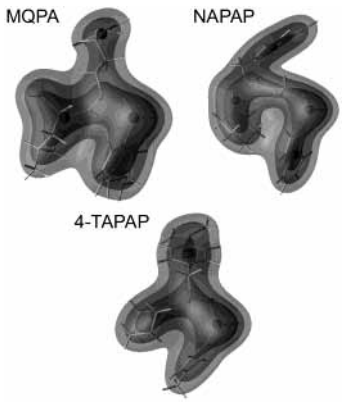

Figure 11. Isocontours of the electron density distribution of thrombin inhibitors calculated with the PASA formalism at $t = 2.5$ bohr^2 (iso = 0.05, 0.075, 0.01 e$^-$/bohr^3). Local maxima of the density are shown as spheres and were obtained by using the hierarchical clustering/merging approach. Figures were generated by using DataExplorer.[48]

for TZQ wherein there is no Cl atom which acts as an attraction center toward atom 6, as in the three other molecules. In this case, motif III is obtained, i.e., a merging occurs between substructure I and the carbonyl group at $t = 1.15$ bohr^2. Similarly, motif IV (Figure 14) appears at $t = 0.9-1.05$ bohr^2 for the first three molecules only, after merging of Cl−C at $t = 0.6-0.75$ bohr^2. This last motif stays unchanged through a large range of resolution values, i.e., up to $t = 3.2-3.25$ bohr^2. In benzodiazepines, the carbonyl groups also stay isolated from the other substructures on a large range of resolution values. Their corresponding ED peaks are not merged with their adjacent N or C atom (atom number 9 in Figure 1) before $t = 1.1$ bohr^2, except for compound QUIZ wherein the chemical bond between atoms 8 and 9 is shorter ($d_{at8-at9} = 1.39$ Å) than in the other compounds (Diazepam, $d_{at9-at11} = 1.51$ Å; Ro5-4864, $d_{at9-at11} = 1.51$ Å; TZQ, $d_{at9-at11} = 1.41$ Å), due to electron delocalization.

High-density atoms are attraction centers which lead to closely related local ED maxima, even in low-resolution ED distributions. Various resulting high-density centers are displayed in Figure 15 (motifs V−VIII) together with the t value at which they occur. Regardless of the molecular structure, each category of substructures is formed at very similar resolution values: 0.7−0.95, 0.9−1.05, 1.2−1.4, and 1.45−1.70 bohr^2 for motifs V, VI, VII, and VIII, respectively. At $t = 1.6$ bohr^2, the molecular structures are eventually clearly partitioned into four groups which are associated with H-acceptor (V), steric/hydrophobic (VI), H-acceptor (VII), and steric/hydrophobic groups (VIII) as possible pharmacophore elements[35] (Figure 15). The local maxima corresponding to those groups are located close to the chlorine atoms or the benzo group (VI), the carbonyl group (VII), and the phenyl group (VIII) and, with the exception of compounds QUIZ and TZQ, a fourth peak is observed in the proximity of the imine group of the molecular structures V (Figure 5). At $t = 2.5$ bohr^2, there is only a small number of substructures left, as shown in Figure 7. Final complete merging of the whole molecular structures into a single peak occurs at $t = 3.2$ for Diazepam, 4.5 for Ro5-4864, 3.25 for QUIZ, and 3.6 bohr^2 for TZQ.

The analysis of the peak density values (Table 2) shows that, as expected, the highest values are associated with the Cl atoms, followed by the carbonyl and N-containing groups, and the lowest peaks are those originating from groups that are constituted of C atoms only. That tendency is reversed at the

TABLE 2: Values of the Electron Density (e$^-$/bohr^3) at Local Maxima Locations in Benzodiazepine-Related Electron Density Distributions Calculated with the PASA Approximation at Various Resolution Values t (bohr^2)

	approximate peak location			
	$t = 1.1$ Figure 4	$t = 1.6$ Figure 5[a]	$t = 1.9$ Figure 6	$t = 2.5$ Figure 7
		diazepam		
Cl (benzo)	0.2765	0.1778 (I)	0.1478	0.1106
carbonyl	0.1852	0.1464 (II)	0.1330	0.1145
amine	0.1827			
imine	0.1713	0.1358 (III)	0.1244	
phenyl	0.1590	0.1289 (IV)	0.1175	0.1000
phenyl	0.1522			
phenyl	0.1521			
phenyl	0.1519			
fused rings	0.1603			
$\rho_{max} - \rho_{min}$	0.1246	0.0489	0.0303	0.0145
		Ro5-4864		
Cl (benzo)	0.2777	0.1788 (III)	0.1478	0.1114
carbonyl	0.1854	0.1462 (IV)	0.1326	0.1145
amine	0.1820			
imine	0.1719	0.1365 (V)	0.1251	
phenyl	0.1584	0.1293 (I)		
phenyl	0.1543			
Cl (phenyl)	0.2760	0.1873 (II)	0.1464	0.1103
fused rings	0.1620			
$\rho_{max} - \rho_{min}$	0.1234	0.0580	0.0227	0.0042
		QUIZ		
Cl (benzo)	0.2767	0.1780 (I)	0.1470	0.1107
carbonyl	0.1898	0.1503 (II)	0.1358	0.1156
amine	0.1839			
phenyl	0.1600	0.1299 (III)	0.1185	0.1007
phenyl	0.1526			
phenyl	0.1537			
phenyl	0.1538			
fused rings	0.1624			
$\rho_{max} - \rho_{min}$	0.1241	0.0481	0.0285	0.0149
		TZQ		
fused phenyl	0.1520			
fused phenyl	0.1517			
carbonyl	0.1906	0.1524 (I)		
amine	0.1835			
triazolo	0.2017	0.1634 (II)	0.1468	0.1229
phenyl	0.1589	0.1286 (III)	0.1172	0.0994
phenyl	0.1523			
phenyl	0.1523			
phenyl	0.1518			
fused rings	0.1625			
$\rho_{max} - \rho_{min}$	0.0500	0.0348	0.0296	0.0235

[a] Roman numbers correspond to the labels shown in Figures 2, 5, and 12.

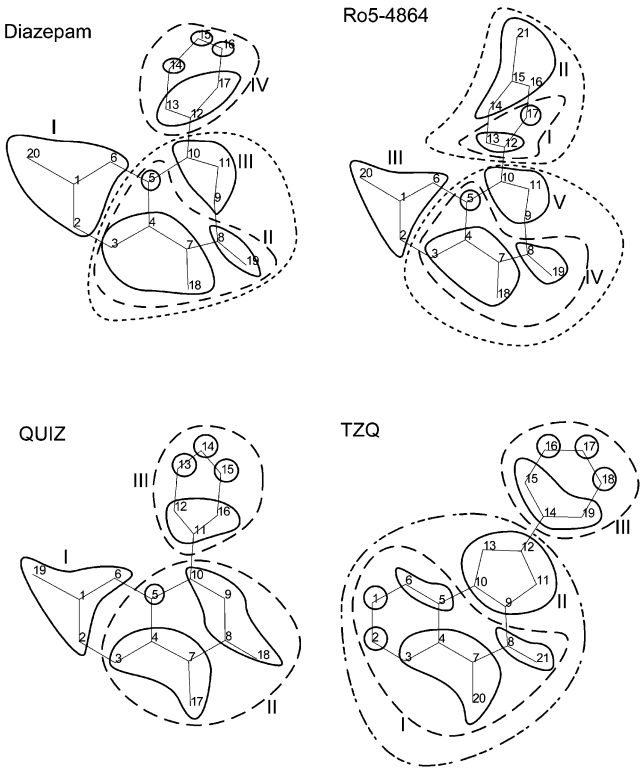

Figure 12. Contours showing the molecular fragments obtained at resolution values $t = 1.1$ (plain line), 1.6 (long dashed line), 1.9 (dash−dot line), and 2.5 (dotted line) bohr^2 with the hierarchical clustering/merging approach for benzodiazepine-related molecules. Roman numbers correspond to the labels shown in Figures 2 and 5.

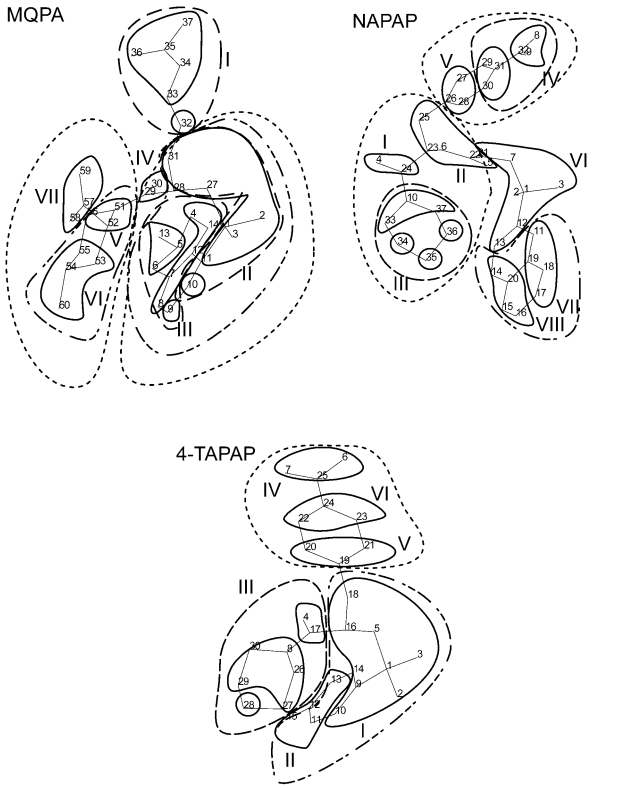

Figure 13. Contours showing the molecular fragments obtained at resolution values $t = 1.1$ (plain line), 1.6 (long dashed line), 1.9 (dash−dot line), and 2.5 (dotted line) bohr^2 with the hierarchical clustering/merging approach for thrombin inhibitors. Roman numbers correspond to the labels shown in Figures 3 and 9.

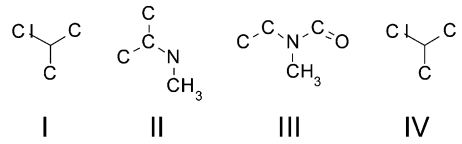

Figure 14. Selected intermediate motifs occurring in benzodiazepine-related compounds during the clustering/merging procedure.

lowest value of t, i.e., at $t = 2.5$ bohr^2; the density values associated with the chlorine atoms ρ_{Cl} becomes the lowest density value at that particular resolution value (Table 2). At that resolution, all discrepancies among the peaks are leveled out as shown by the $\rho_{max} - \rho_{min}$ values reported in Table 2.

Thrombin Inhibitors. A similar analysis of the dendrograms was carried out for the three selected thrombin inhibitors (Figure 3). As in Figure 2, atom labels correspond to the numbering given in Figure 1. After initial merging of H and C═O groups, as described before, all SO_2 groups appear as a single substructure at $t = 0.5$ bohr^2. Guanidinium, as in MQPA, and amidino groups are fully merged at $t = 0.8$ bohr^2. The partitioning of the three molecular structures leads to well-defined substructures as in benzodiazepine-related compounds (Figure 16). Local maxima in the ED occur on the hydrophilic guanidinium and amidino groups as shown by motifs I, which appear at values of t between 0.8 and 0.95 bohr^2. Then these groups are merged with their adjacent hydrophobic part, at $t = 1.15$ in MQPA and at $t = 2.35-2.5$ bohr^2 in the two other compounds (motifs II). In the sulfonyl-containing branches, the hydrophilic SO_2 moeities are rapidly merged with their adjacent NH and CH groups at $t = 0.95-1.15$ bohr^2 (motifs III). Then, at still lower resolution values, one observes a merging between these groups and the hydrophobic end groups, again in narrow ranges of t values: 1.75−1.8 bohr^2 as displayed in motifs IV, except for compound NAPAP wherein substructure III stays unchanged up to $t = 3.25$ bohr^2. Finally, piperidine rings (motifs V) are reduced to one single point at values of 1.2 bohr^2 for 4-TAPAP, 1.4 bohr^2 for NAPAP, and 1.65 bohr^2 for MQPA, a larger value for a substituted ring. That phenomenon was also observed in benzodiazepine-related molecules, wherein the free substituted aromatic ring of compound Ro5-4864 is fully merged at $t = 1.7$ bohr^2, a value that is 0.2 bohr^2 larger than for unsubstituted phenyl rings as in Diazepam, QUIZ, and TZQ (motifs VIII in Figure 15). The hydrophobic piperidine rings are merged with their connected hydrophobic region in the molecule at various values of t: 2.1 in MQPA, 2.4 in NAPAP, and 1.35 bohr^2 in 4-TAPAP (motifs VI). At a resolution value of $t = 1.6$ bohr^2, one gets peaks located on the main chemical groups involved in the binding with thrombin: motifs I (Figure 16) are involved in H-bonding with the biological receptor and/or water molecules; motifs IV and V are involved in interactions with the so-called aryl bonding site and S2 hydrophobic pocket of thrombin, respectively.[45]

Each branch of the thrombin inhibitor compounds is eventually reduced to one single point, except for compound NAPAP (Figure 11). Indeed, NAPAP includes a double ring, as MQPA does. However, the spatial arrangement of the double ring involves the presence of a second local maxima only in the ED in compound NAPAP wherein the double ring's longest axis is in the prolongation of the molecular branch. With the decrease of resolution (increase of t value), peaks are left only on piperidine N atoms, guanidinium or amidino central C, and SO_2 groups. As in benzodiazepines, carbonyl groups stay isolated from other substructures on a large range of resolution values. Those ED points are not merged with their adjacent atom before $t = 1.9$ bohr^2, a value larger than in benzodiazepine-related compounds wherein C═O are part of rings. Finally, molecules

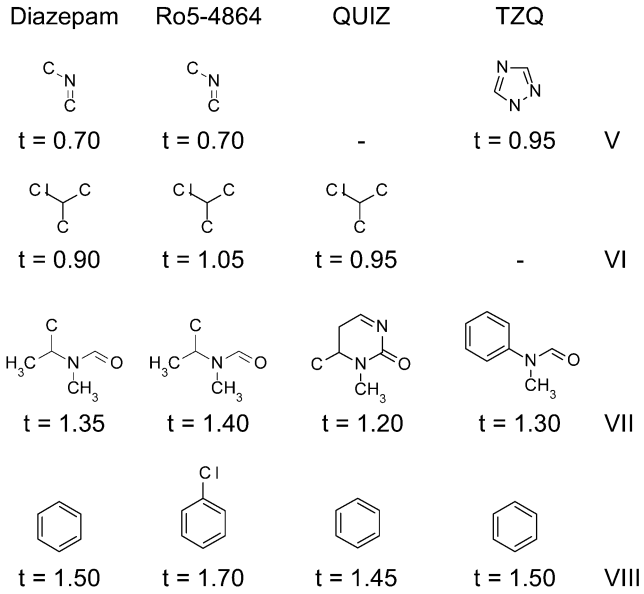

Figure 15. Selected motifs occurring in benzodiazepine-related compounds together with the value of t (bohr^2) at which they appear during the clustering/merging procedure.

are reduced to a single point at t values equal to 5.3 for MQPA, 11.15 for NAPAP, and 3.55 bohr^2 for 4-TAPAP.

In Table 3, it is shown that the sulfonyl groups bear the highest density points, around 0.32 e$^-$/bohr^3 and this remains valid through the whole set of investigated t values, i.e., from 1.1 to 2.5 bohr^2. In the lowest resolution images, i.e., in the ED generated using t greater or equal to 1.6 bohr^2, one observes that the alkyl part of molecule MQPA bears no peaks at all. The presence of peaks on the alkyl chain of compound MQPA is observed only at the highest resolution level, i.e., at $t = 1.1$ bohr^2. But even in that particular case, this chemical group leads to one CP with a rather low-density value of 0.1476 e$^-$/bohr^3.

In conclusion, $t = 1.6$ bohr^2 seems a rather good choice for partitioning molecules in chemically meaningful subgroups useful in terms of pharmacophore design. This was also the case for benzodiazepine-related molecules. We can mention here that a similar conclusion had already been reached in previous works that were carried out independently on benzodiazepines ($t = 1.58$ bohr^2)[23] and thrombin inhibitors ($t = 1.5$ bohr^2).[24] The comparison of the values reported in Tables 2 and 3 shows that, at low resolution, molecules of different sizes have a similar number of peaks with close density values. This thus favors the search for similarities among molecules of different structures.

Conclusions and Perspectives

The present work is part of a more global research work on pharmacophore determination of biopharmacological compounds using their electron density (ED) distribution topological properties. In previous works,[23−26,35] molecules were superposed using density values and distances between critical points of low-resolution ED maps. The interpretation of the matched chemical groups was achieved "by hand". We have shown in the present work that it is possible to computationally associate a chemical group to, at least, each local maxima of the ED functions. The partitioning of a molecular structure into fragments, which in some cases are similar to known chemical functions, is a way to attribute a physicochemical property to local maxima of the ED and at any resolution level. In this paper, we have presented a method which consists of a hierarchical clustering/merging

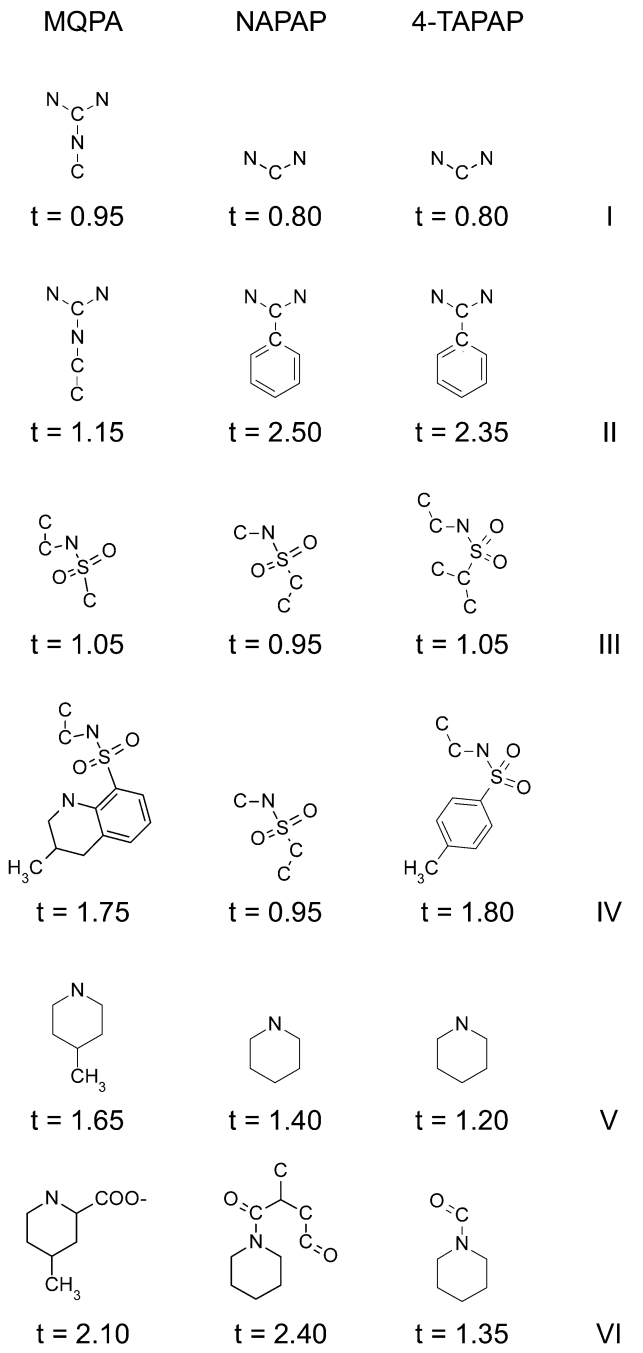

Figure 16. Selected motifs occurring in thrombin inhibitors together with the value of t (bohr^2) at which they appear during the clustering/merging procedure.

procedure wherein local maxima of a promolecular ED distribution established at a given resolution value t are related to individual atoms of the original molecular structure. For each of the molecules, the local maxima of the ED function calculated at various resolution values are determined by using a hierarchical clustering algorithm wherein peaks obtained at a given resolution are used as starting points for discovering peaks at the next lower resolution level. In the so-obtained hierarchical description, peaks at a given resolution level are linked to those obtained at the next lower resolution through gradient trajectories of the ED function. Results can be presented in terms of dendrograms wherein each node corresponds to a well-defined molecular substructure.

It was shown that $t = 1.6$ bohr^2 is a good value for molecular partitioning of benzodiazepine-related molecules and thrombin

TABLE 3: Values of the Electron Density (e^-/bohr^3) at Local Maxima Locations in Thrombin Inhibitor Electron Density Distributions Calculated with the PASA Approximation at Various Resolution Values *t* (bohr^2)

approximate peak location	$t = 1.1$ Figure 8	$t = 1.6$ Figure 9[a]	$t = 1.9$ Figure 10	$t = 2.5$ Figure 11
		MQPA		
carboxyl	0.1911	0.1455 (VII)	0.1288	
piperidine N	0.1831	0.1470 (V)	0.1343	0.1032
piperidine Me	0.1572	0.1291 (VI)		
carbonyl	0.1886	0.1481 (IV)	0.1341	
alkyl chain	0.1476			
guanidinium	0.1817	0.1415 (I)	0.1251	0.1163
sulfonyl	0.3246	0.2314 (II)	0.1992	0.1530
quinoline N	0.1744	0.1426 (III)		
quinoline C	0.1585			
quinoline C	0.1557			
quinoline C	0.1551			
$\rho_{max} - \rho_{min}$	0.1770	0.1023	0.0741	0.0498
		NAPAP		
piperidine N	0.1846	0.1470 (III)	0.1331	
piperidine C	0.1508			
piperidine C	0.1504			
piperidine C	0.1491			0.1139
carbonyl	0.1865	0.1474 (II)	0.1332	
phenyl	0.1648	0.1303 (V)	0.1205	0.1028
phenyl	0.1602			
amidino	0.1757	0.1374 (IV)	0.1223	
carbonyl	0.1852	0.1451 (II)	0.1297	
sulfonyl	0.3261	0.2316 (VI)	0.1985	0.1555
naphthyl	0.1661	0.1369 (VII)	0.1254	0.1081
naphthyl	0.1658	0.1369 (VIII)		
$\rho_{max} - \rho_{min}$	0.1770	0.1013	0.0780	0.0527
		4-TAPAP		
piperidine N	0.1878	0.1504 (III)	0.1364	0.1163
piperidine C	0.1567			
carbonyl	0.1836			
phenyl	0.1645	0.1328 (VI)	0.1205	0.1032
phenyl	0.1607	0.1309 (V)	0.1201	
amidino	0.1740	0.1362 (IV)	0.1214	
sulfonyl	0.3227	0.2283 (I)	0.1954	0.1530
toluene	0.1574	0.1273 (II)		
$\rho_{max} - \rho_{min}$	0.1660	0.1010	0.0753	0.0498

[a] Roman numbers correspond to the labels shown in Figures 3, 9, and 13.

inhibitors into subgroups that are characterized by conventional pharmacophore descriptors, i.e., H-bond acceptor/donor, steric, hydrophobic, ... The association of such physicochemical properties with molecular substructures is done through "good chemical sense" as there are currently no direct ways to relate low-resolution ED topological data to such properties.

The analysis of the obtained dendrograms should be automated in order to facilitate the application of the described approach to large sets of molecules and to larger systems. More precisely, the consideration of the information they contain in a molecular superposition algorithm should require to define a proper and fast way to represent the molecular substructures and how they are associated with the maxima of an ED map. Many references are available in the literature such as in refs 49 and 50. This could come in addition to or in replacement of the descriptors we usually considered for molecular superposition purposes,[23−26,35] i.e., the peak locations and their interdistances.

It should also be interesting to determine what would be the impact of using molecules composed of interacting atoms rather than promolecular systems. A solution to this question may reside in working with nonorthogonal wavelet filtering of ED images computed by quantum mechanical approaches. Let us, however, mention that in a recent work published by Bultinck et al.,[51] it is shown that the ASA approximation level is extremely good to model molecular ED distributions, especially in similarity applications. We also carried out preliminary calculations on thrombin inhibitors which showed that the number and location of critical points do not differ significantly between promolecular and nonpromolecular ED; only the density values are different, especially on heavy atoms.

So far, our algorithm can be applied to the analysis of properties that can be analytically represented, e.g., ASA ED and interaction potentials.[52] However, it could be possible to extend the present approach to grid representations of a property. In such a way, any property could a priori be analyzed in terms of molecular substructures. Results may likely vary depending upon the property that is analyzed.

Further developments of our algorithm are planned toward the location of other kinds of critical points in ED maps, particularly saddle points (bond critical points), which are necessary to properly connect the local maxima or peaks. Moreover, although low CPU time was not our primary goal, an improvement of our algorithm in that direction is also required.

Acknowledgment. We thank L. Piela for fruitful discussions on analytical smoothing, R. Carbó-Dorca and A. Savin for their interest in our original approach, and the FUNDP for the use of the Namur Scientific Computing Facility (SCF) Center. L.D. thanks the FRIA for his Ph.D. Fellowship.

References and Notes

(1) Koritsanszky, T. S.; Coppens, Ph. *Chem. Rev.* **2001**, *101*, 1583.
(2) Connolly, M. L. *NetSci's Science Center: Computational Chemistry*; 1996, http://www.netsci.org/Science/Compchem/feature14.html.
(3) Good, A. C.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112.
(4) Grant, J. A.; Pickup, B. T. *J. Phys. Chem.* **1995**, *99*, 3503.
(5) Grant, J. A.; Pickup, B. T. In *Computer Simulation of Biomolecular Systems. Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer/Escom: Dordrecht, 1997; Vol. 3, p 150.
(6) Maggiora, G. M.; Rohrer, D. C.; Mestres, J. *J. Mol. Graphics Model* **2001**, *19*, 168.
(7) Duncan, B. S.; Olson, A. J. *Biopolymers* **1993**, *33*, 231.
(8) Tsirelson, V. G.; Avilov, A. S.; Abramov, Y. A.; Belokoneva, E. L.; Kitaneh, R.; Feil, D. *Acta Crystallogr. B* **1998**, *54*, 8.
(9) Tsirelson, V.; Abramov, Y.; Zavodnik, V.; Stash, A.; Belokoneva, E.; Stahn, J.; Pietsch, U.; Feil, D. *Struct. Chem.* **1998**, *9*, 249.
(10) Gironés, X.; Amat, L.; Carbó-Dorca, R. *J. Mol. Graphics Model* **1998**, *16*, 190.
(11) Gironés, X.; Carbó-Dorca, R.; Mezey, P. G. *J. Mol. Graphics Model* **2001**, *19*, 343.
(12) Mitchell, A. S.; Spackman, M. A. *J. Comput. Chem.* **2000**, *21*, 933.
(13) Downs, R. T.; Gibbs, G. V.; Boisen, M. B., Jr.; Rosso, K. M. *Phys. Chem. Minerals* **2002**, *29*, 369.
(14) Irle, S.; Lin, H. L.; Niu, J. E.; Schwarz, W. H. E. *Ber. Bunsen-Ges. Phys. Chem.* **1992**, *96*, 1545.
(15) Amat, L.; Carbó-Dorca, R. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1188.
(16) Saunders, V. R.; Dovesi, R.; Roetti, C.; Causa, M.; Harrison, N. M.; Orlando, R.; Zicovich-Wilson, C. M. *CRYSTAL98*; University of Turin, Torino, Italy, and CLRC Daresbury Laboratory, Daresbury, UK.
(17) Gatti, C. *TOPOND96*; CNR-CSRSRC, Milano, Italy.
(18) Bader, R. W. *Atoms in Molecules—A Quantum Theory*; Oxford University Press: Oxford, 1990.
(19) Spackman, M. A.; Maslen, E. N. *J. Phys. Chem.* **1986**, *90*, 2020.
(20) Maksic, Z. B.; Kovacek, D.; Vidic, B. *Chem. Phys. Lett.* **1986**, *129*, 619.
(21) Gironés, X.; Amat, L.; Carbó-Dorca, R. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 847.
(22) Kostrowicki, J.; Piela, L.; Cherayil, B. J.; Scheraga, H. A. *J. Phys. Chem.* **1991**, *95*, 4113.
(23) Leherte, L.; Meurice, N.; Vercauteren, D. P. In *Mathematics and Computers in Modern Science. Acoustics and Music, Biology and Chemistry, Business and Economics*; Mastorakis, N., Ed.; World Scientific Engineering Society: Athens, 2000; p 158.

(24) Leherte, L. *J. Math. Chem*. **2001**, *29*, 47.
(25) Meurice, N.; Leherte, L.; Vercauteren, D. P.; Bourguignon, J.-J.; Wermuth, C. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H.; Testa, B.; Folkers, G., Eds.; Verlag: Basel, 1997; p 497.
(26) Meurice, N.; Leherte, L.; Vercauteren, D. P. In *SAR and QSAR in Environmental Research*; Devillers, J., Ed.; OPA: Amsterdam, 1998; p 195.
(27) Johnson, C. K. *ORCRIT. The Oak Ridge Critical Point Network Program*, Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, TN, 1977.
(28) Zimpel, Z.; Mezey, P. G. *Int. J. Quantum Chem*. **1996**, *59*, 379.
(29) Takahashi, Y.; Sukekawa, M.; Sasaki, S.-I. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 639.
(30) Rarey, M.; Dixon, J. S. *J. Computer-Aided Mol. Des*. **1998**, *12*, 471.
(31) Dury, L.; Leherte, L.; Vercauteren, D. P. In *Principles of Data Mining and Knowledge Discovery*; Zighed, D. A., Komorowski, J., Zytkow, J., Eds.; Springer-Verlag: Berlin, 2000; p 388.
(32) Dury, L.; Latour, T.; Leherte, L.; Barberis, F.; Vercauteren, D. P. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1437.
(33) Gillet, V. J.; Downs, G. M.; Ling, A.; Lynch, M. F.; Venkataram, P.; Wood, J. V.; Dethlefsen, W. *J. Chem. Inf. Comput. Sci*. **1987**, *27*, 126.
(34) Gillet, V. J.; Downs, G. M.; Holliday, J. D.; Lynch, M. F.; Dethlefsen, W. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 260.
(35) Leherte, L.; Meurice, N.; Vercauteren, D. P. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 816.
(36) Amat, L.; Carbó-Dorca, R. *J. Comput. Chem*. **1997**, *18*, 2023.
(37) Coefficients and exponents can be seen and downloaded from the Web site: http://iqc.udg.es/cat/similarity/ASA/funcset.html.
(38) Leung, Y.; Zhang, J.-S.; Xu, Z.-B. *IEEE Trans. Pattern Anal. Machine Intelligence* **2000**, *22*, 1396.
(39) Gilbert, D. G. *Phylodendron, for Drawing Phylogenetic Trees*, Version 0.8d; Indiana University, 1996. Software at http://iubio.bio.indiana.edu/soft/molbio/java/apps/trees/, Web form at http://iubio.bio.indiana.edu/treeapp/treeprint-form.html.
(40) Dury, L. *DENDRO*, Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium, 2002.
(41) Web site: http://www.borland.com.
(42) Allen, F. H.; Kennard, O. *Chem. Design Automation News* **1993**, *8*, 31.
(43) Meurice, N.; Norberg, B.; Durant, F.; Didier, B.; Bourguignon, J.-J.; Vercauteren, D. P. Unpublished results.
(44) Nissink, J. W. M.; Verdonk, M. L.; Kroon, J.; Mietzner, T.; Klebe, G. *J. Comput. Chem*. **1997**, *18*, 638.
(45) Brandstetter, H.; Turk, D.; Hoeffken, H. W.; Grosse, D.; Stürzebecher, J.; Martin, P. D.; Edwards, B. F. P.; Bode, W. *J. Mol. Biol.* **1992**, *226*, 1085.
(46) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucl. Acids Res.* **2000**, *28*, 235. Web site: http://www.rcsb.org/pdb/.
(47) *Insight II*, Version 4.0.0, Molecular Simulation Inc., San Diego CA, 1998.
(48) *IBM Visualization Data Explorer*, IBM Corporation, 1996.
(49) Downs, G. M.; Willett, P. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D. B., Eds.; VCH Publishers: New York, 1996; Vol. 7, p 1.
(50) Arteca, G. A. In *Reviews in Computational Chemistry*; Lipkowitz, K., Boyd, D. B., Eds.; VCH Publishers: New York, 1996; Vol. 9, p 191.
(51) Bultinck, P.; Carbó-Dorca, R.; Van Alsenoy, Ch. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1208.
(52) (52) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem*. **1989**, *93*, 3339.