Journal of Mathematical Chemistry Vol. 29, No. 1, 2001

Application of multiresolution analyses to electron density maps of small molecules: Critical point representations for molecular superposition

Laurence Leherte

Facultés Universitaires Notre-Dame de la Paix, Laboratoire de Physico-Chimie Informatique, Rue de Bruxelles 61, B-5000 Namur, Belgium
E-mail: laurence.leherte@fundp.ac.be

Received 11 December 2000

Three different methods are applied to generate low resolution molecular electron density (ED) distribution functions: a crystallography-based formalism, an analytical approach which allows the calculation of a promolecular ED distribution in terms of a weighted summation over atomic ED distributions, and a wavelet-based multiresolution analysis approach. Critical point graph representations of the molecular ED distributions are then generated by locating points where the gradient of the density is equal to zero, and further considered for pairwise molecular superpositions of thrombin inhibitors using a Monte Carlo/Simulated Annealing technique.

KEY WORDS: multiresolution, electron density, *ab initio* and ASA density functions, critical points, molecular similarity, wavelets

1. Introduction

Computer resources have now become sufficiently powerful to enable simulations of large systems (10,000–50,000 atoms) with realistic mathematical models at a classical level. Therefore, biological macromolecules and supramolecular complexes, for example, can be studied in atomic detail using simulation techniques. However, when systems include large numbers of degrees of freedom and/or environment considerations, a systematic search of energy minima is still untractable. Large biomolecules, for instance, hence require simplified models [1]. Atomic descriptions of large scale conformational changes, for example, those inherent in protein folding, remain computationally inaccessible due to the large number of conformational states, the multiple minima problem, and the multiple time scales: from 1 fs for a chemical bond vibration to about 1 s for the folding process. Thus, constraining degrees of freedom and/or reducing the level of resolution are helpful ways to circumvent some problems.

0259-9791/01/0100-0047$19.50/0

In proteins, amino acid residues can be depicted using a lower level representation, i.e., two or three pseudo-atoms rather than by an all-atom representation. The advantage of using non-atomic representations is, however, not limited to the increase of the speed of computations. Simplified representations of protein geometry have also been used by many groups to reduce sensitivity to small perturbations in conformation, e.g., when docking a ligand versus a receptor [2,3]. Janin et al. [2] replaced amino acid residues with spheres of varying size and performed docking to maximize the buried surface area. Simulating proteins at low resolution is also a way to overcome structural inaccuracies issued from NMR data or from an approximate model [4]. In their paper, Vakser et al. [4] digitized a protein image onto a three-dimensional grid. Any grid point outside the molecular volume determined was set equal to 0, otherwise it was set equal to a numerical value corresponding to the protein surface or to the protein core [5]. The structural elements smaller than the grid interval were thus eliminated from the initial protein structure. In molecular graphics, fast comparisons of protein structures at a predetermined level of detail can be achieved using wavelet analysis [6]. In this work, Carson applied B-spline filters to the modeling of protein backbones, folds, as well as surfaces (represented using Nurbs, i.e., non-uniform rational B-spline surfaces). He also suggested an implication in structure-based drug design. In that sense, Ritchie et al. [7] later suggested that surface representations which contain too high resolution details may not be particularly convenient to search for regions of similarity or complementarity between two molecules. The authors therefore proposed the use of low resolution real spherical harmonics to represent and compare macromolecular surface shapes. Rotations of a molecular surface can thus be simulated by rotating only the harmonic expansion coefficients. Particularly, it was determined that, contrarily to small molecules for which low order multipoles ($L = 2$ or 3) can be used, higher order harmonic expansions (between $L = 5$ and $L = 9$) are needed for a good surface matching involving globular molecules. Smooth molecular shapes can also be obtained by convolving a molecular surface with a three-dimensional Gaussian function of appropriate variance. Duncan and Olson [8] proposed applications of such a method to automated molecular docking procedures.

In DNA, structural elements such as chains and bases can be modeled as rigid segments connected through energetic terms [9–11]. Particularly, Butzlaff et al. [9,10] described an approach which reduces the number of free variables by assembling certain groups of atoms into configurational structures with less degrees of freedom. Corresponding potential energy functions were constructed with respect to these new variables using methods from the theory of splines and radial basis functions. DNA is, under such a description, represented by two kinds of rigid substructures: the bases and the phosphate groups, and one rotational group: the ribose subunit.

The idea of grouping atoms together was also achieved in a completely different approach. For example, Takahashi et al. [12] described organic structures in terms of graphs wherein each node represents a functional group and the edge between any two nodes is weighted by the topological path length (the number of bonds). In such a way, searching for common or structural features between those reduced-graph representa-

tions is similar to finding a common pharmacophore model shared by biologically active molecules.

Low resolution representations are also extremely useful in the refinement or interpretation of images generated by experimental approaches such as electron microscopy or X-ray diffraction. For example, a current challenge in structural biology is to establish the structure of complex systems. Such structure elucidations require high resolution structural determination methods to generate atomic models of the individual components. Then, complexes created from the well-resolved individual components are imaged at lower resolution to validate the interpretation of the experimental low resolution global structure [13]. Another potential application to low resolution experimental data is the restoration of noisy images. The principle of multiresolution analysis based on wavelet theory (WMRA) was shown to be adapted to such problems [14]. Performing a WMRA consists in the decomposition of an image into its details corresponding to a range of resolution levels. A particular application in the field of chemistry can be found in [15] wherein a WMRA was applied to restore two-dimensional X-ray topographs damaged by Gaussian noise.

The purpose of the present paper consists in assessing low resolution representations of molecular structures in their ability to carry enough information to determine similarities between various molecular structures described in terms of their electron density (ED) distribution. In the proposed methodology, molecular structures are represented at various levels of resolution in terms of critical point (CP) graphs obtained using a topological analysis procedure of their ED distribution functions.

In the next section, the theoretical background of CP analysis is presented. It is followed, in section 3, by the mathematical background of the three methods selected to generate low resolution ED maps. These three methods are based on a crystallography formalism, on a WMRA technique, and on a smoothing procedure of analytical ED functions, respectively. Applications to the comparison of three thrombin inhibitors and results are then reported in section 4.

2. Theoretical background

In this section, the concept of topological analysis of electron density maps is presented through the project “Molecular Scene Analysis” which was initiated in the field of protein crystallography. Then, the three approaches selected to generate low resolution molecular electron density maps are detailed at a mathematical level.

2.1. Molecular scene analysis

The concept of “scene analysis” is normally used in the context of machine vision to refer to the set of processes associated with the classification and understanding of complex images. Such analyses rely on the availability of *a priori* information, both in the form of structural templates and in the form of rules or heuristics, to locate and

identify features in a scene. By analogy, Glasgow et al. [16] use the phrase "molecular scene analysis" (MSA) to refer to the processes associated with the interpretation of molecular structures and molecular interactions. A goal of the MSA project is to facilitate the image reconstruction processes for protein electron density maps [17–20]. This requires a simplified representation of a protein structure, one that preserves relevant shape, connectivity and distance information. One of the particular MSA applications consists in the identification of structural motifs through a similarity search procedure between templates as established from a statistical analysis of PDB structures on one side, and observed critical point graph segments on the other [18].

2.2. Critical point analysis

Electron density (ED) and molecular electrostatic potential (MEP) are seen as properties of prime importance in the description of molecular interactions. Several molecular similarity evaluation approaches are thus based on refined molecular field matching techniques. For example, similarity quantifiers applicable to steric volume and electrostatic fields have been implemented in the program MIMIC [21]. This program calculates the so-called Carbó index [22], one of the most widely used molecular similarity measure between molecules. This index was originally dedicated to the comparison of molecular ED distributions.

Orientation-independent descriptors possess the obvious advantage of avoiding molecular translation and reorientation operations when matching two or more molecular structures. In this context, reference studies were achieved by Mezey [23] through his Shape Group method which allows the characterization of molecular surfaces (van der Waals envelopes, iso-density contours, MEP iso-contours, etc.) using curvature information. A molecular surface is partitioned into domains whose specific arrangement is described in terms of homology groups characterized by their ranks, or Betti numbers [24,25]. Bader [26] established a topological analysis method of three-dimensional ED distributions in terms of the number and kind of their critical points (CP), i.e., the points where the gradient of the density is equal to zero. Each CP can be identified by its corresponding 3×3 Hessian matrix $H(\vec{r})$ which is built on the local second derivatives of the ED function. In Bader's theory of atoms in molecules (AIM), each atom is associated with an attractor and its basin bounded by a zero-flux surface over which many atomic properties can be integrated. This yields a unique partitioning of a total system into a set of bounded spatial regions. The CPs are linked through a gradient vector field analysis to generate a graph whose vertices and edges are CPs and gradient trajectories, respectively. Popelier [27] has later extended this approach to include the concept of molecular similarity. The author proposed to use the properties of bond critical points (BCP), i.e., the density value, Laplacian, and ellipticity, extracted from quantum mechanical *ab initio* wavefunctions. Similarly to Bader's approach, Johnson [28] developed a CP analysis method, based on Morse theory, for the location, identification, and connection of CP trees in experimental protein ED maps. His method was aimed at the automated interpretation of X-ray diffraction data for protein structures.

MEP functions have also been the subject of topological/topographical analysis studies [29,30]. For example, Willett and coworkers [31,32] developed an approach where grid points with MEP values ranging below or beyond given cut-off values are merged into single points which are further connected depending upon their closeness. Such graph representations facilitate the alignment of MEP fields using genetic algorithms. Recently, Leboeuf et al. [33] implemented a new algorithm in their program ALLCHEM for the topological analysis of MEP functions determined using Density Functional theory.

Details regarding the CP analysis approach applied in the present work have been presented elsewhere [34–36]. However, a brief overview of the approach is described in this paper as a reminder of the methodology.

In the proposed topological approach to low resolution ED molecular distributions, a molecule is segmented into its meaningful parts through the location and identification of the points in the ED map where the gradients vanish. The second derivatives are used to determine the nature of the CP. For each CP, the corresponding Hessian matrix is constructed:

$$\mathbf{H}(\vec{r}) = \begin{pmatrix} \dfrac{\partial^2\rho}{\partial x^2} & \dfrac{\partial^2\rho}{\partial x\partial y} & \dfrac{\partial^2\rho}{\partial x\partial z} \\ \dfrac{\partial^2\rho}{\partial y\partial x} & \dfrac{\partial^2\rho}{\partial y^2} & \dfrac{\partial^2\rho}{\partial y\partial z} \\ \dfrac{\partial^2\rho}{\partial z\partial x} & \dfrac{\partial^2\rho}{\partial z\partial y} & \dfrac{\partial^2\rho}{\partial z^2} \end{pmatrix}. \tag{1}$$

This matrix is then diagonalized:

$$\mathbf{H}'(\vec{r}) = \begin{pmatrix} \dfrac{\partial^2\rho}{\partial x'^2} & 0 & 0 \\ 0 & \dfrac{\partial^2\rho}{\partial y'^2} & 0 \\ 0 & 0 & \dfrac{\partial^2\rho}{\partial z'^2} \end{pmatrix}. \tag{2}$$

The three nonzero diagonal elements of array $\mathbf{H}'(\vec{r})$, the eigenvalues, are used to determine the type of CPs of the ED map. Four possible cases are considered depending upon the number of negative eigenvalues n_E. When $n_E = 3$, the CP corresponds to a local maximum or *peak*. A point where $n_E = 2$ is a saddle point or *pass*. $n_E = 1$ corresponds to a saddle point or *pale*, while $n_E = 0$ characterizes a *pit*. This method is implemented in the program ORCRIT [28] which was initially aimed at the automated interpretation of X-ray diffraction data for protein structures.

3. Multiresolution analysis

In previous works, the program XTAL [37] was the main computation tool that was used to generate low resolution ED distributions of various molecular systems [34–36,38–40]. Hypothetical crystal parameters had to be assumed to generate such promolecular charge distributions. In the present work, two different approaches, which do not require neither the knowledge nor the assumption of any crystal parameter, are also applied. They are based on the calculation of $\rho(\vec{r})$ using quantum mechanics formalisms. In the first approach, full ED distributions were generated using the program Gaussian94 [41] and were further smoothed using a wavelet-based multiresolution analysis. In the second approach, a promolecular ED distribution ρ_{M} was calculated as a weighted summation over atomic ED distributions ρ_a which are described in terms of series of squared Gaussian functions [42]. Smoothed ED versions were generated analytically as solutions of the diffusion equation $\nabla^2\rho = \partial\rho/\partial t$ [43].

3.1. The concept of resolution in crystallography

The intensity of X-rays diffracted by a crystalline structure is proportional to the modulus of their corresponding structure factor $F(\vec{h})$:

$$F(\vec{h}) = \sum_{j=1}^{n_{\mathrm{at}}} f_j \mathrm{e}^{-B_j(\sin\theta/\lambda)^2} \mathrm{e}^{-2\pi \mathrm{i}\vec{h}\cdot\vec{r}_j}, \tag{3}$$

where f_j and B_j are the atomic scattering factor and the thermal agitation factor of atom j, respectively, 2θ is the angle between the diffracted and the primary beams of wavelength λ, and $\vec{h}$ is a reciprocal space vector. Within the crystallographic approach, the electron density (ED) distribution function $\rho(\vec{r})$ is calculated as the Fourier transform of $F(\vec{h})$:

$$\rho(\vec{r}) = \widehat{F}(\vec{r}) = \frac{1}{V_{\text{unit-cell}}} \sum_{\vec{h}=-\infty}^{+\infty} F(\vec{h}) \mathrm{e}^{-2\pi \mathrm{i}\vec{h}\cdot\vec{r}}. \tag{4}$$

In practice, the number of known structure factors is not infinite and varies with resolution.

In crystallography, the resolution factor $d_{\min}$ is a well-known concept which is theoretically defined using Bragg's law:

$$\left(\frac{\sin\theta}{\lambda}\right)_{\max} = \frac{1}{2d_{\min}}, \tag{5}$$

where $d_{\min}$ depends on different factors such as the quality of the crystal, the chemical composition, the radiation used, the temperature of the experiment, etc.

3.2. Wavelet-based multiresolution analysis

To our knowledge, only few applications of wavelet-based multiresolution analysis (WMRA) to chemical sciences, particularly structural chemistry, exist. In practice, wavelet theory is most commonly applied to the treatment of signal processing in analytical chemistry [44] and for the representation of wavefunctions and orbitals in electronic structure calculations [45–47].

3.2.1. Wavelet transforms

A wavelet transform (WT) is a localized transform in both space (time) and frequency which uses integration kernels called wavelets. A basis set of wavelet functions $\{\Psi_{ab}(x)\}$ is built on translated and dilated versions of a so-called *mother wavelet* $\Psi(x)$:

$$\Psi_{ab}(x) = \frac{1}{\sqrt{|a|}}\Psi\left(\frac{x-b}{a}\right) \quad \text{with } a \in \mathbb{R}_0,\ b \in \mathbb{R}, \tag{6}$$

where a is the scaling parameter which allows to capture changes in frequency, and b is the shift along the x axis applied to analyze space (time)-dependent variations of a signal. A detailed mathematical introduction to the continuous wavelet transform can be found in [48,49].

The projection $\mathcal{W}f$ of a square integrable signal f, i.e., $\int_{-\infty}^{+\infty} |f(x)|^2\,\mathrm{d}x \neq \infty$, onto this basis according to

$$\mathcal{W}f(a,b) = \int_{-\infty}^{+\infty} f(x)\Psi^*_{ab}(x)\,\mathrm{d}x \tag{7}$$

$$= \langle f, \Psi_{ab}\rangle \tag{8}$$

is the result from the wavelet transform. * denotes the complex conjugate.

If the wavelet function Ψ satisfies the *admissibility condition*, i.e.:

$$C_\Psi = \int_{-\infty}^{+\infty} \frac{|\widehat{\Psi}(\omega)|^2}{\omega}\,\mathrm{d}\omega < +\infty, \tag{9}$$

where $\widehat{\Psi}(\omega)$ is the Fourier transform of the function $\Psi(x)$, then the continuous wavelet transform $\mathcal{W}f(a,b)$ is invertible such as $f(x)$ can be reconstructed using

$$f(x) = \frac{1}{C_\Psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathcal{W}f(a,b)\Psi_{ab}(x)\frac{\mathrm{d}a\,\mathrm{d}b}{a^2}. \tag{10}$$

From the admissibility condition, it follows that $\widehat{\Psi}(0)$ must be equal to zero, i.e., Ψ has a mean equal to zero. Ψ is thus an oscillating function, and this, together with the decay property, justifies the name *wavelet*.

In addition, Ψ is often required to have a certain number p of *vanishing moments*:

$$\int_{-\infty}^{+\infty} x^n\Psi(x)\,\mathrm{d}x = 0, \quad n = 0, 1, \ldots, p-1. \tag{11}$$

This property improves the efficiency of Ψ at detecting discontinuities in the $(p-1)$th derivative of a signal.

For numerical purposes, the continuous WT can be discretized, by restricting the parameters a and b to the points of a dyadic lattice. Thus, if a and b/a are set equal to 2^{-j} and 1, respectively, equations (6) and (10) are written as

$$\Psi_{jk} = 2^{j/2}\Psi\left(2^j x - k\right), \quad \text{with } j, k \in \mathbb{Z}, \tag{12}$$

$$f(x) = \sum_{j,k} \langle f, \Psi_{jk}\rangle \Psi_{jk}(x) \tag{13}$$

using the bracket notation introduced in equation (8).

3.2.2. Multiresolution analysis

One of the successes of the WT was the discovery that it is possible to construct functions Ψ for which $\{\Psi_{jk},\ j, k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$ [50]. The construction of the so-called Daubechies' wavelet functions is derived from a multiresolution analysis which is a mathematical construction used to express an arbitrary function $f \in L^2(\mathbb{R})$ at various levels of detail. Function $f(x)$ is developed as in equation (13) where $\langle f, \Psi_{jk}\rangle$, also written d_{jk}, are called the wavelet coefficients.

Mallat [51] proposed a *multiresolution theory* where a fast and nonredundant wavelet transform was demonstrated. The fast wavelet transform is performed by a set of bandpass filters constructed using a combination of low- and high-pass filters. The low-pass filter corresponds to the so-called *scaling function* Φ, and the high-pass filter corresponds to the wavelet function Ψ. In practice, the wavelet expansion (like the Fourier expansion) is truncated at some scale J:

$$f(x) = \sum_k c_{Jk}\Phi_{Jk}(x) + \sum_{j=J}^{J_0-1}\sum_k d_{jk}\Psi_{jk}(x), \tag{14}$$

where it is chosen here to set the resolution of the original signal J_0 equal to zero. Thus, lower resolution signals are characterized by negative values for J. Relation (14) can be understood as follows. The first sum is a coarse representation of f, where f is replaced by a linear combination of a finite number of translations of the scaling functions $\Phi_{J0}(x)$. The remaining terms are refinements (details) determined at each scale j by translations of the wavelet $\Psi_{j0}(x)$ that are added to obtain a successively more detailed approximation of f.

Coefficients c_{Jk} are projections of the function $f \in L^2(\mathbb{R})$ onto a space built on the basis set $\{\Phi_{Jk}(x)\}$:

$$c_{Jk} = \langle f, \Phi_{Jk}\rangle. \tag{15}$$

Let us consider a function Φ such as the family $\{\Phi(x-k), k \in \mathbb{Z}\}$ is an orthogonal basis of a vector space denoted V_0. It is convenient to define the approximation space V_j, i.e., the space of all combinations of Φ's at level j. The space of all combinations of Ψ's at level j is denoted W_j. A wavelet-based multiresolution analysis (WMRA) will thus

be defined as a sequence of closed subspaces V_j of $L^2(\mathbb{R})$, $j \in \mathbb{Z}$, with the following properties [50]:

$$V_j \subset V_{j+1} \tag{16}$$

such that

$$\bigcap_{j=-\infty}^{+\infty} V_j = \emptyset \quad \text{and} \quad \bigcup_{j=-\infty}^{+\infty} V_j = L^2(\mathbb{R}), \quad \text{with } j \in \mathbb{Z} \tag{17}$$

and

$$\forall f \in L^2(\mathbb{R}), \quad f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}. \tag{18}$$

Since the set $\{\Phi(x-k)\}$ is an orthonormal basis for V_0, it follows by repeated applications of axiom (18) and the property of shift invariance

$$\forall f \in L^2(\mathbb{R}),\ \forall k \in \mathbb{Z}, \quad f(x) \in V_0 \Leftrightarrow f(x-k) \in V_0 \tag{19}$$

that $\{\Phi(2^j x - k)\}$ is an orthonormal basis for V_j.

The detail spaces. Given the nested subspaces V_j, W_j is defined to be the orthogonal complement of V_j in V_{j-1}, i.e., $V_j \perp W_j$ and

$$V_j = V_{j-1} \oplus W_{j-1} \tag{20}$$

which is the key statement of the WMRA [49].

In terms of approximation and detail spaces, relation (14) becomes

$$V_{J_0} = V_J \oplus W_J \oplus W_{J+1} \oplus W_{J+2} \oplus \cdots \oplus W_{J_0-1} = V_J \oplus \left(\bigoplus_{j=J}^{J_0-1} W_j \right). \tag{21}$$

The refinement equation. Let us consider V_0 and V_1, such that $V_0 \subset V_1$. These two subspaces are generated by integer translates of $\Phi(x)$ and $\Phi(2x)$, respectively. The subspace relation implies that $\Phi(x)$, $\Phi_{00}(x)$, must be generated by the finer scale functions $\Phi(2x-k)$:

$$\Phi(x) = \sqrt{2} \sum_k h_k \Phi(2x-k), \tag{22}$$

where the numbers h_k are called the filter coefficients. Equation (22), known as the *dilation* or *refinement equation*, is the principal relation determining multiresolution.

Since $\Psi \in W_0$ and $W_0 \subset V_1$, it is also known that the wavelet satisfies an equation that is similar to relation (22):

$$\Psi(x) = \sqrt{2} \sum_k g_k \Phi(2x-k), \tag{23}$$

where the wavelet coefficients g_k are obtained directly from the filter coefficients h_k [50]:

$$g_k = (-1)^k h_{k_{\max}-k}, \quad \text{where } k = 0, 1, \ldots, k_{\max}. \tag{24}$$

Table 1
Scaling and wavelet coefficients associated with orthogonal Daubechies wavelets with one, two, and ten vanishing moments [50,52].

	k	h_k	g_k	k	h_k	g_k
D_1	0	0.707107	0.707107			
	1	0.707107	−0.707107			
D_2	0	0.482963	−0.129410			
	1	0.836516	−0.224144			
	2	0.224144	0.836516			
	3	−0.129410	−0.482963			
D_{10}	0	0.026670	−0.000013	10	−0.029458	−0.071394
	1	0.188177	−0.000094	11	0.033213	−0.093057
	2	0.527201	−0.000116	12	0.003607	0.127369
	3	0.688459	0.000686	13	−0.010733	0.195946
	4	0.281172	0.001992	14	0.001395	−0.249846
	5	−0.249846	−0.001395	15	0.001992	−0.281172
	6	−0.195946	−0.010733	16	−0.000686	0.688459
	7	0.127369	−0.003607	17	−0.000116	−0.527201
	8	0.093057	0.033213	18	0.000094	0.188177
	9	−0.071394	0.029458	19	−0.000013	−0.026670

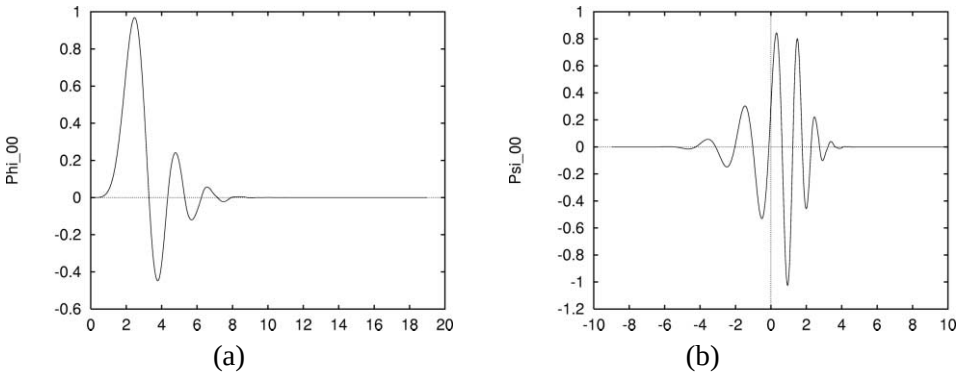


Figure 1. Scaling (a) and wavelet (b) Daubechies' functions of order 10, with $j = k = 0$.

This relationship was proposed by Daubechies [50] to calculate h_k's and g_k's associated with scaling and wavelet functions with a given number of vanishing moments. As examples, Daubechies' wavelets with one, two, and ten vanishing moments are presented in table 1. The D_{10} filter and wavelet functions, generated using the program Wavelib [52], are also displayed in figure 1. D_1, also known as the Haar filter, is the only real-valued wavelet that is compactly supported, symmetric, and orthogonal. For this wavelet system, the coefficients h_k and g_k are equal to $\sqrt{2}(1/2, 1/2)$ and $\sqrt{2}(1/2, -1/2)$, respectively. The associated scaling function Φ actually corresponds to the box function which is characterized by only one vanishing moment.

Fast wavelet transform. The orthogonality of scaling functions and wavelets together with the dyadic coupling between wavelet-based multiresolution analysis (WMRA) spaces lead to a relation between scaling function coefficients and wavelet coefficients on different scales. This yields a fast and accurate algorithm due to Mallat [51] named the *pyramid algorithm* or the *Fast Wavelet Transform* (FWT). The goal is to derive a mapping between the sequence $\{c_{jl}\}$ and the sequences $\{c_{j-1,l}\}$ and $\{d_{j-1,l}\}$. The key to this mapping is the dilation equation (22) and the wavelet equation (23). As demonstrated in [53,54], the following identities which define a partial wavelet transform can be derived:

$$c_{j-1,l} = \sum_k h_k c_{j,2l+k}, \tag{25}$$

$$d_{j-1,l} = \sum_k g_k c_{j,2l+k}. \tag{26}$$

In order to further decompose the space V_{j-1}, one applies the mapping to the sequence $\{c_{j-1,l}\}$ in order to obtain the new sequences $\{c_{j-2,l}\}$ and $\{d_{j-2,l}\}$. This procedure is repeated until the full FWT is achieved. Equations (25) and (26) show that once the elements $\{d_{j-1,l}\}$ are computed, they are not modified in the subsequent steps. Equation (26) also shows that the calculation of coefficients $\{c_{j-1,l}\}$ from coefficients $\{c_{j,2l+k}\}$ implies a downsampling, i.e., the number of coefficients is reduced by 2. On the contrary, reconstruction implies an upsampling procedure, i.e., the number of data point is multiplied by 2 at each level of resolution. Figure 2 illustrates an application of the FWT algorithm to a finite one-dimensional signal composed of 8 data points. In this example, the signal can be extended beyond and below the data points by periodicity, or by padding with zeroes. The number of h_k and g_k coefficients is equal to 4, as in a FWT using D_2. A first application of the filter and wavelet transformation generates four coefficients which compose the first averaged version of the initial signal, and four detail coefficients. A second application of the filter coefficients to the averaged signal yields 2 points for the filtered version, and 2 detail values.

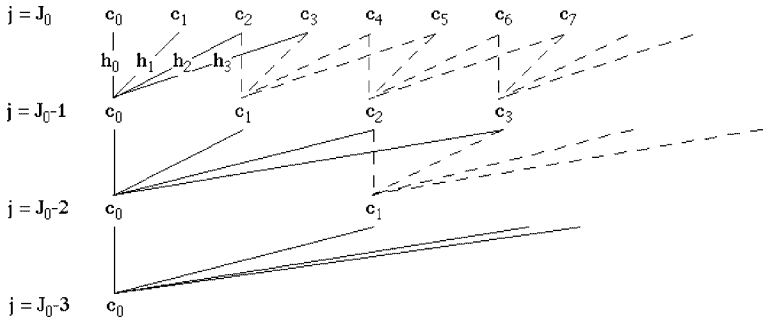


Figure 2. Schematic representation of the application of the FWT cascade algorithm to the transformation of a 8-point signal over three resolution levels. The averaged version is obtained using h_k coefficients. Details are obtained using the corresponding wavelet coefficients g_k.

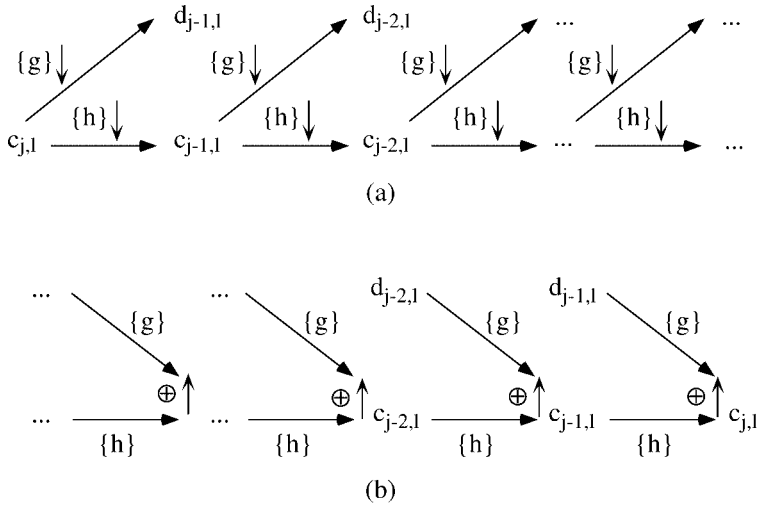


Figure 3. The Mallat multiresolution algorithm. The down- and up-arrows indicate decimation and up-sampling, respectively: (a) decomposition scheme, (b) reconstruction scheme.

The inverse mapping can be derived to obtain

$$c_{jl} = \sum_k h_{l-2k} c_{j-1,k} + \sum_k g_{l-2k} d_{j-1,k}. \tag{27}$$

The *inverse fast wavelet transform* is obtained by repeated application of this equation for $j = J + 1, J + 2, \ldots, J_0$. A general scheme for the pyramid decomposition and reconstruction algorithms is presented in figure 3.

Multidimensional cases and smoothing. A simple way to obtain wavelets in dimensions higher than 1 is to carry out one-dimensional wavelet decomposition for each variable separately [55]. The *standard decomposition*, as described in [55], consists in the application of a one-dimension Fast Wavelet Transform (FWT) to each row of data values. This operation gives an average signal along with detail coefficients for each row. Next, these transformed rows are treated as if they formed an image, and a one-dimensional FWT is applied to each column. The resulting values are all detail coefficients except for a single overall averaged image.

Since we want to obtain an image at various levels of resolution, a smoothing procedure is required, which is applied before reconstructing the original signal as follows: all details generated after a given number of decomposition levels using a FWT are set equal to zero before a full reconstruction procedure is applied to restore the original number of data points. Reconstruction is required to keep the initial number of data values. This is actually a procedure used in signal compression, one of the biggest application of wavelets.

3.3. 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 [42]:

$$\rho_a\left(\vec{r}-\vec{R}_a\right)=\sum_{i=1}^{5} w_{a,i}\left[\left(\frac{2\zeta_{a,i}}{\pi}\right)^{3/4} \mathrm{e}^{-\zeta_{a,i}|\vec{r}-\vec{R}_a|^2}\right]^2, \tag{28}$$

where $\vec{R}_a$ is the position vector of atom a, and $w_{a,i}$ and $\zeta_{a,i}$ are the fitted parameters, respectively, as reported in [56]. ρ_{M} is then calculated as

$$\rho_{\mathrm{M}}=\sum_a Z_a\rho_a, \tag{29}$$

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 2. (Coefficients and exponents can be seen and downloaded from [56].)

As asserted by Gironés et al. [57], the ASA of the ED function allows to describe molecular shapes in almost the same way as *ab initio* quantum mechanical computations. In this 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. [43]:

$$\rho_{a,t}\left(\vec{r}-\vec{R}_a\right)=\sum_{i=1}^{5} a_{a,i}(1+4b_{a,i}t)^{-3/2}\exp\left(\frac{-b_{a,i}|\vec{r}-\vec{R}_a|^2}{1+4b_{a,i}t}\right) \tag{30}$$

where:

$$b_{a,i}=2\zeta_{a,i} \quad \text{and} \quad a_{a,i}=w_{a,i}\left(\frac{b_{a,i}}{\pi}\right)^{6/4}. \tag{31}$$

Table 2
$w_{a,i}$ and $\zeta_{a,i}$ atomic shell approximation coefficients as reported in [56] for C atom ($a=6$).

i	$w_{a,i}$ (e^-)	$\zeta_{a,i}$ (bohr^{-2})
1	0.5079501021	0.2065113939
2	0.1888265284	0.5553428097
3	0.1765106174	7.3531983076
4	0.1125973447	22.0875917625
5	0.0141154074	107.3463850250

In this context, t is seen as the product of a diffusion coefficient with time. On a similar basis, Duncan and Olson [8] generated low resolution molecular surfaces by convolving the density function with a 3D Gaussian function of appropriate variance.

4. Applications

In previous applications of low resolution molecular representations to the comparison of small molecules, rigid benzodiazepine-related structures were considered [38, 40]. In the present work, a new set of drug molecules was selected after the work of Nissink et al. [58]. These authors compared three thrombin inhibitors, MQPA, NAPAP, and 4-TAPAP [59], at the particular resolution value of 2.5 or 3 Å (a value of 2.5 Å is mentioned in the text, but a value of 3 Å is reported in figure 4 of [58]). All three selected molecules adopt a star shape with three branches terminated either by a cyclic substructure linked to a sulfonyl function (branch I), a piperidine ring (branch II), or an amidino group (branch III) (figure 4). The three-dimensional atom coordinates of the considered protein ligands were derived from crystallographic data reported in the Protein Data Bank (PDB) [60,61]. The PDB codes for MQPA, NAPAP, and 4-TAPAP are 1etr, 1ets, and 1ett, respectively. H atoms were added to the crystalline PDB structures using InsightII [62] in order to generate the complete structures as described in [59].

Considering the resolution values reported in the work of Nissink et al. and the geometry of the three molecules, this set of molecules was considered as relevant to test the superposition approach based on graph representations of low resolution electron density maps. Indeed, one could argue that if a critical point (CP) representation does not permit to properly differentiate between the various chemical functions, many possible superposition results could be obtained. The superposition algorithm that was used in the present work is based on a Monte Carlo/Simulated Annealing (MC/SA) approach. With respect to previous works on drug molecules which are known to bind to benzodiazepine receptors [38,40], a slightly modified function, which is described later, was used to evaluate the superposition results.

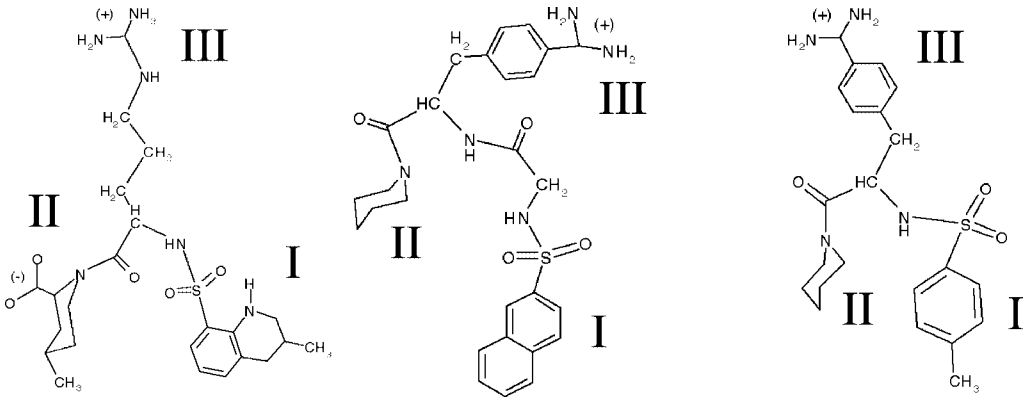


Figure 4. Planar structure formula of compounds MQPA, NAPAP, and 4-TAPAP.

4.1. Electron density map calculations

All three approaches described above were applied to 128 × 128 × 128 (16 × 16 × 16 Å) ED maps. Within the crystallographic approach, the program XTAL [37] was used to calculate maps at resolutions of 2.5 and 3 Å, using the P1 space group operations (figure 5) as in Nissink et al. work [58]. Atomic Shell Approximation (ASA) ED functions were computed using an in-house program with a smoothing parameter t equal to 1.1 and 1.5 bohr^2 (figure 6), and finally, full ED maps for the three molecules were generated using *ab initio* 6-31G* RHF MO-LCAO-SCF calculations. At this point, computing valence or full ED distribution functions was a point to discuss. Indeed, valence ED may be a more relevant property to consider when describing intermolecular interactions. However, full ED functions were analyzed for two main reasons. The first one is that only full ED distributions will give a view that is comparable to promolecular full ED representations as generated through the crystallography and ASA formalisms. The second one is that a CP analysis of a valence ED at atomic level leads to a huge number of CPs. Indeed, ρ maxima are found both at atom and bond locations [64]. This disappears in lower resolution images wherein important structural features vanish too. This is for example observed for benzodiazepine-related molecules wherein chlorine atoms, which are essential steric/lipophilic functions, are not seen at all in smoothed ED maps. Therefore, as emphasized by Popelier [27] in his Bond Critical Point (BCP) analysis paper, core density is essential to the representation of molecular structures.

In the case of the ASA approach, a molecular ED distribution is the summation over squared Gaussian functions convolved with a normal Gaussian distribution function whose standard deviation is equal to $\sqrt{2t}$ [65]. Thus, values of t = 1.1 and 1.5 bohr^2 correspond to deviations equal to 1.48 bohr (0.78 Å) and 1.73 bohr (0.92 Å), respectively.

For the wavelet-based multiresolution analysis (WMRA) approach, the only representation that is not based on a promolecular description of the molecular ED, the filter Φ that was selected to smooth the original *ab initio* quantum mechanical images is the Daubechies' filter of order 10 as presented in table 1. The procedure is implemented in an in-house program. Four levels of decomposition followed by reconstruction with all detail coefficients set equal to zero were needed to generate images that are similar to those obtained using the two other approaches (figures 7 and 8). In figure 7, it is clearly seen that images obtained after one and two levels of decomposition/reconstruction are extremely noisy, while at further decomposition levels, images are smooth averages of the original ED distribution functions. This is due to the scale of the ED fluctuations as illustrated in figure 9. In this figure, a one-dimensional ASA representation of the ED function associated with the C atom is shown together with its corresponding smoothed versions calculated after two and four levels of decomposistion/reconstruction with the filter D_{10}. A way to circumvent this problem would be to use non-orthogonal filters such as spline functions [66,67] or convolution products of Daubechies' filters [68]. This last approach is now considered for applications to the three molecules studied in this work. The use of non-orthogonal wavelets presents the great advantage to avoid decimation during the decomposition procedure. The transform is also translationaly invariant.

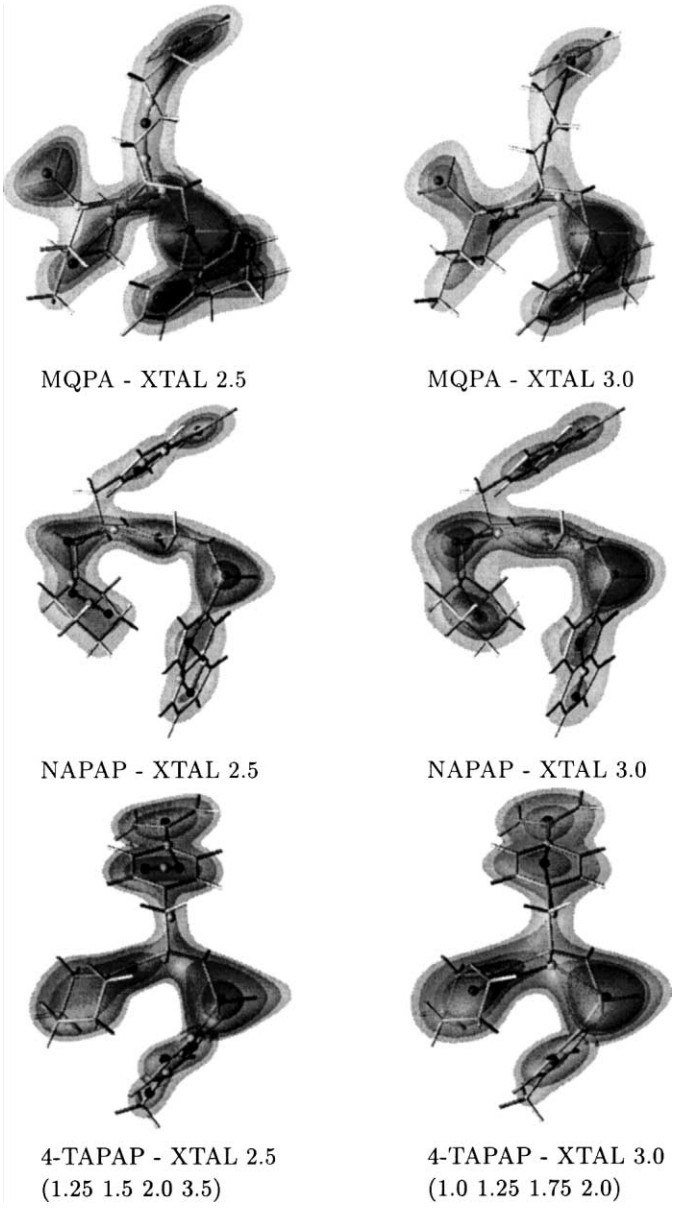


Figure 5. Iso-contours of the electron density distributions calculated for molecules MQPA, NAPAP, and 4-TAPAP, using XTAL at a resolution of 2.5 Å (iso = 1.25, 1.75, 2.0, 3.5 $e^-/Å^3$) and 3.0 Å (iso = 1.0, 1.5, 1.75, 2.0 $e^-/Å^3$) superimposed to their molecular skeleton. Critical points graph representations were obtained from topological analyses using ORCRIT (peaks: large spheres, passes: small spheres). Figures were generated using DataExplorer [63].

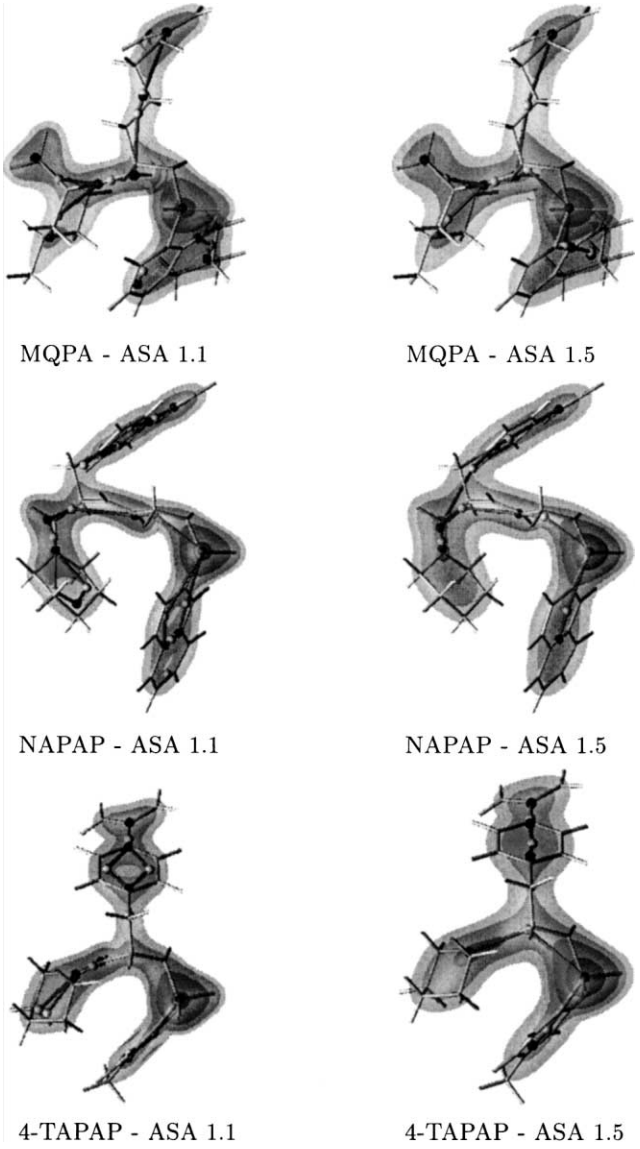


Figure 6. Iso-contours of the electron density distributions calculated for molecules MQPA, NAPAP, and 4-TAPAP, using the ASA approach with $t = 1.1$ bohr^2 (iso = 0.125, 0.150, 0.200, 0.300 e$^-$/bohr^3) and 1.5 bohr^2 (iso = 0.100, 0.125, 0.150, 0.200 e$^-$/bohr^3) superimposed to their molecular skeleton. Critical points graph representations were obtained from topological analyses using ORCRIT (peaks: large spheres, passes: small spheres). Figures were generated using DataExplorer [63].

Figure 7 also shows that, due to the size of the grid interval, a drastic change in the shape of the ED contours does not appear before the fourth level of decomposition/reconstruction, i.e., in the image associated with variations of scale of 2 Å. Indeed, smoothed images at the first, second, and third levels of decomposition/reconstruction

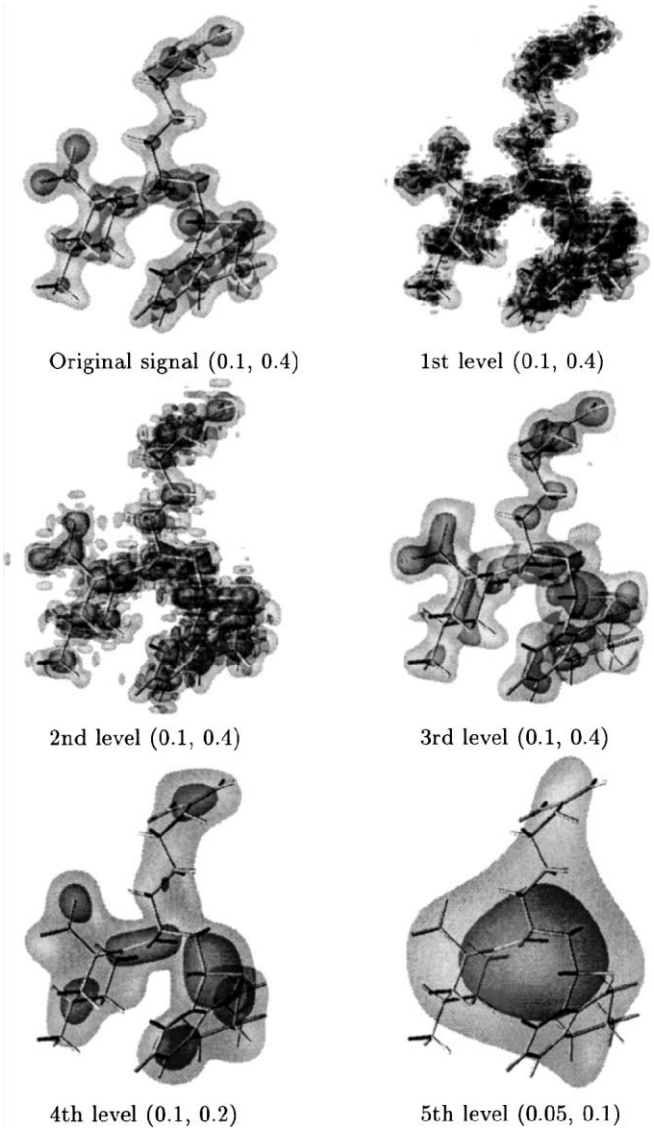


Figure 7. Iso-contours (in parentheses and in $e^-/bohr^3$) of the electron density distributions calculated for molecule MQPA using the WMRA approach with $\Phi = D_{10}$ superimposed to their molecular skeleton. Figures were generated using DataExplorer [63].

are associated with variations still at the atomic scale, i.e., 0.25, 0.5, and 1 Å. Atomic details are definitely lost at the fourth level, and contours thus appear a lot less corrugated than in the original image. The fifth level is a representation of the averaged ED function wherein all details at a scale lower than 4 Å were discarded.

On a computational point of view, the three approaches are characterized by very different computation times. All values reported hereafter correspond to calculations of

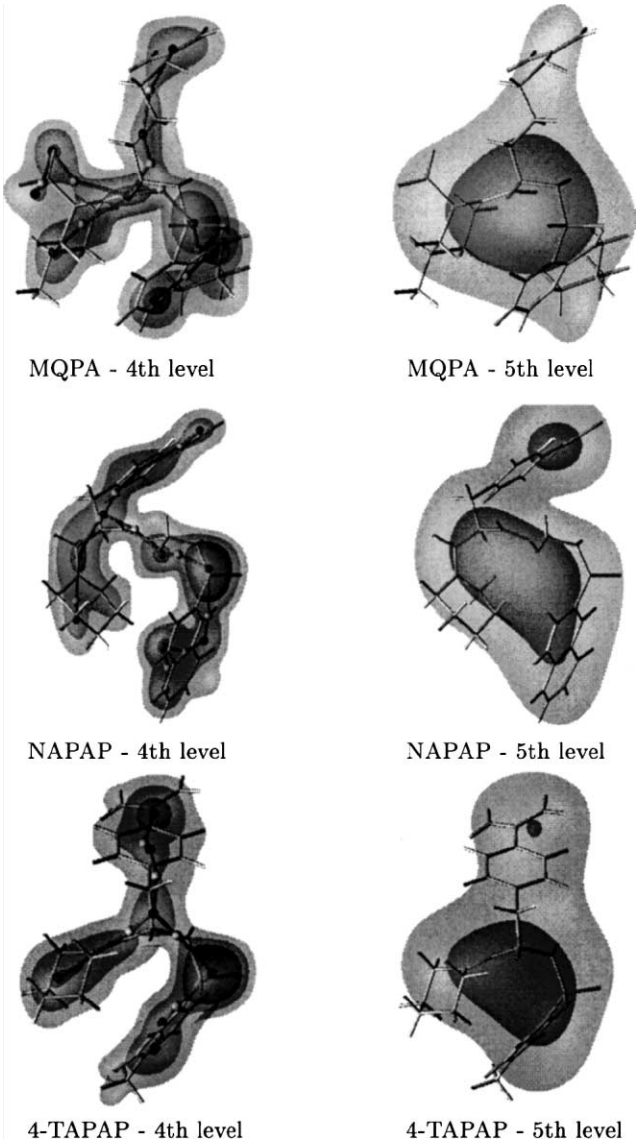


Figure 8. Iso-contours of the electron density distributions calculated for molecules MQPA, NAPAP, and 4-TAPAP, using the WMRA approach with $\Phi = D_{10}$ and $J = -4$ (iso = 0.100, 0.170, 0.250 e^-/bohr^3) or $J = -5$ (iso = 0.050, 0.100 e^-/bohr^3) superimposed to their molecular skeleton. Critical points graph representations were obtained from topological analyses using ORCRIT (peaks: large spheres, passes: small spheres). Figures were generated using DataExplorer [63].

$128 \times 128 \times 128$ ED grids carried out on an IBM R6000 model 580 for the largest compound NAPAP, otherwise specified. The crystallographic approach is very fast (about $2'$ at 2.5 Å resolution) but presents the disadvantage of generating areas of negative ED values resulting from the Fast Fourier Transform truncation error. It is also limited to

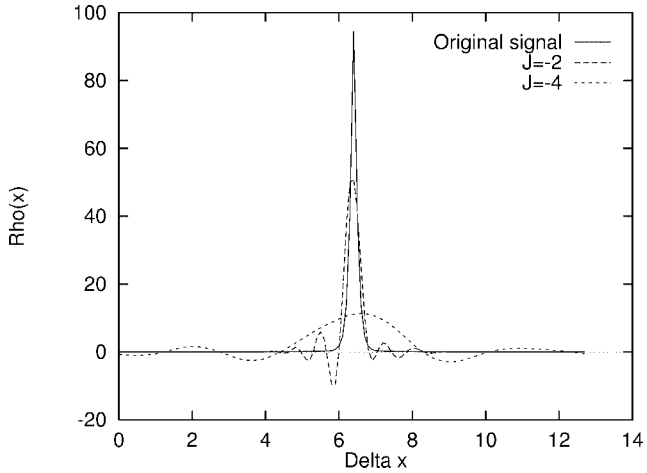


Figure 9. Original and smoothed ED functions (in e^-/bohr^3) of the C atom calculated using the ASA formalism and coefficients reported in table 2. $\triangle x$ is equal to 0.5 Å.

the generation of promolecular descriptions of molecular ED distributions. The WMRA procedure is reasonably fast (about 450″ using filter D_{10} over four levels of decomposition/reconstruction), but it follows the lengthy calculation of an *ab initio* quantum mechanical ED map (about 3 hours on an IBM SP2 120 MHz processor). The resolution levels that can be modeled are also strongly dependent on the grid interval of the ED map. But the great advantage of the WMRA method is that it can be applied to any n-dimensional molecular property, i.e., electron density, electrostatic potential, etc. The analytical ASA approach requires medium total computing times (about 3300″), and the resolution, which is determined by the value of t, can be varied continuously. However, it requires an analytical expression of the molecular property that is considered.

4.2. Critical point analysis results

After the generation of all electron density (ED) maps (figures 5, 6, and 8 left), a critical point (CP) analysis program, ORCRIT, was applied to locate the peaks and passes above a selected density cut-off value. This cut-off was applied to each map in order to eliminate CPs which originate from the background noise. This is necessary, especially for maps generated using XTAL wherein the procedure gives rise to ripples due to fast Fourier transform truncation errors. For such maps, a cutoff of 1.0 e^-/Å^3 was applied. In the frame of the wavelet-based multiresolution analysis (WMRA) approach, negative density areas originates from the fluctuations of the wavelet function itself (figure 9) and, in this last case, a cut-off value of 0.100 e^-/bohr^3 was selected.

The main locations of the peaks observed in the various smoothed ED maps are reported in table 3 and density values are presented in figures 10–12. In these figures, the numbering of the CPs is presented in the decreasing order of their density value. Also, connections between CPs of the same kind, i.e., peak–peak or pass–pass, are not drawn since they are not physically significant.

Table 3
Composition of the critical point graphs (+ = peaks, * = passes) and location of the peaks for compounds MQPA, NAPAP, and 4-TAPAP, obtained with the program ORCRIT [28] applied to smoothed electron density maps using the crystallographic (XTAL), analytical (ASA), and wavelet (WMRA) approaches. For each of the three molecular branches (I, II, and III), the number of peaks and their label (in parentheses) are reported (see figures 10–12).

Method Resolution	XTAL 2.5 Å	XTAL 3.0 Å	ASA $t = 1.1$ bohr^2	ASA $t = 1.5$ bohr^2	WMRA D_{10} 4th level
MQPA	9+, 8*	7+, 6*	10+, 7*	8+, 7*	9+, 8*
(I) SO_2 group	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
(I) Quinoline	3 (4, 6, 9)	2 (2, 3)	3 (7, 9, 11)	2 (7, 11)	2 (2, 3)
(II) Carbonyl	0	1 (7)	0	1 (2)	1 (4)
(II) Carboxyl	1 (2)	1 (5)	1 (2)	1 (3)	2 (5, 15)
(II) Piperidine	1 (8)	1 (8)	2 (4, 10)	2 (4, 13)	1 (6)
(III) Amidino	1 (5)	1 (4)	1 (5)	1 (6)	1 (7)
(III) Alkyl chain	1 (12)	0	1 (15)	0	1 (8)
Other	1 (3)	0	1 (3)	0	0
NAPAP	9+, 8*	8+, 6*	10+, 13*	7+, 6*	10+, 9*
(I) SO_2 group	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
(I) Naphtyl	2 (6, 13)	2 (7, 13)	2 (7, 8)	1 (8)	3 (4, 6, 7)
(I) Carbonyl	1 (2)	1 (2)	1 (3)	1 (5)	0
(II) Carbonyl	1 (3)	1 (3)	1 (2)	1 (2)	1 (3)
(II) Piperidine	2 (7, 8)	1 (4)	2 (4, 20)	1 (4)	1 (14)
(III) Amidino	1 (4)	1 (8)	1 (6)	1 (6)	1 (13)
(III) Phenyl	1 (11)	1 (6)	2 (9 13)	1 (10)	1 (5)
Other	0	0	0	0	2 (2, 10)
4-TAPAP	8+, 6*	5+, 3*	7+, 9*	6+, 5*	7+, 6*
(I) SO_2 group	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
(I) Toluene	3 (7, 8, 11)	1 (5)	1 (9)	1 (9)	2 (4, 6)
(II) Piperidine	1 (2)	1 (2)	1 (2)	1 (2)	1 (3)
(III) Amidino	1 (3)	1 (3)	1 (5)	1 (3)	1 (2)
(III) Phenyl	2 (4, 5)	1 (4)	2 (6, 8)	2 (5, 7)	1 (9)
Other	0	0	1 (3)	0	1 (5)

Depending upon the resolution, various chemical groups were detected. In every case, the sulfonyl, the carbonyl and/or carboxy, and the amidino groups led to a peak located on the sulfur atom, on the carbon atom, and on the central C atom, respectively. Six-membered rings are characterized by either 1 or 2 peaks. In the ASA application with $t = 1.5$ bohr^2, a low resolution representation, there is only one peak for the whole naphtyl group of compound NAPAP. The presence of peaks on the alkyl chain of compound MQPA is observed only at the highest resolution levels, i.e., in XTAL 2.5 Å and ASA $t = 1.1$ images. In this very last case, N atoms located in rings are also detectable.

It is also observed that the sulfonyl groups bear the highest density points. Such groups should thus easily match together in a superposition algorithm using a density-based similarity measure. In the lowest resolution images, i.e., in the ED generated using

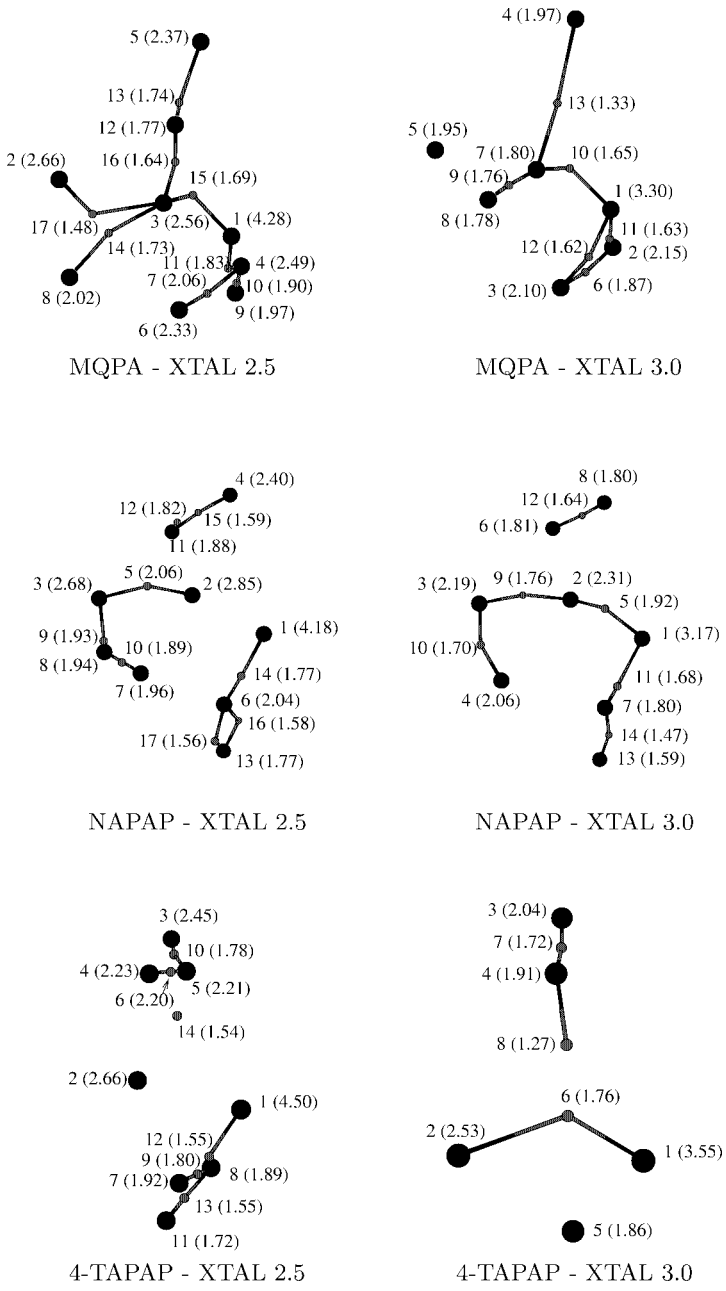


Figure 10. Critical point number and density value in $e^{-}/Å^{3}$ (in parentheses) of the electron density distributions calculated for molecules MQPA, NAPAP, and 4-TAPAP, using XTAL at a resolution of 2.5 and 3.0 Å. Critical points graph representations were obtained from topological analyses using ORCRIT (peaks: large spheres, passes: small spheres). Structures were generated using XMol [69].

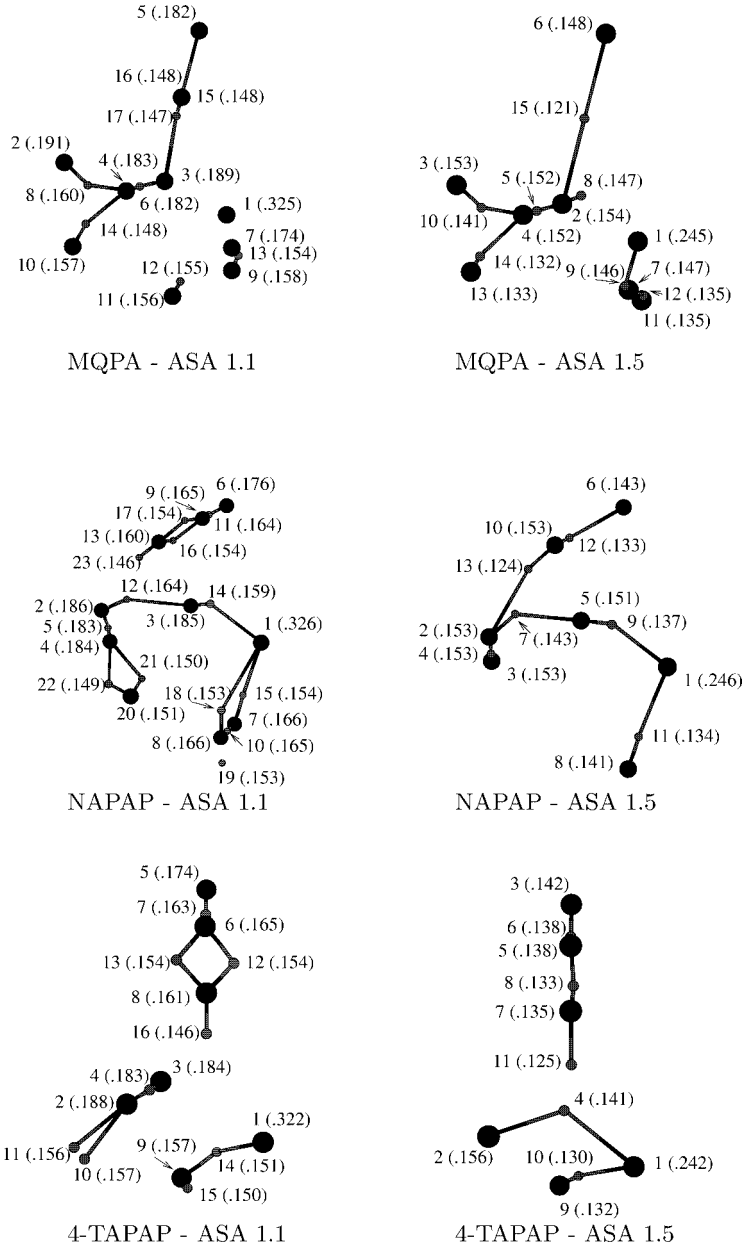
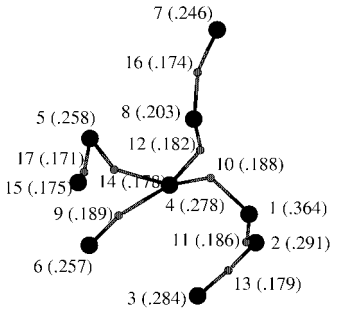
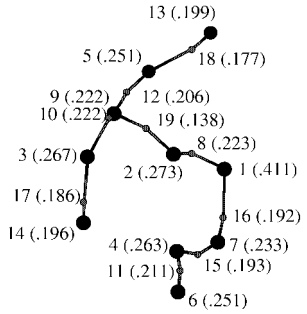


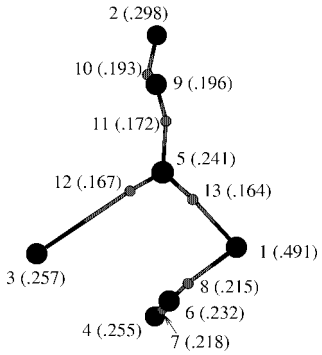
Figure 11. Critical point number and density value in $e^-/bohr^3$ (in parentheses) of the electron density distributions calculated for molecules MQPA, NAPAP, and 4-TAPAP, using the ASA approach with $t = 1.1$ and 1.5 $bohr^2$. Critical points graph representations were obtained from topological analyses using ORCRIT (peaks: large spheres, passes: small spheres). Structures were generated using XMol [69].



MQPA - 4th level



NAPAP - 4th level



4-TAPAP - 4th level

Figure 12. Critical point number and density value in e^-/bohr^3 (in parentheses) of the electron density distributions calculated for molecules MQPA, NAPAP, and 4-TAPAP, using the WMRA approach with $\Phi = D_{10}$ and $J = -4$. Critical points graph representations were obtained from topological analyses using ORCRIT (peaks: large spheres, passes: small spheres). Structures were generated using XMol [69].

XTAL at a 3 Å resolution and ASA with $t = 1.5$ bohr^2, one observes that the alkyl part of molecule MQPA bears no peaks at all. Even in the highest resolution images, this chemical group leads to one CP with a rather low density values. This implies that chemical groups composed of C atoms only do not always appear at the same resolution level: a ring will still show up when an alkyl chain is undetected. The order of the peak density values can be approximately summarized as: sulfonyl $>$ (CO(O) $>$ or $<$ amidino) $>$ alkyl chain. C-rings are not included in this sequence due to the highly variable distribution of their peak density values. Since CPs, regardless of the molecular structure, tend to be located on similar chemical groups (sulfonyl, amidino, phenyl, etc.) it is concluded that there may be a possible theoretical assignment of a chemical group to a peak, but this assertion needs to be validated through further works. In a first step to comfort this assertion, the CP graph representations were considered for the superposition of the three molecules under study, and results were described and compared with experiments in the following section.

4.3. Pairwise superposition results

As already mentioned in the previous section, only peaks and passes with a density value above a pre-determined cut-off value were considered. Passes were included in the procedure since they allow to catch the molecular skeleton more completely than peaks alone. A Monte Carlo/Simulated Annealing (MC/SA) algorithm was used to match each pair of critical point (CP) graphs based on the density and distance values associated with all critical points (CP). The use of these two properties, ρ and d, was initially proposed in [70] and further implemented in a genetic algorithm dedicated to graph superposition [71].

A SA simulation consists of a sequence of several Monte Carlo (MC) sampling procedures that are carried out at progressively decreasing rates of acceptance:

$$p = \exp(-\beta \cdot \mathrm{RMS}), \tag{32}$$

where β is the parameter that controls the rate of acceptance and modulates the evaluation function RMS:

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\rho_i - \rho_i^{\mathrm{ref}}\right)^2 + \frac{1}{\mathrm{nbn}}\sum_{i=1}^{\mathrm{nbn}}\left(d_i - d_i^{\mathrm{ref}}\right)^2}, \tag{33}$$

where, in a completely connected graph composed of n points, the number of connections between the points is given by

$$\mathrm{nbn} = \frac{n(n-1)}{2}. \tag{34}$$

In formula (33), the ρ- and d-dependent terms have a similar weight. However, as already mentioned, standardization was applied to ρ values and distances separately in order to avoid a summation over two different contributions with different units and

magnitudes. In a standardization procedure, each data d of a data set is transformed according to

$$d_{\text{standardized}} = \frac{d - \overline{d}}{\sigma_d}, \tag{35}$$

where $\overline{d}$ and σ_d are the mean and the standard deviation of the data set, respectively. This approach will reveal itself to be as efficient, if not more, than the previous one wherein no standardization was applied [38,40]. This operation is thus similar to a reweighting of the density and distance values, in favour of the density contribution, especially for data sets obtained using the Atomic Shell Approximation (ASA) and wavelet-based multiresolution analysis (WMRA) approaches wherein, due to the unit system, density values are extremely small with respect to distance values (table 4). After standardization, and in order to force the matching of CPs of the same kind, each density value ρ (peak) value was incremented by 10, while ρ (pass) values remained unchanged. This increment has a null effect once two CPs of the same kind are matched together because, as described later, the evaluation function is based on the difference of the two corresponding density values.

Each SA run consisted of 20 MC procedures (each of 200000 iterations) that were carried out with β ranging between 0.05 and 1. In the largest and smallest cases (17 pairs of CPs and 8 pairs of CPs), the SA procedure took about 1500″ and 70″on an IBM R6000 model 580, respectively. Best CP matching results were then verified by an additional MC run of 10^6 iterations with a β value selected in order to obtain a number of accepted moves that is close to the number of rejected moves. No improvement was observed in any of the 15 superposition calculations. Matching results were then translated into visual molecular superimpositions using the program QUATFIT [72]. Superposition results corresponding to the lowest RMS values obtained within each strategy are presented in tables 5–9 and figures 13–15 wherein the reference graph (the largest one of each pair) is shown in bold.

Examination of the SA results shows that all, but two, superimposition patterns were correctly predicted, i.e., all three branches were properly matched. When using unstandardized data, three optimal superposition patterns were wrong. With standardized data, the two wrong superimposition results are observed for the pair 4-TAPAP/MQPA at the lowest resolutions, i.e., using XTAL at 3.0 Å (figure 13) and ASA with $t = 1.5$ bohr^2 (figure 14). A careful analysis of the matched CP pairs for these two cases shows that none of the CPs located on the amidino branch of compound MQPA, i.e., points 4 and 13 at 3 Å and points 6 and 15 at ASA $t = 1.5$ bohr^2 are present in the matched CP pairs. The problem probably originates from the fact that, at those two low resolution representations, the alkyl chain of compound MQPA does not bear any CP. With the same two approaches applied at a higher resolution level, the superposition results are in agreement with experimental structures and modelled patterns [58].

With unstandardized data, a third mismatch is observed for the same pair of molecules, described at the ASA representation level with $t = 1.1$ bohr^2. In this case, the

Table 4
Density (ρ in $e^-/Å^3$ for XTAL, $e^-/bohr^3$ for ASA and WMRA) and distance (d in Å) means and standard deviations associated with all critical point (CP) data sets generated for compounds MQPA, NAPAP, and 4-TAPAP, obtained with the program ORCRIT [28] applied to smoothed electron density maps using the crystallographic (XTAL), analytical (ASA), and wavelet (WMRA) approaches.

Method	XTAL	XTAL	ASA	ASA	WMRA D_{10}
Resolution	2.5 Å	3.0 Å	$t = 1.1$ bohr^2	$t = 1.5$ bohr^2	$J = -4$
MQPA					
$\overline{\rho}$	2.148	1.917	0.174	0.149	0.224
σ_ρ	0.633	0.452	0.041	0.027	0.055
$\overline{d}$	4.828	4.392	4.767	4.353	4.775
σ_d	2.022	1.965	2.063	2.313	1.991
NAPAP					
$\overline{\rho}$	2.112	1.855	0.170	0.150	0.227
σ_ρ	0.622	0.470	0.035	0.029	0.055
$\overline{d}$	5.591	5.534	5.944	4.895	5.572
σ_d	2.609	2.504	2.640	2.313	2.643
4-TAPAP					
$\overline{\rho}$	2.143	2.080	0.173	0.147	0.238
σ_ρ	0.735	0.646	0.041	0.031	0.082
$\overline{d}$	4.969	5.010	4.874	4.639	4.498
σ_d	2.629	2.140	2.313	2.346	2.107

Table 5
Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs obtained using ORCRIT [28] for electron density maps calculated at a resolution of 2.5 Å with XTAL [37]. For each pair of graphs, all n points of the smallest molecular graph were considered. Numbering of critical points is presented in figure 10. The largest CP graph of each pair is in bold. Wrong CP matches are in italic.

	RMS																	
MQPA	0.716	1	2	3	4	5	*6*	7	8	9	10	11	12	13	14	*15*	16	17
NAPAP		1	3	2	6	4	*7*	17	8	13	16	14	11	15	10	*5*	12	9
4-TAPAP	0.701	1	2	3	*4*	5	6	7	8	9	10	11	12	13	*14*			
MQPA		1	3	5	*2*	12	16	4	6	7	13	9	11	10	*17*			
4-TAPAP	0.640	1	*2*	3	*4*	5	6	7	*8*	9	10	11	12	13	14			
NAPAP		1	*2*	4	*3*	11	12	6	*7*	16	15	13	14	17	5			

corresponding RMS value is equal to 3.817, a value which reflects the superposition of CPs of different kinds.

A detailed analysis of all CP pairs reported in tables 5–9 shows that, even for good superposition results, some CP pairs are structurally inconsistent. For example, in table 5, CP pair (6,7) in compounds (MQPA, NAPAP) corresponds to the superposition

Table 6
Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs obtained using ORCRIT [28] for electron density maps calculated at a resolution of 3.0 Å using XTAL [37]. For each pair of graphs, all n points of the smallest molecular graph were considered. Numbering of critical points is presented in figure 10. The largest CP graph of each pair is in bold. Wrong CP matches are in italic.

	RMS													
MQPA	0.918	1	*2*	3	4	*5*	*6*	7	8	9	10	11	12	13
NAPAP		1	*4*	7	8	*6*	*10*	2	3	9	5	11	14	12
4-TAPAP	0.774	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*					
MQPA		1	*3*	*5*	*8*	*2*	*12*	*9*	*10*					
4-TAPAP	0.596	1	2	3	4	5	6	7	8					
NAPAP		1	4	8	6	7	5	12	9					

Table 7
Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs obtained using ORCRIT [28] for electron density maps calculated using the ASA method with $t = 1.1$ bohr^2. For each pair of graphs, all n points of the smallest molecular graph were considered. Numbering of critical points is presented in figure 11. The largest CP graph of each pair is in bold. Wrong CP matches are in italic.

	RMS																	
MQPA	0.717	1	*2*	3	4	5	*6*	7	8	9	10	*11*	*12*	13	14	15	16	17
NAPAP		1	*13*	3	2	6	*14*	7	12	8	4	*20*	*21*	10	5	9	11	17
4-TAPAP	3.494	1	2	3	*4*	5	*6*	*7*	*8*	9	10	11	12	13	14	15	*16*	
MQPA		1	10	2	*11*	5	*3*	*15*	*4*	7	14	8	16	17	13	12	*6*	
4-TAPAP	0.395	1	2	3	*4*	5	6	7	8	9	10	11	12	13	14	15	16	
NAPAP		1	4	3	*14*	6	9	11	13	8	21	22	16	17	15	18	23	

Table 8
Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs obtained using ORCRIT [28] for electron density maps calculated using the ASA method with $t = 1.5$ bohr^2. For each pair of graphs, all n points of the smallest molecular graph were considered. Numbering of critical points is presented in figure 11. The largest CP graph of each pair is in bold. Wrong CP matches are in italic.

	RMS													
NAPAP	0.702	1	2	3	4	5	6	*7*	8	9	*10*	11	12	*13*
MQPA		1	4	13	14	2	6	5	11	8	*3*	12	15	*10*
4-TAPAP	0.606	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	9	10	*11*		
MQPA		1	*7*	*3*	*9*	*13*	*10*	*4*	*14*	11	12	*8*		
4-TAPAP	0.560	1	2	3	4	5	6	*7*	8	9	10	*11*		
NAPAP		1	3	6	9	10	12	*5*	13	8	11	*7*		

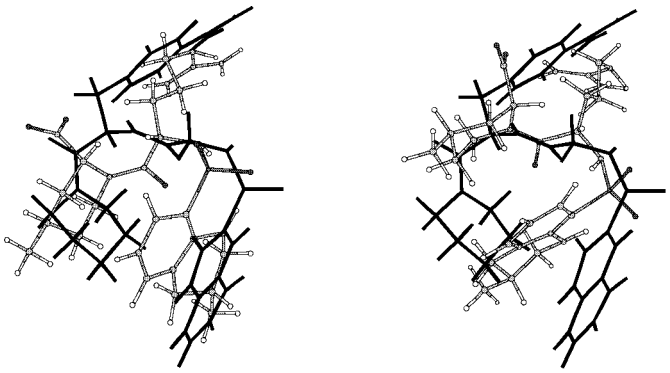
Table 9
Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs obtained using ORCRIT [28] for electron density maps calculated using the WMRA method with $\Phi = D_{10}$ and 4 levels of decomposition/reconstruction. For each pair of graphs, all n points of the smallest molecular graph were considered. Numbering of critical points is presented in figure 12. The largest CP graph of each pair is in bold. Wrong CP matches are in italic.

	RMS																	
MQPA	0.873	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	*17*
NAPAP		1	4	6	2	9	3	13	5	17	8	16	19	11	10	14	18	*12*
4-TAPAP	0.749	1	2	3	4	5	6	7	8	9	10	11	12	13				
MQPA		1	7	6	3	4	2	13	11	8	16	12	14	10				
4-TAPAP	0.678	1	2	3	4	5	6	7	8	9	10	11	12	13				
NAPAP		1	5	14	6	2	7	11	16	9	12	10	17	8				

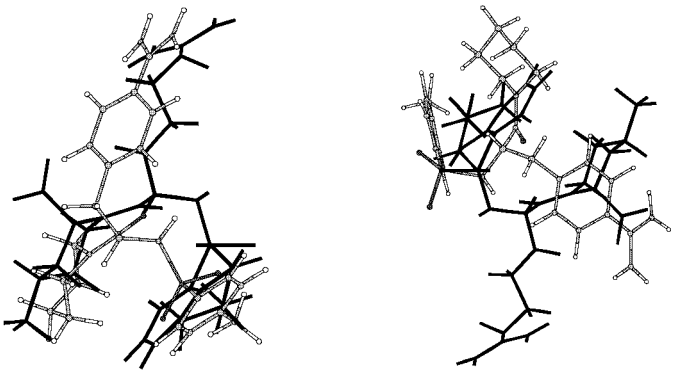
of points in branch II of MQPA and branch I in NAPAP (figure 10). Another inconsistency is also raised due to the fact that CP 15 of compound MQPA does not have any structurally corresponding point in NAPAP, and was thus matched with CP 5 of this last compound. All found inconcistencies are shown in italics in tables 5–9. Such wrong matchings are inevitable when one considers all CPs in the superposition procedures. Indeed, a given molecular branch contains different numbers of CP depending upon the molecular structure. This difference in the number of points is large at atomic resolution and only reduced, but not totally cancelled, at lower resolution levels. The number of inconsistencies (wrong matchings) is however low with respect to the number of right ones, and does not affect the global superimposition patterns. One can imagine that, rather than superposing all CPs, finding a common substructure between two compounds would be more appropriate. However, this should be achieved under constraints in order to avoid the risk to eliminate structurally important points. In the two wrong superposition patterns, which occurred at the lowest resolution levels (RMS = 0.774 in table 6 and RMS = 0.606 in table 8, respectively), almost all pairs of matched CP are wrong ones.

Another consequence of the difference in the number of CPs between the three molecules leads to a extremely high RMS value, 3.494, for pair MQPA/4-TAPAP described at the ASA level with $t = 1.1$ bohr^2 (table 7). This value originates from two matchings between CPs of different types. Points 4 and 7 of compound 4-TAPAP are passes, while points 11 and 15 of compound MQPA are peaks. On the contrary, 4-TAPAP and NAPAP have very similar graph structures, and the RMS value associated to the best superposition pattern is very low, 0.395.

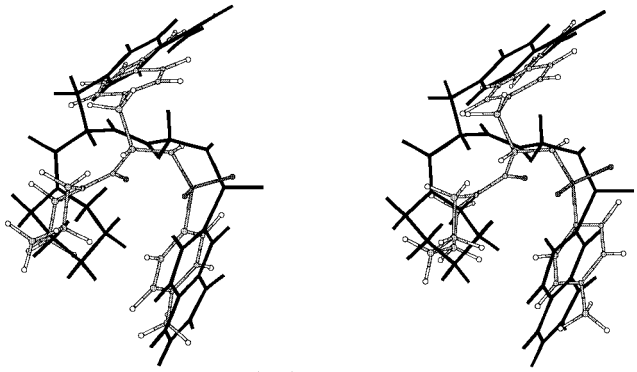
In conclusion, the method that presents the least number of inconsistencies is the WMRA one. There is only one wrong CP pair obtained when superposing compounds MQPA and NAPAP. The question to determine whether this is related to a non-promolecular representation of the full ED distribution is raised. Further analyses are thus required to answer this point.



MQPA - **NAPAP**



4-TAPAP - **MQPA**



4-TAPAP - **NAPAP**

Figure 13. Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs of molecules MQPA, NAPAP, and 4-TAPAP obtained from electron density maps generated using XTAL at a resolution of 2.5 (left) and 3.0 Å (right). Figures were generated using XMol [69].

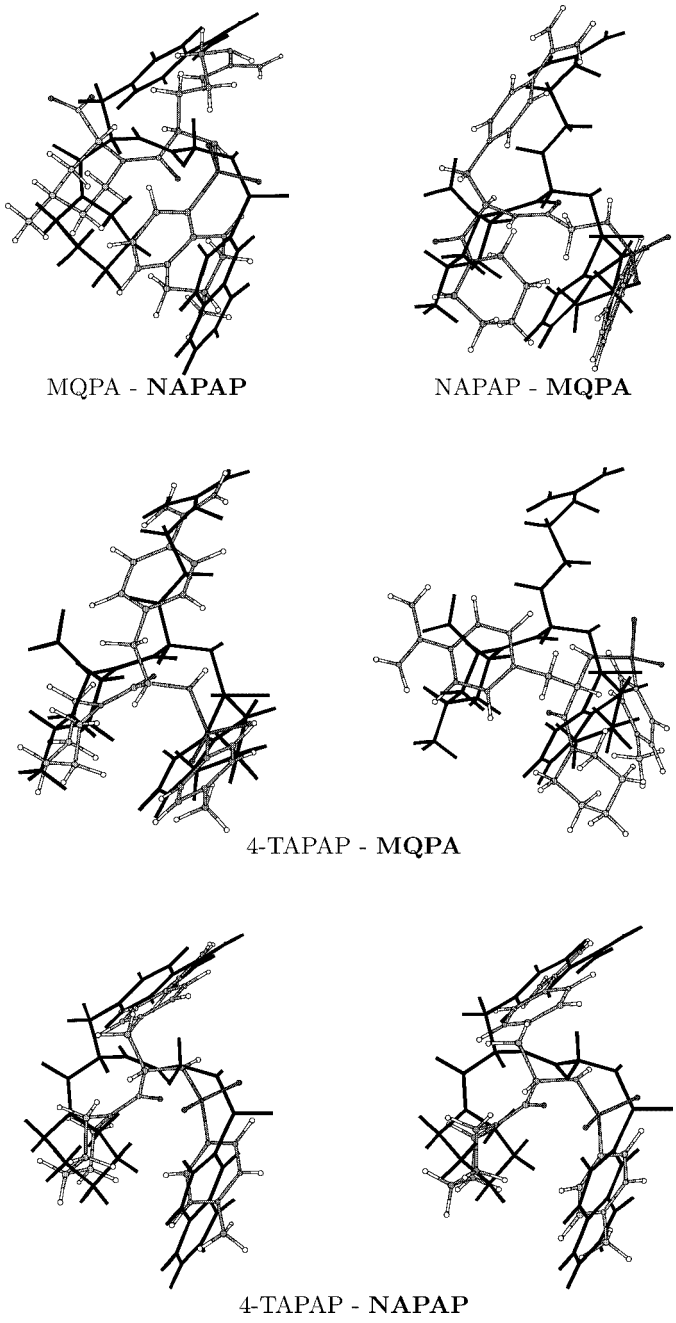
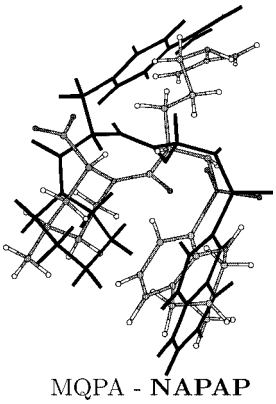
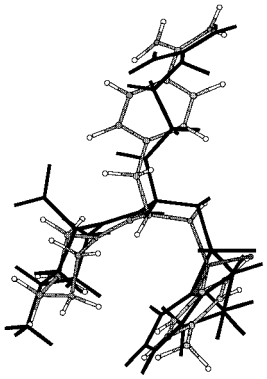


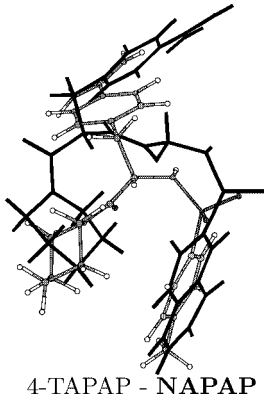
Figure 14. Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs of molecules MQPA, NAPAP, and 4-TAPAP obtained from electron density maps generated using the ASA method with $t = 1.1$ (left) and $t = 1.5$ bohr^2 (right). Figures were generated using XMol [69].



MQPA - **NAPAP**



4-TAPAP - **MQPA**



4-TAPAP - **NAPAP**

Figure 15. Best superposition results obtained using a SA algorithm for the comparison of the critical point graphs of molecules MQPA, NAPAP, and 4-TAPAP obtained from electron density maps generated using the WMRA method with $\Phi = \mathrm{D}_{10}$ and $J = -4$. Figures were generated using XMol [69].

5. Conclusions and perspectives

A simulated annealing procedure was applied to critical point (CP) graphs representations of smoothed full electron density (ED) distribution functions in order to superpose small biological molecules. A set of three thrombin inhibitors, already considered in superposition calculations using low resolution data [58], was selected as reference materials.

In the present work, three different smoothing approaches were applied which led to very similar results. Two methods, a crystallography-based algorithm and an analytical approach, allowed the calculation of promolecular ED distributions. The third one, a wavelet-based multiresolution analysis (WMRA), consisted in smoothing original ED functions through the cancellation of details at a predetermined number of scales. The crystallography-based method is the fastest, but the most time-consuming wavelet-based multiresolution analysis is applicable to any grid representation of a molecular property. The analytical approach presents the advantage to generate functions that are continuously scalable, but requires an analytical description of the molecular property considered.

A topological analysis program was then applied to the low resolution ED maps and the obtained CP graphs revealed a structure of the ligands described in terms of chemical groups rather than in terms of individual atoms. CP representations of the molecular skeleton are thus interesting in the sense that reduction of chemical groups to simple points is a procedure that is commonly used by chemists for comparing molecules, for example, in the search for a common substructure such as a pharmacophore model.

Among the three methods used in this paper, the WMRA approach led to the most consistent superposition results: at low resolution, structures of the different compounds become similar, and superpositions did not present any strong inconsistencies.

An interesting perspective of the present work would consist in the comparative analysis of full promolecular and non-promolecular ED distributions at low resolution. Both kinds of ED distributions can be calculated using quantum mechanical formalisms with the same basis set and be smoothed using the same WMRA approach. Non-orthogonal wavelets will also be tested. Their use should facilitate the analysis of detail images which were not considered in the present work.

Acknowledgements

The author wishes to thank C.K. Johnson for providing the program ORCRIT, as well as S. Fortier, J.I. Glasgow, and F.H. Allen for fruitful discussions. The author also thanks D.P. Vercauteren, director of the Laboratoire de Physico-Chimie Informatique, for careful reading and discussion of the manuscript, J.-P. Antoine for discussing non-orthogonal wavelet applications, H. Chermette for useful advise on valence electron density maps, L. Piela for discussing the approach of analytical smoothing, and the FUNDP for the use of the Namur Scientific Computing Facility (SCF) Center, a common project between the FNRS, IBM-Belgium, and the FUNDP.

References

[1] J.M. Troyer and F.E. Cohen, Simplified models for understanding and predicting protein structure, in: *Reviews in Computational Chemistry*, Vol. 2, eds. K.B. Lipkowitz and D.B. Boyd (VCH Publishers, New York, 1990) pp. 57–80.

[2] J. Cherfils, S. Duquerroy and J. Janin, Protein–protein recognition analyzed by docking simulation, Proteins 11 (1991) 271–280.

[3] M.J.E. Sternberg, H.A. Gabb and R.M. Jackson, Predictive docking of protein–protein and protein–DNA complexes, Curr. Opin. Struct. Biol. 8 (1998) 250–256.

[4] I.A. Vakser, O.G. Matar and C.F. Lam, A systematic study of low-resolution recognition in protein-protein complexes, Proc. Natl. Acad. Sci. USA 96 (1999) 8477–8482.

[5] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo and I.A. Vakser, Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl. Acad. Sci. USA 89 (1992) 2195–2199.

[6] M. Carson, Wavelets and molecular structure, J. Comp.-Aided Mol. Des. 10 (1996) 273–283.

[7] D.W. Ritchie and G.J.L. Kemp, Fast computation, rotation and comparison of low resolution spherical harmonic molecular surfaces, J. Comput. Chem. 20 (1999) 383–395.

[8] B.S. Duncan and A.J. Olson, Shape analysis of molecular surfaces, Biopolymers 33 (1993) 231–238.

[9] M. Butzlaff, W. Dahmen, S. Diekmann, A. Dress, E. Schmitt and E. von Kitzing, A hierarchical approach to force field calculations through spline approximations, J. Math. Chem. 15 (1994) 77–92.

[10] E. von Kitzing and E. Schmitt, Configurational space of biological macromolecules as seen by semi-empirical force fields: Inherent problems for molecular design and strategies to solve them by means of hierarchical force fields, J. Mol. Struct. (Theochem) 336 (1995) 245–259.

[11] W.K. Olson, Simulating DNA at low resolution, Curr. Opin. Struct. Biol. 6 (1996) 242–256.

[12] Y. Takahashi, M. Sukekawa and S.-I. Sasaki, Automatic identification of molecular similarity using reduced-graph representation of chemical structure, J. Chem. Inf. Comput. Sci. 32 (1992) 639–643.

[13] T.S. Baker and J.E. Johnson, Low resolution meets high: Towards a resolution continuum from cells to atoms, Curr. Opin. Struct. Biol. 6 (1996) 585–594.

[14] J.-L. Stark, F. Murtagh and A. Bijaoui, *Image Processing and Data Analysis – The Multiscale Approach* (Cambridge University Press, Cambridge, UK, 1997).

[15] M. Pilard and Y. Epelboin, Multiresolution analysis for the restoration of noisy X-ray topographs, J. Appl. Cryst. 31 (1998) 36–46.

[16] J.I. Glasgow, S. Fortier and F.H. Allen, Molecular scene analysis: crystal and molecular structure determination through imagery, in: *Artificial Intelligence and Molecular Biology*, ed. L. Hunter (AAAI Press, Menlo Park, CA, 1993) pp. 433–458.

[17] K. Baxter, E. Steeg, R. Lathrop, J. Glasgow and S. Fortier, From electron density and sequence to structure: Integrating protein image analysis and threading for structure determination, in: *Proc. 4th International Conference on Intelligent Systems for Molecular Biology*, eds. D.J. States, P. Agarwal, T. Gaasterland, L. Hunter and R.F. Smith (AAAI Press, Menlo Park, CA, 1996) pp. 25–33.

[18] L. Leherte, J. Glasgow, K. Baxter, E. Steeg and S. Fortier, Analysis of three-dimensional protein images, J. Artificial Intelligence Research 7 (1997) 125–159.

[19] K. Edgecombe, A. Ableson, K. Baxter, A. Chiverton, J. Glasgow and S. Fortier, Protein model determination from crystallographic data, in: *Proceedings of the Pacific Symposium on Biocomputing'98*, eds. R. Altman, A.K. Dunker, L. Hunter and T.E. Klein (1998).

[20] A. Ableson and J.I. Glasgow, Crystallographic threading, in: *Proc. 7th International Conference on Intelligent Systems for Molecular Biology*, eds. T. Lengauer, R. Schneider, P. Bork, D. Brutlag, J. Glasgow, H.-W. Mewes and R. Zimmer (AAAI Press, Menlo Park, CA, 1999) pp. 2–9.

[21] J. Mestres, D.C. Rohrer and G.M. Maggiora, MIMIC: A molecular-field matching program. Exploiting applicability of molecular similarity approaches, J. Comput. Chem. 18 (1997) 934–954.

[22] R. Carbó, B. Calabuig, L. Vera and E. Besalú, Molecular quantum similarity: Theoretical framework, ordering principles, and visualization techniques, Adv. Quantum Chem. 25 (1994) 253–313.

[23] P.G. Mezey, Three-dimensional topological aspects of molecular similarity, in: *Concepts and Applications of Molecular Similarity*, eds. M.A. Johnson and G.M. Maggiora (Wiley, New York, 1990) pp. 321–368.
[24] P.G. Mezey, The shape of molecular charge distribution: group theory without symmetry, J. Comput. Chem. 8 (1987) 462–469.
[25] G.A. Arteca and P.G. Mezey, Shape group studies of molecular similarity and regioselectivity in chemical reactions, J. Comput. Chem. 9 (1988) 608–619.
[26] R.W. Bader, *Atoms in Molecules – A Quantum Theory*, 2nd ed. (Clarendon Press, Oxford, 1995).
[27] P.L.A. Popelier, Quantum molecular similarity. 1. BCP space, J. Phys. Chem. A 103 (1999) 2883–2890.
[28] C.K. Johnson, ORCRIT. The Oak Ridge critical point network program, Technical report, Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, TN (1977).
[29] R.N. Shirsat, S.V. Bapat and S.R. Gadre, Molecular electrostatics. A comprehensive topographical approach, Chem. Phys. Lett. 200 (1992), 373–378.
[30] S.R. Gadre and S.S. Pundlik, Topographical analysis of electron density and molecular electrostatic potential for cyclopropa- and cyclobutabenzenes, J. Am. Chem. Soc. 117 (1995) 9559–9563.
[31] D.J. Wild and P. Willett, Similarity searching in files of three-dimensional chemical structures: Alignment of molecular electrostatic potential fields with a genetic algorithm, J. Chem. Inf. Comput. Sci. 36 (1996) 159–167.
[32] D.A. Thorner, D.J. Wild, P. Willett and P.M. Wright, Similarity searching in files of three-dimensional chemical structures: Flexible field-based searching of molecular electrostatic potentials, J. Chem. Inf. Comput. Sci. 36 (1996) 900–908.
[33] M. Leboeuf, A.M. Köster, K. Jug and D.R. Salahub, Topological analysis of the molecular electrostatic potential, J. Chem. Phys. 111 (1999) 4893–4905.
[34] L. Leherte and F.H. Allen, Shape information from a critical point analysis of calculated electron density maps: Application to DNA–drug systems, J. Comput.-Aided Mol. Des. 8 (1994) 257–272.
[35] L. Leherte, S. Fortier, J. Glasgow and F.H. Allen, Molecular scene analysis: application of a topological approach to the automated interpretation of protein electron density maps, Acta Cryst. Sec. D 50 (1994) 155–166.
[36] L. Leherte and D.P. Vercauteren, Critical point analysis of calculated electron density maps at medium resolution: Application to shape analysis of zeolite-like systems, J. Molec. Model. 3 (1997) 156–171.
[37] *XTAL 3.0 User's Manual*, eds. S.R. Hall and J.M. Stewart (Universities of Western Australia and Maryland, 1990).
[38] L. Leherte, N. Meurice and D.P. Vercauteren, Critical point representation of electron density maps for the comparison of benzodiazepine-type ligands, J. Chem. Inf. Comput. Sci. 40 (2000) 816–832.
[39] L. Leherte, Th. Latour and D.P. Vercauteren, Topological analysis of electron density maps of chiral cyclodextrin–guest complexes: A steric interaction evaluation, Supramolecular Sci. 2 (1995) 209–217.
[40] L. Leherte, N. Meurice and D.P. Vercauteren, Application of multiresolution analyses to electron density maps: A critical point analysis approach for the comparison of molecules, in: *Mathematics and Computers in Modern Science – Acoustics and Music, Biology and Chemistry, Business and Economics*, ed. N. Mastorakis (World Scientific Engineering Society, 2000) pp. 158–164.
[41] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle and

J.A. Pople, Gaussian 94, Revision A.7 (Gaussian Inc., Pittsburgh, PA, 1994).
[42] L. Amat and R. Carbó-Dorca, Quantum similarity measures under atomic shell approximation: First order density fitting using elementary Jacobi rotations, J. Comput. Chem. 19 (1997) 2023–2039.
[43] J. Kostrowicki, L. Piela, B.J. Cherayil and H.A. Scheraga, Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard-Jones atoms, J. Phys. Chem. 95 (1991) 4113–4119.
[44] A.K.-M. Leung, F.-T. Chau and J.-B. Gao, A review on applications of wavelet transform techniques in chemical analysis: 1989–1997, Chemometrics and Intelligent Laboratory Systems 43 (1998) 165–184.
[45] P. Fischer and M. Defranceschi, The wavelet transform: a new mathematical tool for quantum chemistry, in: *Conceptual Trends in Quantum Chemistry*, eds. E.S. Kryachko and J.-L. Calais (Kluwer Academic, Dordrecht, 1994) pp. 227–247.
[46] J.-L. Calais, Wavelets – something for quantum chemistry?, Int. J. Quant. Chem. 58 (1996) 541–548.
[47] T.A. Arias, Multiresolution analysis of electronic structure: Semicardinal and wavelet bases, Rev. Mod. Phys. 71 (1999) 267–311.
[48] C.K. Chui, *Wavelet Analysis and its Applications*, Vol. 1, *An Introduction to Wavelets* (Academic Press, San Diego, CA, USA, 1992).
[49] G. Strang and T. Nguyen, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, Wellesley, MA, USA, 1997).
[50] I. Daubechies, Orthonormal bases of compactly supported wavelets, Commun. Pure Appl. Math. 41 (1988) 909–996.
[51] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Machine Intell. 11 (1989) 674–693.
[52] M. Bourges-Sévenier, WaveLib 1.1 – a C and Matlab library of wavelet functions, Réalisation d'une bibliothèque C de fonctions ondelettes, Publication interne no. 864, IRISA, Rennes, France (1994).
[53] B. Jawerth and W. Sweldens, An overview of wavelet based multiresolution analyses, Technical Report, University of South Carolina, USA (1993), ftp://maxwell.math.scarolina.edu:/pub/wavelet/papers/varia/sirev-36-3.tex.
[54] O.M. Nielsen, Wavelets in scientific computing, Ph.D. thesis, Technical University of Denmark, Lyngby (1998), http://www.imm.dtu.dk/∼omni/thesis.ps.gz.
[55] E.J. Stollnitz, T.D. DeRose and D.H. Salesin, Wavelets for computer graphics: a primer, part 1, IEEE Computer Graphics and Applications 15 (1995) 76–84.
[56] WWW site http://iqc.udg.es/cat/similarity/ASA/funcset.html.
[57] X. Gironés, L. Amat and R. Carbó-Dorca, A comparative study of isodensity surfaces using ab initio and ASA density functions, J. Mol. Graphics Model. 16 (1998) 190–196.
[58] J.W.M. Nissink, M.L. Verdonk, J. Kroon, T. Mietzner and G. Klebe, Superposition of molecules: Electron density fitting by application of Fourier transforms, J. Comput. Chem. 18 (1997) 638–645.
[59] H. Brandstetter, D. Turk, H.W. Hoeffken, D. Grosse, J. Stürzebecher, P.D. Martin, B.F.P. Edwards and W. Bode, Refined 2.3 angstroms X-ray crystal structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA: A starting point for improving antithrombotics, J. Mol. Biol. 226 (1992) 1085–1099.
[60] F.C. Berstein, T.F. Koetzle, G.J. Williams, E.E. Meyer, Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanovishi and M. Tasumi, The Protein Data Bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535.
[61] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne, The Protein Data Bank, Nucleic Acids Research 28 (2000) 235–242, http://www.rcsb.org/pdb/.
[62] Insight II, Version 4.0.0 (MSI, San Diego, 1998).
[63] IBM Visualization Data Explorer (IBM Corporation, 1996).
[64] M. Hô, H. Schmider, K.E. Edgecombe and V.H. Smith, Jr., Topological analysis of valence electron

charge distributions from semiempirical and ab initio methods, Int. J. Quant. Chem. Quant. Chem. Symp. 28 (1994) 215–226.

[65] E. Steiner, *The Chemistry Maths Book* (Oxford Science Publications, Oxford, 1998).

[66] S. Mallat and S. Zhong, Wavelet maxima representation, in: *Wavelet and Applications*, ed. Y. Meyer (Masson, Paris, 1991) pp. 207–284.

[67] S. Mallat and W.L. Hwang, Singularity detection and processing with wavelets, IEEE Trans. Inform. Theory 38 (1992) 617–643.

[68] M.J. Shensa, The discrete wavelet transform: Wedding the A Trous and Mallat algorithms, IEEE Trans. Signal Process. 40 (1992) 2464–2482.

[69] The molecular graphics program XMol, Version 1.3.1 (Minnesota Supercomputer Center, Minneapolis, MN, 1993), ftp://ftp.msc.edu/pub/xmol.

[70] F. Buyck, Analyse topologique de cartes de densité électronique de ligands pour les récepteurs benzodiazépiniques centraux, M.Sc. thesis, Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium (1994).

[71] N. Meurice, Développement d'un algorithme génétique original pour l'alignement de graphes moléculaires issus de la densité électronique. Application à l'étude comparative de composés d'intérêt pharmacologique, Ph.D. thesis, Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium (2000).

[72] D.J. Heisterberg, Technical report, Ohio Supercomputer Center, Columbus, OH; translation from FORTRAN to C and Input/Output by J. Labanowski, Ohio Supercomputer Center, Columbus, OH (1990).