Critical Point Representations of Electron Density Maps for the Comparison of Benzodiazepine-Type Ligands

Laurence Leherte,* Nathalie Meurice,† 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 September 9, 1999

A procedure for the comparison of three-dimensional electron density distributions is proposed for similarity searches between pharmacological ligands at various levels of crystallographic resolution. First, a graph representation of molecular electron density distributions is generated using a critical point analysis approach. Pairwise as well as multiple comparisons between the obtained graphs of critical points are then carried out using a Monte Carlo/simulated annealing technique, and results are compared with genetic algorithm solutions.

1. INTRODUCTION

The conception of new potent lead compounds relies on the knowledge of the interaction mode between the ligands and their receptor sites. Unfortunately, very often, the direct study of ligand−receptor binding modes through experimentation remains difficult. Also, direct evaluations of steric, electronic, and hydrophobic interactions are hard to tackle when the three-dimensional (3D) structure of the receptor is unavailable. Therefore the investigation of molecular binding to a given macromolecule must be achieved using alternative methods which usually lead to the description of a pharmacophore model. A pharmacophore is a set of geometrical, steric, and/or electronic features which are required to justify the activity and/or the affinity of a set of molecules presenting a common pharmacological behavior with respect to a common receptor. The design of such a model usually results from the determination of similarities between the molecular structures of several ligands which can be closely related in terms of chemical functions, but also very often relatively different. The most basic comparison approaches usually require the knowledge of the atom types and their connectivity. In this sense, the use of 2D representations overcomes the more time-consuming adjustment procedures of relative molecular orientations,[1,2] especially when matching 3D grids of points, i.e., electron density distribution functions or molecular electrostatic potential maps, is required.

Electron density (ED) and molecular electrostatic potential (MEP) are seen as properties of prime importance in the description of noncovalent interactions of a drug molecule with a receptor. Consequently, molecular similarity evaluation approaches are generally based on refined molecular field matching techniques. For example, several similarity quantifiers applicable to steric volume and electrostatic fields are implemented in the program MIMIC.[3,4] This program calculates the so-called Carbó index,[5] one of the most widely used molecular similarity measurements between various molecules. This index was originally dedicated to the comparison of molecular ED distributions. More generally, it can be applied to compare any grid property P between two prealigned molecules A and B:

$$C_{AB} = \frac{\sum_i P_A^i P_B^i}{[\sum_i (P_A^i)^2]^{1/2} [\sum_i (P_B^i)^2]^{1/2}} \quad (1)$$

where the summations extend over all grid points i. While C_{AB} is sensitive to the shape of the property P, a Gaussian approximation of C_{AB},[6] the Hodgkin index H_{AB}, is known to be more sensitive to the magnitude of P:

$$H_{AB} = \frac{2\sum_i P_A^i P_B^i}{\sum_i (P_A^i)^2 + \sum_i (P_B^i)^2} \quad (2)$$

The need of an alignment procedure can be overcome by avoiding any explicit reference to 3D atomic coordinates or to grid points or by using a reduction of the number of representation data. These two aspects are considered in topological analysis techniques. In this context, a reference work was achieved by Mezey[7] through his shape group method which allows the characterization of molecular surfaces (i.e., van der Waals, density or MEP isocontours, etc.) using curvature information. The local curvature properties of such contour surfaces are characterized by the eigenvalues of 2×2 Hessian matrixes $\mathbf{H}(\vec{r})$ defined at each point $\vec{r}$ on the contour surface within local tangential planes. Points with zero, one, and two negative eigenvalues of $\mathbf{H}(\vec{r})$ belong to domains that are locally concave, of saddle type, and convex, respectively. A molecular surface is thus partitioned into domains whose specific arrangement is described in terms of homology groups characterized by their ranks, or Betti numbers.[8,9] Domain decompositions of Connolly surfaces in terms of convex, saddle-shaped, and concave domains were also carried out by Lin et al.[10] In their method, the authors reduce each domain of the solvent

* Corresponding author. Telephone: +32-81-724560. Fax: +32-81-724530. E-mail: firstname.surname@fundp.ac.be.

† FRIA Ph.D. Fellowship.

10.1021/ci990112d CCC: $19.00
Published on Web 02/15/2000

accessible surface to a critical point determined as the projection of the gravity center of the domain onto the molecular surface. Another aspect of the topological approaches lies in the 3D analysis of ED distributions and MEP maps. According to Bader,[11] the topological properties of ED distributions can be described in terms of the number and kind of their critical points, i.e., points where the gradient of the density is equal to zero. Each critical point can be identified by its corresponding 3×3 Hessian matrix $\mathbf{H}(\vec{r})$ 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 critical points are linked through a gradient vector field analysis to generate a graph whose vertexes and edges are critical points and gradient trajectories, respectively. Similarly to Bader's approach, Johnson[12] developed a critical point analysis method, based on Morse theory, for the location, identification, and connection of critical point 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 studies.[13,14] In a brute force approach, the locations and connections of extrema in MEP were carried out by Willett and co-workers[15,16] in order to generate the so-called field graphs. In their approach, grid points with MEP values ranging below or beyond given cutoff values are merged into single points which are further connected depending on their closeness. Such graph representations facilitate the alignment of MEP fields using genetic algorithms. This is particularly important for a similarity search in large databases where speed and efficiency are always simultaneously required.

In this work, we present a theoretical approach for molecular comparison of electronic properties of benzodiazepines and related compounds at various levels of the structural hierarchy. Molecular structures are represented using critical point graphs obtained using a topological analysis procedure of their ED distribution functions.

In section 2, we describe the set of molecular structures that was considered for developing our approach. We further emphasize, in section 3, the critical point analysis approach and the concept of crystallographic resolution of an ED map. Results from the topological analyses and details relative to the Monte Carlo/simulated annealing and genetic algorithm procedures allowing pairwise or simultaneous comparisons of critical point graphs are presented in section 4.

2. MATERIALS

Benzodiazepines and related compounds form an important family of molecules with various pharmacological effects. They are known to interact with the so-called benzodiazepine receptors (BZR) which include two main classes named according to their tissue distribution: central receptors (CBR) coupled to the GABA-gated chloride channels in the central nervous system[17] and mitochondrial receptors (MBR) located in peripheral tissues such as kidney and heart and even in brain.[18] To our knowledge, both 3D receptor structures have not been solved through crystallographic techniques so far.

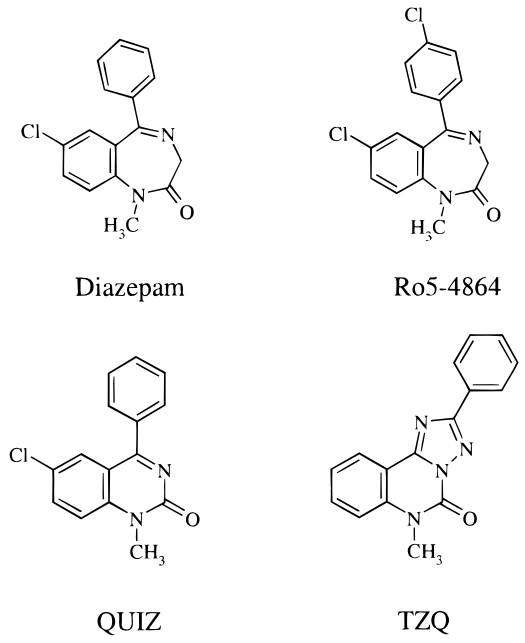


Figure 1. Planar structure formulas of Diazepam, Ro5-4864, QUIZ, and TZQ.

Many benzodiazepine-like molecules are commercially available as drugs used in the treatment of anxiety, depression, or insomnia. CBR ligands can functionally been classified as (i) agonists, i.e., positive modulators of the inhibition of the neuron activity which induce anxiolitic and sedative effects, (ii) inverse agonists, i.e., negative modulators which generate convulsant effects, and (iii) antagonists, i.e., which present binding but without any functional consequence. The MBR ligands are known to mediate the delivery of cholesterol to the inner mitochondrial membrane for further oxidation.[19] If we assume that all ligands bind to the same domain of the receptor protein, then they should present common characteristics that justify recognition and binding regardless of the type of activity.

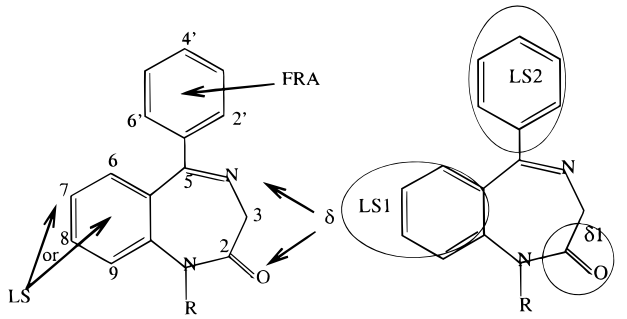
Interaction mechanisms between the guest molecules and their host receptors are still unclear, especially since some non-benzodiazepine compounds have been found to also show high affinities for the BZRs. Actually, the absence of any correlation between therapeutic properties and affinity values for MBRs makes the binding studies very difficult. The elaboration of a pharmacophore model is thus a major concern for scientists in this field of research.

In this study, we have selected a limited number of ligands, considering their affinities for both the CB and MB 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-N1-methyl-4-phenylquinazolin-2-one), and TZQ ([1,2,4]triazolo[1,5-*c*]quinazolin-5(6*H*)-one) (Figure 1).

Ligands are often characterized by an affinity value, IC_{50}, which is a measurement of the ligand concentration required to inhibit 50% of the binding between an isotopically labeled ligand and its receptor. These values are reported in Table 1 for the four chosen compounds.[20] A study of MBR ligands[20] also shows that compound Ro5-4864 is about 10 times more affine for MBRs than Diazepam. This result is, at first, rather surprising considering that Ro5-4864 differs from Diazepam only by a chlorine atom at position 4′ (numbering of the

Table 1. IC_{50} Values Relative to the Binding of Diazepam, Ro5-4864, TZQ, and QUIZ (nM) to CBRs and MBRs[20]

	CBR	MBR
Diazepam	6	80
Ro5-4864	5000	6
TZQ	non-significant	2800
QUIZ	–	–



(a) (b)

Figure 2. Schematic representation of (a) the most frequently described pharmacophore features common to BZR ligands: LS = lipophilic and/or steric group, δ = proton acceptor group, FRA = freely rotating aromatic group. (b) The benzodiazepine pharmacophore determined using the genetic algorithm superposition results based on the critical point representation of electron density distributions. The three pharmacophore elements are superimposed on a benzodiazepin-2-one.

atoms is shown in Figure 2a). General structure−affinity relationships have been deduced from binding experiments, which show that the removal of the methyl group at position R reduces the affinity toward MBRs. However, as reported by Colotta et al.,[21] group R appears to be important for potent inverse agonist activity but is not a requirement for high in vitro affinity. The absence of a Cl atom at position 7, the replacement of the imine by a secondary amine group, or the change of the Cl atom from position 4′ to position 2′ also induces a decrease in affinity. Ro5-4864 is affine and selective to MBRs, while Diazepam, an affine molecule as well, is not selective to either of both receptors. Compound TZQ is not a benzodiazepine compound but presents a significant affinity for the MBRs. Compound QUIZ is by now synthesized, but its affinity values are still under measurement.

Several pharmacophore models have already been published.[21−28] Besides the presence of a rigid planar region constituted by the benzo rings, all models possess each of the following features or a selected combination of them: (a) a proton acceptor region (δ); (b) a lipophilic or steric group (LS), i.e., the benzo ring and/or its chlorine atom; (c) a free rotating aromatic region, i.e., a phenyl group (FRA) (Figure 2a).

3. METHODS

In this section, we detail the methodology used to generate critical point representations of the ligands at various levels of resolution. In a first step, electron density maps (EDMs) are calculated at a given resolution value using crystallography relationships. Then the corresponding critical point representations are obtained using a topological analysis approach which characterizes the critical points using density magnitude and curvature data. The connectivity pattern between the points is defined using Euclidean distances, density values, and density gradient vectors.

The three-dimensional (3D) atomic coordinates of the considered ligands are derived from crystallographic data. Two structures are reported in the Cambridge Crystallographic Database:[29] Diazepam (DIZPAM10) and Ro5-4864 (FULWUE). Atom coordinates of molecules TZQ and QUIZ were obtained from X-ray diffraction experiments in our institution.[30] Each molecule is embedded in a 16 × 8 × 16 Å box (grid interval is set equal to 0.25 Å). This guarantees that packing effects, which do not occur under biological conditions, do not affect the critical point representation of individual molecules. EDMs are then calculated using the crystallographic package XTAL.[31]

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 $F(\vec{h})$:

$$F(\vec{h}) = \sum_{j=1}^{\text{nat}} f_j e^{-B((\sin\theta)/\lambda)^2} e^{-2\pi i \vec{h}\cdot\vec{r}_j} \tag{3}$$

where f_j and B_j are the atomic form factor and the thermal agitation factor of atom j, respectively. 2θ is the angle between the diffracted X-ray and the primary beam of wavelength λ, and $\vec{h}$ is a reciprocal space vector.

The ED distribution function is calculated as the Fourier transform of $F(\vec{h})$:

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

In practice, the number of known structure factors is not infinite and varies with resolution. If the structure factors are known in both modulus and phase, the atomic positions are unequivocally determined using a maxima detection algorithm of the ED function $\rho(\vec{r})$.

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 chemical composition, the radiation used, and the temperature of the experiment.

The resolution values that were selected to calculate EDMs are 2.5 and 3.0 Å following previous works on biological macromolecules, i.e., DNA[32] and proteins,[33] which showed that corresponding chemical groups (i.e., phosphate, ribose, and amino acid residues) reduce to uniquely defined *peaks*.

Critical Point Analysis of Electron Density Maps. In this subsection, we overview a methodology that transforms a 3D EDM into a graph of critical points.

By first locating and then connecting critical points, the procedure generates a graph representation which captures the skeleton and the volumetric features of a molecular ED distribution function. The critical point approach is based on the representation of an image in terms of the critical points of the ED function, i.e., the points where the gradient

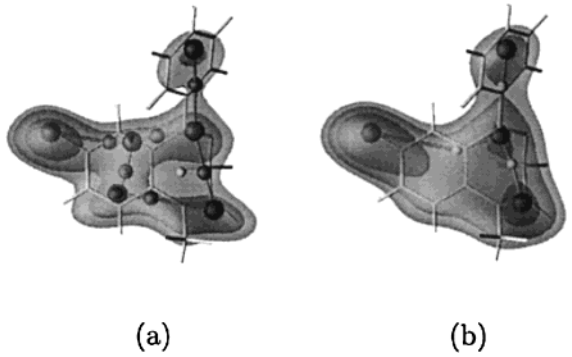


Figure 3. Isocontours of the electron density distribution of Diazepam calculated using XTAL at a resolution of (a) 2.5 Å (contours: 1.5, 2.0, 3.0 $e^-/Å^3$) and (b) 3.0 Å (contours: 1.2, 1.5, 1.9 $e^-/Å^3$). Critical point graph representations are obtained from a topological analysis using ORCRIT: (a) peaks, large spheres; passes, medium spheres; pale, small sphere; and (b) peaks, large spheres; passes, small spheres. Figures were generated using DataExplorer.[53]

of $\rho(\vec{r})$ is zero. For each critical point, the corresponding Hessian matrix $\mathbf{H}(\vec{r})$ is constructed and then diagonalized:

$$\mathbf{H}(\vec{r}) = \begin{pmatrix} \partial^2\rho/\partial x^2 & \partial^2\rho/\partial x\,\partial y & \partial^2\rho/\partial x\,\partial z \\ \partial^2\rho/\partial y\,\partial x & \partial^2\rho/\partial y^2 & \partial^2\rho/\partial y\,\partial z \\ \partial^2\rho/\partial z\,\partial x & \partial^2\rho/\partial z\,\partial y & \partial^2\rho/\partial z^2 \end{pmatrix} \qquad (6)$$

The three non-zero diagonal elements of the diagonalized Hessian matrix, i.e., the eigenvalues, are used to determine the type of critical point in the EDM. Four possible cases are considered, depending on the number of negative eigenvalues, n_E. When $n_E = 3$, the critical point corresponds to a local maximum or *peak*; a point where $n_E = 2$ is a saddle point or *pass*; $n_E = 1$ corresponds to another saddle point or *pale*, while $n_E = 0$ characterizes a *pit*.

A topological analysis algorithm has been implemented by Johnson in the computer program ORCRIT,[12] which originally generates minimal spanning tree representations, i.e., connected acyclic graphs of minimal weight. In this work, we allow the presence of cycles in the critical point connectivity patterns. The only restriction that is imposed is that a link cannot occur between two critical points which differ, in sign, by more than one eigenvalue.[34]

As detailed in Bader's book,[11] the analysis of an EDM shows peaks and passes at atom positions and chemical bonds, respectively. A molecular structure at atomic resolution is thus visualized as a graph of alternating peaks and passes, with pales and pits located in rings and cages, respectively. At lower resolutions, results similar to those obtained for biological macromolecules are expected, i.e., critical points become representative of groups of atoms rather than the atoms themselves.

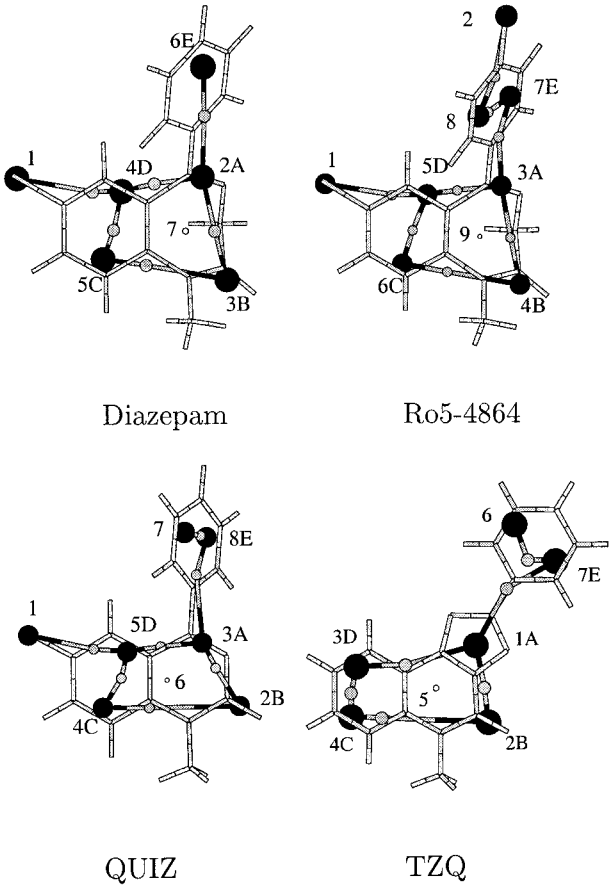


Figure 4. Critical point graph representations obtained from topological analyses of 2.5 Å resolution maps of the electron density using ORCRIT (peaks, black spheres; passes, gray spheres; pale, white sphere), superimposed on the corresponding 3D molecular structure (sticks). Peaks and pales are numbered according to their density value.

The method, described with more details in a previous work,[34] was applied to EDMs generated using the XTAL crystallographic package.

4. RESULTS

Critical Point Representations. In this section, we first describe the critical point graph representations that were obtained from the analysis of electron density maps (EDMs) generated at resolutions of 2.5 and 3.0 Å using the program ORCRIT.

At 2.5 Å resolution, all critical points characterized by a density value higher than 1.5 $e^-/Å^3$ were considered as

Table 2. Geometrical Parameters (deg) and Density Values ($e^-/Å^3$) of the Critical Point Graph Representations Obtained from a Topological Analysis of 2.5 Å Resolution Maps of the Electron Density Using ORCRIT[a]

	Diazepam	Ro5-4864	TZQ	QUIZ
cycle torsion angle (A−B−C−D)	2.7	1.1	359.9	0.17
FRA-cycle torsion angle (E−A−D−C)	207.6	202.7	182.0	170.3
Cl density values	4.40	4.24−4.79	−	4.43
N-group density values	2.68	2.89	3.36	2.80
C═O density values	2.55	2.85	3.13	3.06
C density values	2.11−2.47	1.98−2.66	2.10−2.59	1.90−2.71
pass density values	1.58−2.40	1.13−2.55	2.07−2.68	1.70−2.70
pale density value	1.63	1.50	2.23	2.36

[a] Peak numbering is presented in Figure 4.

Table 3. Geometrical Parameters (deg) and Density Values ($e^-/Å^3$) of the Critical Point Graph Representations Obtained from a Topological Analysis of 3.0 Å Resolution Maps of the Electron Density Using ORCRIT[a]

	Diazepam	Ro5-4864	TZQ	QUIZ
B−D−C angle	148	142	151	−
A density	2.81	2.78	2.20	2.86
B density	2.07	1.94	2.33	2.43
C density	1.78	2.41	2.01	1.79
D density	1.97	2.00	2.42	−
pass density	1.50−1.88	1.53−1.93	1.61−2.30	1.51−2.01

[a] Peak numbering is presented in Figure 6.

significant of the molecular structure. Below that value, the background noise due to the termination error in the Fourier transform process leads to an excess of unphysical critical points. At a resolution of 2.5 Å, local maxima in the ED distribution do not reflect atom positions, except for the chlorine atoms which are at the origin of peaks. This is illustrated in Figure 3a, where three isodensity contours (1.5, 2.0, and 3.0 $e^-/Å^3$) of the EDM of Diazepam are represented together with the corresponding critical point graph (all other EDMs are provided as Supporting Information). The six carbon atoms of the FRA group form a high-density center composed of one or two peaks, and the benzodiazepine group is represented by one cycle composed of four distinct peaks connected through passes. The low-density area located in the center of this cycle is at the origin of a pale. Critical point graphs are presented in Figure 4 for the four molecules under study. The peak−pass sequences show that all molecules adopt a topological structure with common features. The four-peak cycle is formed by local maxima that are localized (a) at the level of the imine group (peak A), (b) in the region of the carbonyl function (peak B), and (c) on the benzo ring (peaks C and D). A lateral branch appears at the level of the chlorine substituant at position 7, except for compound TZQ which does not contain any Cl atom. Besides these common characteristics, a difference appears in terms of the number of peaks localized on the FRA group. This group in all compounds, except Diazepam, is characterized by two peaks separated by a distance of 1.76 (Ro5-4864), 1.35 (QUIZ), and 1.88 Å (TZQ). These values are in the limit of detection in a 2.5 Å resolution map. Each pair of close peaks is actually merged by decreasing only slightly the resolution value from 2.5 to 2.6 Å.

The analysis of the peak density values shows that, as expected, the highest values are associated with the Cl atoms,

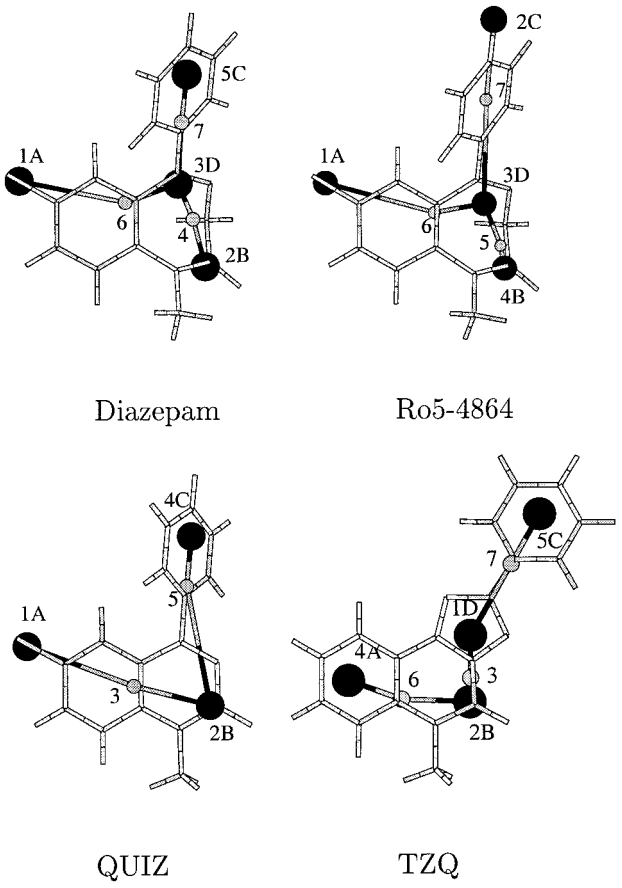


Figure 5. Critical point graph representations obtained from topological analyses of 3.0 Å resolution maps of the electron density using ORCRIT (peaks, black spheres; passes, gray spheres), superimposed on the corresponding 3D molecular structure (sticks). Peaks and passes are numbered according to their density value.

followed by the imine and carbonyl groups, and the lowest peaks are those originating from the C atoms (Figure 4 and Table 2). From the observations reported in Table 2, it is seen that the four-peak cycle is almost planar. The torsion angle of the cycle is equal to 0 ± 3 degrees. Only slight deviations are observed which are probably due to numeric approximations linked to the grid resolution (grid interval). Branches of non-benzodiazepine ligands are almost coplanar with the cycle, in agreement with the molecular geometry of these two molecules.

At the lower 3.0 Å resolution, the density cut-off value was lowered to 0.5 $e^-/Å^3$. The topological graphs obtained

Table 4. Matching Parameters Used for the Monte Carlo Comparison of the Critical Point Graph Representations Obtained from Topological Analyses of 2.5 Å Resolution Maps of the Electron Density Using ORCRIT[a]

set of graphs	no. of points	$1/\beta$	no. of trials per β value	no. of runs
		n Points		
Diazepam/Ro5-4864	7	0.350	10 000 (1.03)	1
QUIZ/Ro5-4864	8	0.350	10 000 (1.18)	1
TZQ/Ro5-4864	7	0.350	10 000 (0.89)	1
Diazepam/QUIZ/TZQ/Ro5-4864	7	0.050−0.500	50 000	11
		n − 1 Points		
Diazepam/Ro5-4864	6	0.350	10 000 (1.05)	1
QUIZ/Ro5-4864	7	0.350	10 000 (1.34)	1
TZQ/Ro5-4864	6	0.450	10 000 (0.79)	1
Diazepam/QUIZ/TZQ/Ro5-4864	6	0.050−0.500	100 000	11

[a] For each set of graphs, *n* or (*n* − 1) points of the smallest molecular graph were considered. The ratio of accepted over rejected trials is given in parentheses for single MC runs.

Table 5. Matching Parameters Used for the Monte Carlo Comparison of the Critical Point Graph Representations Obtained from Topological Analyses of 3.0 Å Resolution Maps of the Electron Density Using ORCRIT[a]

set of graphs	no. of points	$1/\beta$	no. of trials per β value	no. of runs
		n Points		
Diazepam/Ro5-4864	7	1.150	10 000 (0.96)	1
QUIZ/Ro5-4864	5	1.150	10 000 (0.71)	1
TZQ/Ro5-4864	7	1.150	10 000 (1.01)	1
Diazepam/QUIZ/TZQ/Ro5-4864	5	0.050−1.000	50 000	21
		$n-1$ Points		
Diazepam/Ro5-4864	6	1.150	10 000 (0.94)	1
QUIZ/Ro5-4864	4	1.150	10 000 (0.60)	1
TZQ/Ro5-4864	6	1.150	10 000 (1.03)	1
Diazepam/QUIZ/TZQ/Ro5-4864	4	0.050−1.000	50 000	21

[a] For each set of graphs, *n* or $(n-1)$ points of the smallest molecular graph were considered. The ratio of accepted over rejected trials is given in parentheses for single MC runs.

Table 6. Matching Parameters Used for the Genetic Algorithm Comparison of the Critical Point Graph Representations Obtained from Topological Analyses of 2.5 Å Resolution Maps of the Electron Density Using ORCRIT[a]

set of graphs	no. of points	no. of generations	population size	mutation rate	crossover rate	no. of runs
			n Points			
Diazepam/Ro5-4864	7	1 000	100	0.001	0.6	10
QUIZ/Ro5-4864	8	1 000	100	0.001	0.6	10
TZQ/Ro5-4864	7	1 000	100	0.001	0.6	10
Diazepam/QUIZ/TZQ/Ro5-4864	7	10 000	100, elitist	0.05	0.6	10
			$n-1$ Points			
Diazepam/Ro5-4864	6	1 000	100	0.001	0.6	10
QUIZ/Ro5-4864	7	1 000	100	0.001	0.6	10
TZQ/Ro5-4864	6	1 000	100	0.001	0.6	10
Diazepam/QUIZ/TZQ/Ro5-4864	6	10 000	100, elitist	0.10	0.55	10

[a] For each set of graphs, *n* or $(n-1)$ points of the smallest molecular graph were considered.

Table 7. Matching Parameters Used for the Genetic Algorithm Comparison of the Critical Point Graph Representations Obtained from Topological Analyses of 3.0 Å Resolution Maps of the Electron Density Using ORCRIT[a]

set of graphs	no. of points	no. of generations	population size	mutation rate	crossover rate	no. of runs
			n Points			
Diazepam/Ro5-4864	7	1 000	100	0.001	0.6	10
QUIZ/Ro5-4864	5	1 000	100	0.001	0.6	10
TZQ/Ro5-4864	7	1 000	100	0.001	0.6	10
Diazepam/QUIZ/TZQ/Ro5-4864	5	10 000	100, elitist	0.05	0.6	10
			$n-1$ Points			
Diazepam/Ro5-4864	6	1 000	100	0.001	0.6	10
QUIZ/Ro5-4864	4	1 000	100	0.001	0.6	10
TZQ/Ro5-4864	6	1 000	100	0.001	0.6	10
Diazepam/QUIZ/TZQ/Ro5-4864	4	10 000	200, elitist	0.05	0.6	10

[a] For each set of graphs, *n* or $(n-1)$ points of the smallest molecular graph were considered.

at this resolution (Figures 3b and 5 and Table 3) are composed of three local density maxima which are located close to the chlorine atoms or the benzo group (A), the carbonyl group (B), and the FRA group (C), and with the exception of compound QUIZ, a fourth peak is observed at the proximity of the imine group of the molecular structures (D). It is remarkable that all four peaks correspond to the pharmacophore elements that have been emphasized above (Figure 2a). Molecules that are capable of binding to benzodiazepine receptors thus present common topological features at various resolution levels with differences that are more likely to occur at high resolution.

We assume that the properties of these common features are important in determining the selectivity with respect to CBRs or MBRs. More particularly, at 3.0 Å resolution, Diazepam and Ro5-4864, which present a significant affinity toward CBRs and MBRs, are characterized by a similar topology. In addition, the experimental affinity data reported in Table 1 show that TZQ, which presents a different topology, is not recognized by CBRs and has a significant binding affinity for MBR receptors.

The comparison of similar molecular structures (i.e., with similar chemical functions) is obviously easily done by hand. However, several different matching possibilities may appear when the number of superposition constraints diminishes, for example with resolution or when the chemical structures strongly differ.

Critical Point Graphs Matching. In this work, we present two methods for pairwise matching of critical point graphs, which have been implemented in order to determine the topological features that are essential to an effective search of similarity between affine molecules and to the design of a pharmacophore. This is a different approach vs the one adopted by Hodgkin et al.[35] who used a Metropolis Monte

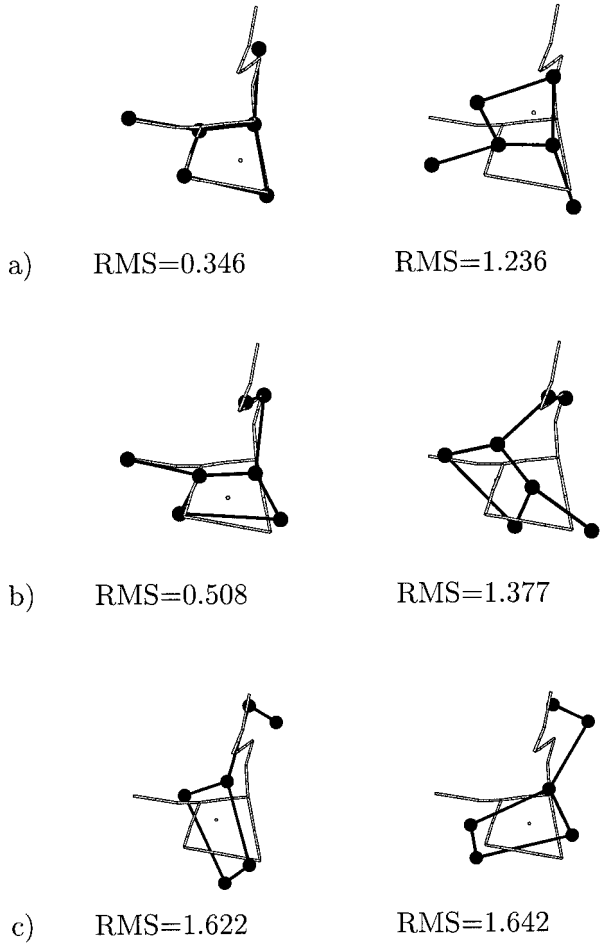


Figure 6. Best (left) and inverted or standard (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of *n*-point graphs obtained using a critical point analysis of 2.5 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecular graphs, (a) Diazepam, (b) QUIZ, and (c) TZQ, are shown using black (peaks) and white spheres (pales).

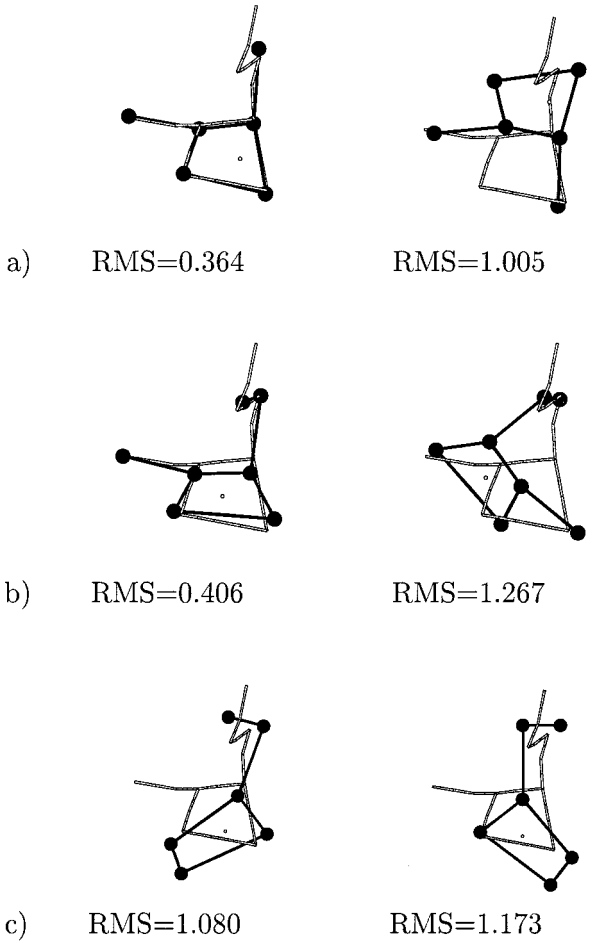


Figure 7. Best (left) and inverted (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of $(n-1)$-point graphs obtained using a critical point analysis of 2.5 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecular graphs, (a) Diazepam, (b) QUIZ, and (c) TZQ, are shown using black (peaks) and white spheres (pales).

Carlo procedure to overlay molecules constrained to a predefined pharmacophore model.

The first method is a Monte Carlo (MC) one. Such an algorithm is intended to generate a sequence of solutions satisfying a predetermined probability distribution function. Those solutions are selected among all possibilities using random numbers. During a Metropolis MC process, only sufficiently probable solutions are kept in order to favor convergence toward the selected probability distribution.[36]

The second approach is a genetic algorithm (GA) one. GAs are inspired by the Darwinian evolution wherein successive generations of populations (a population is a set of possible solutions) evolve through genetic operations (reproduction, mutation, crossover) in order to generate high-quality individuals (highly probable solutions).[37−39]

When matching two molecules, it is important to converge toward both local and global minima. The conformation of a drug molecule in the active state may indeed strongly differ from its minimal energy conformation in a vacuum. MC as well as GA strategies are thus both adequate for such multiple minima optimizations as they provide several possible solutions of high-occurrence probability. In their work, Paretti et al.[40] proposed a probability distribution p:

$$p = \exp(-C\Delta) \tag{7}$$

where Δ is the difference of Carbó similarity index between two successive MC alignments. C stands for a constant. The use of GAs to match 2D chemical graphs was presented by Brown et al.[41] In their work, molecular structures were represented in terms of graphs (atoms and number of bonds connected to each atom). Later, Jones et al.[42] also used a GA for flexible molecular overlay. The fitness function involved evaluation of (a) the vdW intramolecular energy of each molecule and (b) the common volume between each molecule and the reference molecule. Each molecule was represented by a number of rotatable bonds and chemical features (lone pair, H-donor group, etc.). The use of GAs for the superposition of 3D flexible chemical structures was further proposed by Handschuh et al.[43] Once atoms of a structure were matched against atoms of a reference molecule, the difference between the corresponding atom distances in both structures were minimized by changing torsion angle values.

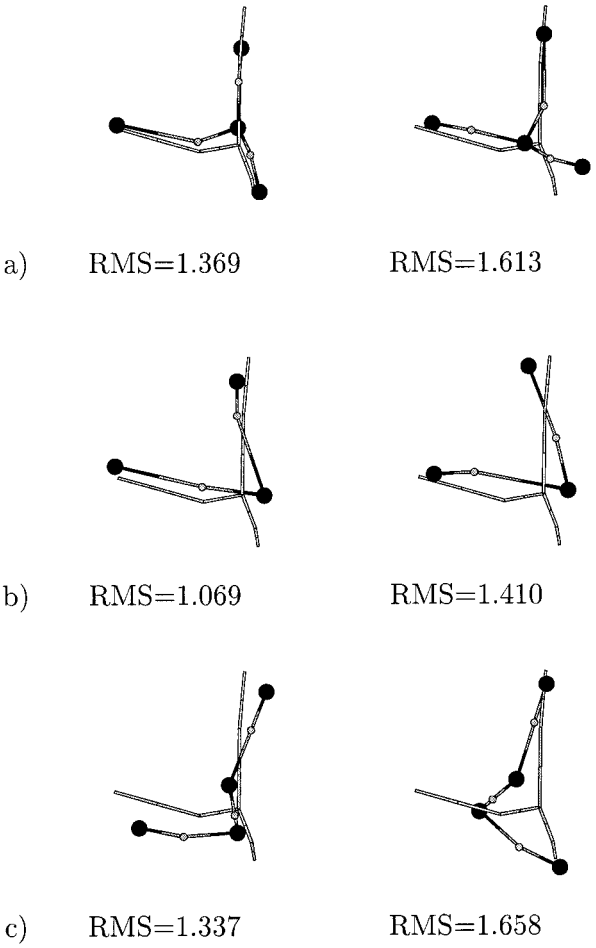


Figure 8. Best (left) and inverted (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of *n*-point graphs obtained using a critical point analysis of 3.0 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecular graphs, (a) Diazepam, (b) QUIZ, and (c) TZQ, are shown using black (peaks) and gray spheres (passes).

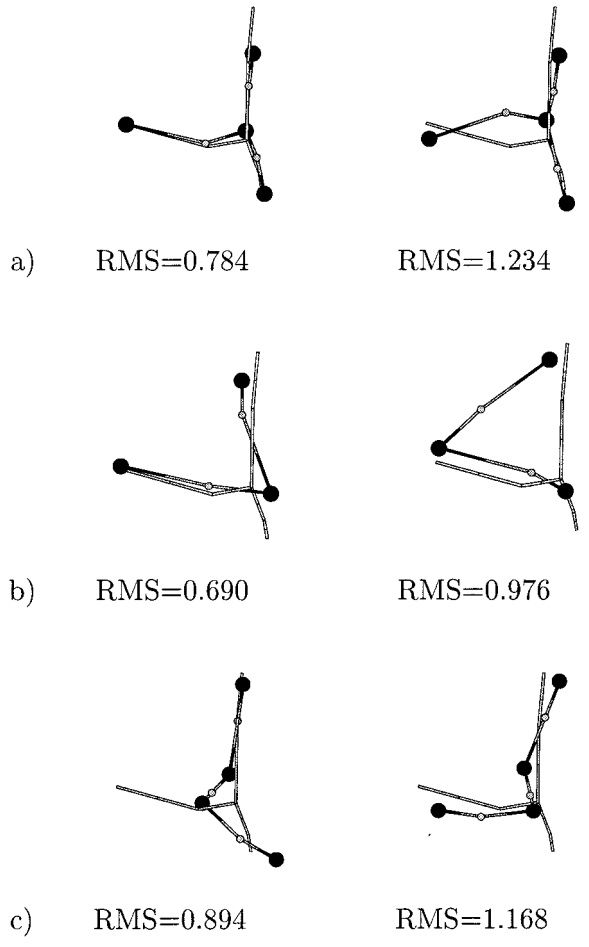


Figure 9. Best (left) and inverted or standard (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of $(n-1)$-point graphs obtained using a critical point analysis of 3.0 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecular graphs, (a) Diazepam, (b) QUIZ, and (c) TZQ, are shown using black (peaks) and gray spheres (passes).

The GA technique that was used in the present study was initially designed for the simultaneous comparison of several molecules.[44] In this original approach called GAGS (genetic algorithm for graph similarity search), the comparison is driven by the similarity between the critical point representations of all ligands, based on their peak heights and connectivity properties. Previous studies[44,45] showed that the obtained superposition results contain enough information to correctly align similar structures such as Diazepam and Ro5-4864. Interestingly, benzodiazepine-like compounds such as QUIZ and TZQ can either adopt the same standard orientation or an inverted one, wherein their FRA moiety corresponds to the benzo rings of the first two compounds. As a result, a general pharmacophore was described, unifying the optimal GA trends (Figure 2b). It is composed of three major elements: LS1 and LS2 standing for two lipophilic and/or steric regions which are bridged by a proton acceptor site, $\delta 1$. In the present study, both the GAGS coding scheme and the evaluation function were adapted in order to allow pairwise matching of critical point graph representations and hence the comparison with the more conventional pairwise MC approach.

(a) Representation Mode. For matching purposes, the original critical point graphs were converted to fully connected graphs. This conversion simplifies the superposition algorithm since the number of bonds connected to each point does not have to be checked.

In both MC and GA adopted coding schemes, a sequence (called chromosome in GAs) is defined as a 2D array where the first dimension is the number of considered structures (two for pairwise comparisons), and the second dimension is defined by the number of matching points. If *n* is the size of the smallest graph, the second dimension will be set to $(n-1)$ or *n*, depending on whether we choose or not a leave-one-out strategy. The sequences (or chromosomes) are then filled with integer numbers, being the labels of the involved critical points. In such a way, each sequence (or chromosome) is a potential correspondence between the critical points of the compared graphs, i.e., a hypothesis of subgraph matching.

(b) Evaluation Function. Each *n*-point graph *G* is thus in a first step represented by a 2D array, called "property matrix" $\mathbf{P}_G$, whose diagonal elements are the density values (in $e^-/Å^3$) of the critical points, and the off-diagonal elements

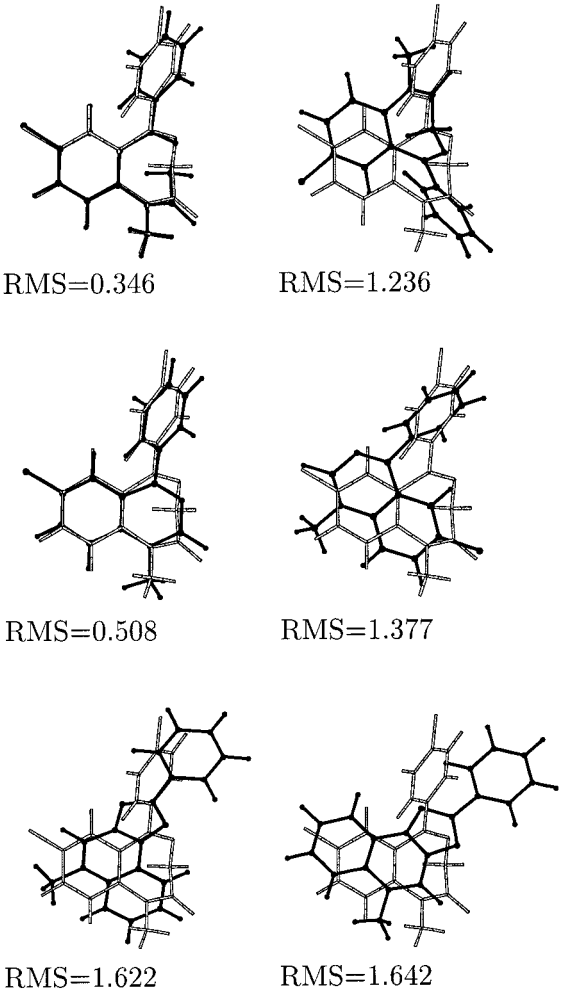


Figure 10. Best (left) and inverted or standard (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of *n*-point graphs obtained using a critical point analysis of 2.5 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecules, (a) Diazepam, (b) QUIZ, and (c) TZQ, are displayed using black sticks and balls.

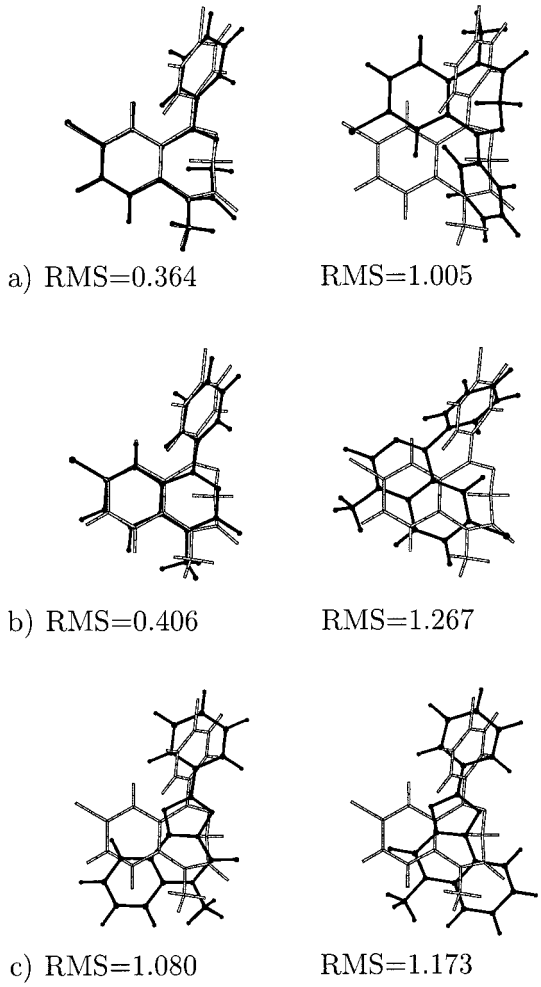


Figure 11. Best (left) and inverted (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of $(n-1)$-point graphs obtained using a critical point analysis of 2.5 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecules, (a) Diazepam, (b) QUIZ, and (c) TZQ, are displayed using black sticks and balls.

correspond to the distances (in Å) between these points:

$$\mathbf{P}_G = \begin{pmatrix} \rho_1 & d_{12} & \cdots & \cdots \\ d_{21} & \rho_2 & d_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \rho_n \end{pmatrix} \tag{8}$$

Several preliminary tests showed that the various kinds of critical points must be clearly differentiated to guarantee the efficiency of the algorithm. Initial peak heights were augmented by 10 units, while pale heights were decreased by the same amount, which forces the superposition of critical points of the same kind. Density values of the passes were not modified. Once critical points of the same type are matched, the bias introduced in the density values will have no effect in the calculation of the corresponding root mean square deviation (rms) value. We have considered the case where all critical points, i.e., peaks, passes, and pales are included in the critical point graphs. Superposition results showed that passes do not prevent improbable matchings, such as connected segments superposed on nonconnected parts of the reference molecule. Consequently, at a resolution of 2.5 Å, the critical points that must be explicitly taken into account are the peaks and pales. However, at 3.0 Å resolution, peaks and passes will be considered. Representations of the fully connected graphs and property matrices of the four compounds under study at 2.5 and 3.0 Å resolution are reported as Supporting Information. For instance, the final representation of Diazepam at a resolution of 2.5 Å contains seven critical points (six peaks, 1, 2A, 3B, 4D, 5C, 6E, and one pale, 7, in Figure 4).

In a second step, the quality of a sequence (or chromosome) needs to be evaluated; i.e., if the critical point matching hypothesis is valuable, then the difference between the corresponding properties is minimal. This is given by a rms value calculated according to

$$\mathrm{rms} = \left[\frac{1}{n \times n_\mathrm{p}} \sum_{i=1}^{n \times n_\mathrm{p}} (\rho_i - \rho_i^{\mathrm{ref}})^2 + \frac{1}{\mathrm{nbn} \times n_\mathrm{p}} \sum_{i=1}^{\mathrm{nbn} \times n_\mathrm{p}} (d_i - d_i^{\mathrm{ref}})^2 \right]^{1/2} \tag{9}$$

where, in a completely connected graph composed of *n*

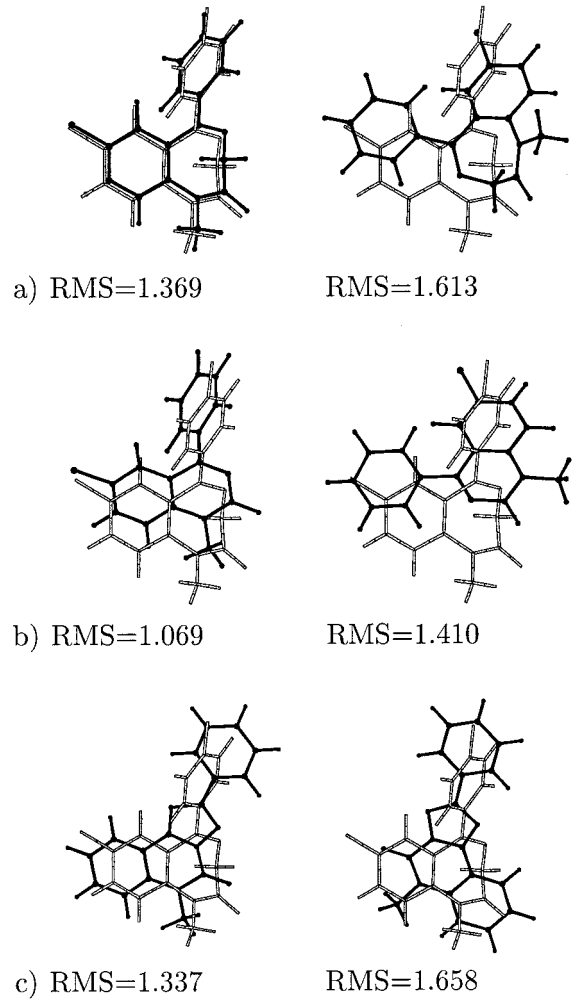


Figure 12. Best (left) and inverted (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of *n*-point graphs obtained using a critical point analysis of 3.0 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecules, (a) Diazepam, (b) QUIZ, and (c) TZQ, are displayed using black sticks and balls.

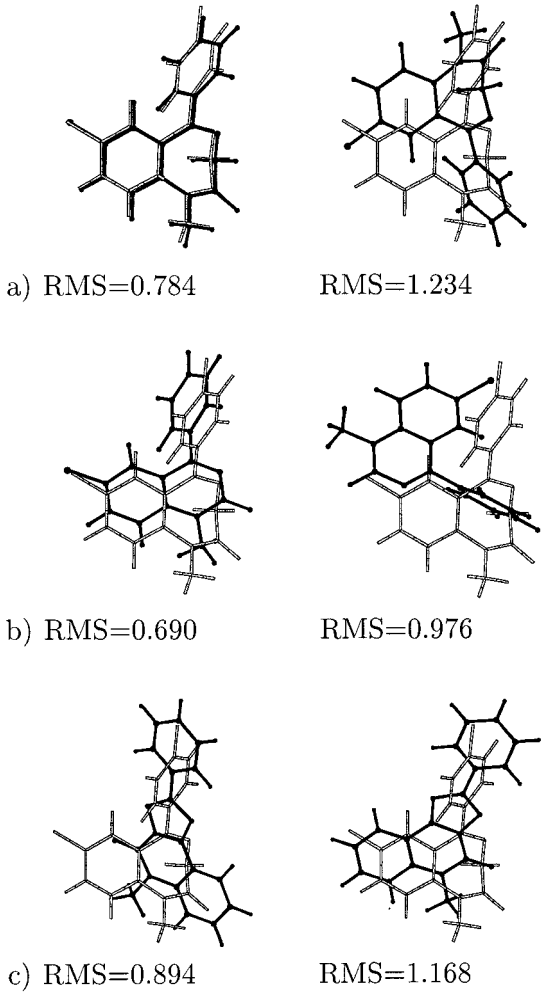


Figure 13. Best (left) and inverted or standard (right) solutions obtained by the Monte Carlo and genetic algorithm matching procedures for pairs of $(n-1)$-point graphs obtained using a critical point analysis of 3.0 Å resolution maps of the electron density. Compound Ro5-4864 is displayed using gray sticks. Fitted molecules, (a) Diazepam, (b) QUIZ, and (c) TZQ, are displayed using black sticks and balls.

points, the number of connections between the points is given by

$$\mathrm{nbn} = n(n-1)/2 \tag{10}$$

and n_p is the total number of pairs of critical point graphs that can be built from the n_s molecules to be compared:

$$n_p = n_s(n_s-1)/2 \tag{11}$$

The GA approach directly minimizes the rms between the pairs of graphs, while the MC method uses a Boltzmann-type probability function which distributes the various solutions on the different accessible energy levels:

$$p = \exp(-\beta(\mathrm{rms})) \tag{12}$$

where β is chosen to roughly balance the number of accepted and rejected trials.

For the structural alignment of drug molecules, other similarity scores can be proposed such as in the program SEAL described by Klebe et al.:[46]

$$A_F = -\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{l} w_k P_{ik} P_{jk} e^{-\alpha_k r_{ij}^2} \tag{13}$$

wherein summations over *i* and *j* run over all atoms of the first and second molecule, respectively. The third summation runs over all considered properties P_k (steric, electrostatic, hydrophobic). This approach requires an optimization of weights w_k and factors α_k to generate the best rms values.

It appeared that both the MC and the GA approaches that we implemented are efficient enough to perform pairwise comparisons of critical point graphs built from electron density functions evaluated at the two levels of crystallographic resolution. However, the matching of a larger number of graphs is only very slowly convergent using the MC approach, especially when chemical graphs are structurally different. This was the case for compound TZQ whose chemical structure differs largely from compounds Diazepam, Ro5-4864, and QUIZ. We thus included our MC algorithm in a simulated annealing (SA) procedure[47] which proved

Table 8. First Best Superposition Results Obtained Using a Genetic Algorithm and a Monte Carlo Algorithm for the Comparison of the Critical Point Graphs[a]

n points		rms								Δ rms
Diazepam			1	2	3	4	5	6	7	
Ro5-4864	standard	0.346	1	3	4	5	6	7	9	
	inverted	1.236	1	3	8	5	6	4	9	0.890
	inverted	1.300	1	3	7	5	8	4	9	0.954
QUIZ			1	2	3	4	5	6	7	8
Ro5-4864	standard	0.508	1	4	3	6	5	9	8	7
	inverted	1.377	4	1	5	6	3	9	7	8 0.869
	inverted	1.382	4	1	5	6	3	9	8	7 0.874
TZQ			1	2	3	4	5	6	7	
Ro5-4864	inverted	1.622	3	5	4	6	9	2	7	
	standard	1.642	3	4	5	6	9	2	7	0.020
	inverted	1.701	3	4	6	1	9	7	8	0.079
Diazepam			1	2	3	4	5	6	7	
QUIZ	standard	1.517	1	3	2	5	4	7	6	
TZQ	inverted		7	3	4	1	2	6	5	
Ro5-4864	standard		1	3	4	5	6	8	9	
QUIZ[b]	standard	1.547	1	3	2	5	4	8	6	
TZQ	inverted		6	3	2	1	7	4	5	
Ro5-4864	standard		1	3	4	5	6	7	9	
QUIZ[b]	standard	1.540	1	7	2	5	4	8	6	
TZQ	standard		4	7	2	1	3	6	5	
Ro5-4864	standard		1	8	4	5	6	7	9	
QUIZ	inverted	1.592	8	5	4	3	2	7	6	
TZQ	inverted		7	3	4	1	2	6	5	
Ro5-4864	inverted		7	5	6	3	4	8	9	

[a] For each set of two or four graphs, all *n* points of the smallest molecular graph were considered. Peaks and pales were obtained using ORCRIT for electron density maps calculated at a resolution of 2.5 Å using XTAL. The numbering of critical points is presented in Figures 4 and 8. [b] GA results only.

Table 9. First Best Superposition Results Obtained Using a Genetic Algorithm and a Monte Carlo Algorithm for the Comparison of the Critical Point Graphs[a]

$n-1$ points		rms								Δ rms
Diazepam			1	2	3	4	5	6	7	
Ro5-4864	standard	0.364	1	3	4	5	6	7	−	
	inverted	1.005	1	3	7	5	8	4	−	0.641
QUIZ			1	2	3	4	5	6	7	8
Ro5-4864	standard	0.406	1	4	3	6	5	−	8	7
	inverted	1.267	4	1	5	6	3	−	8	7 0.861
TZQ			1	2	3	4	5	6	7	
Ro5-4864	standard	1.080	3	4	6	−	9	8	7	
	inverted	1.173	3	6	4	−	9	7	8	0.093
Diazepam			1	2	3	4	5	6	7	
QUIZ	standard	1.039	−	3	2	5	4	8	6	
TZQ	standard		−	1	2	3	4	6	5	
Ro5-4864	standard		−	3	4	5	6	7	9	
QUIZ	inverted	1.106	−	5	4	3	2	7	6	
TZQ	standard		−	1	2	3	4	6	5	
Ro5-4864	standard		−	3	4	5	6	7	9	

[a] For each set of two or four graphs, $(n-1)$ points of the smallest molecular graph were considered. Peaks and pales were obtained using ORCRIT for electron density maps calculated at a resolution of 2.5 Å using XTAL. The numbering of critical points is presented in Figures 4 and 8.

to be more efficient in tackling such multiple matching problems. In this scheme, the range of possible values for parameter β (eq 12) is determined through two limiting values, β_{min} and β_{max}. A MC matching procedure is thus applied at each successive β value in an ascending order. In a physical system, this corresponds to the generation of energy states which obey the Boltzmann distribution function at successively lower temperatures. As the temperature decreases, the Boltzmann distribution tends toward the system configuration with the lowest energy. SA approaches are often used to overcome multiple energy minima problems, such as conformational analysis problems.[48,49]

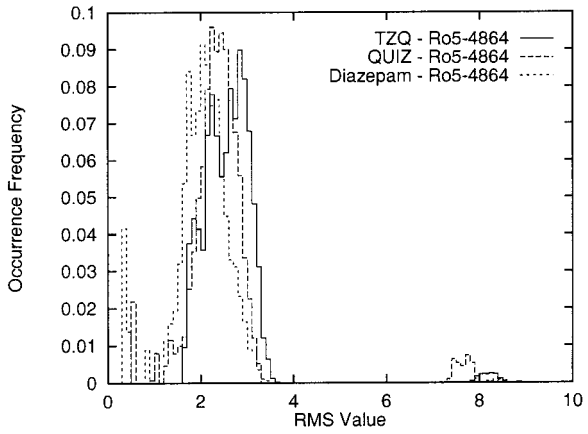
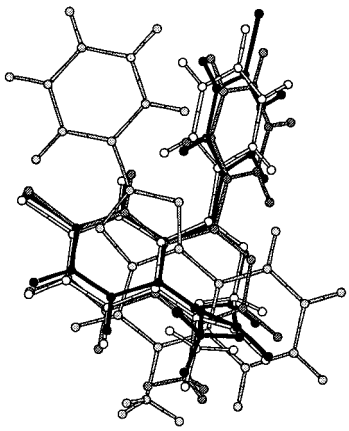


Figure 14. Occurrence frequency distribution of rms values characterizing *n*-point superposition solutions of the three compounds TZQ, QUIZ, and Diazepam, over the reference graph Ro5-4864. The four critical point graphs were obtained by topological analysis of electron density maps calculated using XTAL at a resolution of 2.5 Å.

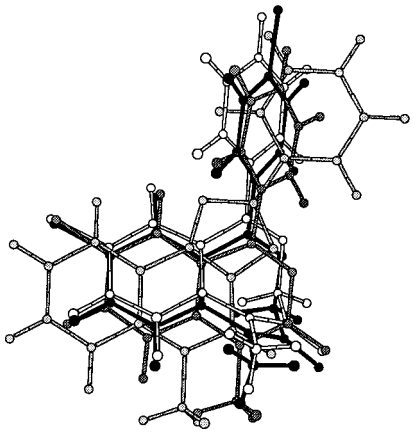
(c) Results. In this subsection, we detail and discuss the results of both MC and GA pairwise and multiple alignments.

As we already mentioned earlier, fully connected graphs (see Supporting Information) were used in all matching applications: all possible links between any pair of critical points were added to the initial connectivity patterns.

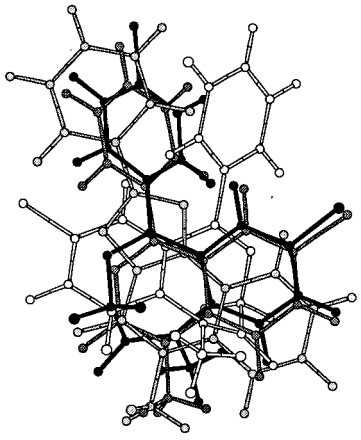
Two matching conditions were considered. On the first hand, the number of points to be matched was determined as the size of the smallest critical point graph to be compared (*n*-point comparison). Unfortunately, in the case of structurally different molecules, this may force the superposition of unrelated points. Therefore, matching experiments using $(n-1)$ points were also carried out, which is generally known as the leave-one-out strategy. Pairwise matchings were achieved using both *n* and $(n-1)$ points. As mentioned previously, the single MC and GA procedures were efficient enough to achieve optimal superpositions. A SA procedure was however requested to replace the MC one in order to superpose all four molecular graphs at a time. MC optimization parameters are reported in Tables 4 and 5 and GA parameters are reported in Tables 6 and 7, for alignments of the 2.5 and 3.0 Å resolution graphs, respectively. For the MC simulations, we first selected a β value to obtain a number of accepted trials as close as possible to the number of rejected ones. It is observed that, for a given β value, the number of accepted moves is reduced when the graph size diminishes. This is justified by the fact that a change in the matched points has a relatively stronger effect on the rms value. In the case of the SA simulations, 50 000 iterations were achieved at each β value. However, with a reduction of the number of matched points, we had to increase either this number or the number of MC runs in order to obtain the optimal solution. For the GA runs, all parameters were kept constant, except for the multiple superpositions. In these cases, the mutation rate and the number of generations had to be increased in order to sample a larger number of potential solutions. The elitist option was also applied in



RMS=1.517

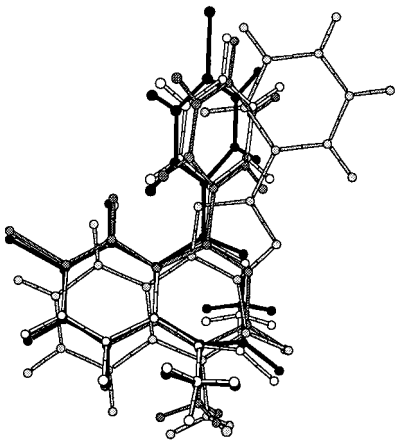


RMS=1.540

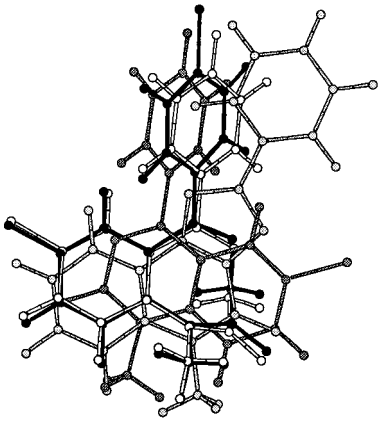


RMS=1.592

Figure 15. Selected *n*-point solutions of Monte Carlo/simulated annealing and genetic algorithm quaternary alignments of graphs obtained using a critical point analysis of electron density distributions at a resolution of 2.5 Å. Compounds Ro5-4864, Diazepam, QUIZ, and TZQ are shown as black, white, dark gray, and light gray sticks, respectively.



RMS=1.039



RMS=1.106

Figure 16. Selected ($n - 1$)-point solutions of Monte Carlo/simulated annealing and genetic algorithm quaternary alignments of graphs obtained using a critical point analysis of electron density distributions at a resolution of 2.5 Å. Compounds Ro5-4864, Diazepam, QUIZ, and TZQ are shown as black, white, dark gray, and light gray sticks, respectively.

order to ensure that the best solution obtained in a generation is copied in the next one. CPU times on an IBM RISC6000 model 580 platform are <1 s for 10^4 MC iterations and lie between 150 and 250 s for 10^4 GA generations (we remind that at each GA generation, every chromosome requires a calculation of its fitness).

Final superimpositions were obtained using QUATFIT,[50] a program designed to superpose specified atoms of two molecules by minimizing the least squares errors. Figures 6 and 7 illustrate solutions of the superposition procedure applied to pairs of graphs at a resolution of 2.5 Å using n and ($n - 1$) points, respectively. Results at 3.0 Å resolution are displayed in Figures 8 and 9. Transformation matrices resulting from the application of the program QUATFIT were used to generate visual representations of the corresponding molecular structures (Figures 10−13).

(c.1) Resolution of 2.5 Å. Optimal matchings between pairs of graphs that are characterized by the lowest rms values are presented in Tables 8 and 9. Other solutions that are representative of various classes of superposition results are also reported. All reported results were selected among the 50 best solutions that were generated during the simulation procedures.

It is first observed that the best matching of the most dissimilar molecules, i.e., TZQ and Ro5-4864, is characterized by a larger rms value than for the two other molecules. Results also show that the optimal matchings occur without any change of orientation, except for compound TZQ

Table 10. Best Superposition Results Obtained Using a Monte Carlo and a Genetic Algorithm for the Comparison of the Critical Point Graphs[a]

n points		rms							Δ rms
Diazepam			1	2	3	4	5	6	7	
Ro5-4864	standard	1.369	1	4	3	5	2	6	7	
	inverted	1.613	2	4	3	5	1	7	6	0.244
	inverted	1.699	2	1	3	6	4	7	5	0.330
	inverted	1.806	1	2	3	7	4	6	5	0.437
	inverted	1.958	4	1	3	6	2	5	7	0.589
QUIZ			1	2	3	4	5			
Ro5-4864	standard	1.069	1	3	6	2	7			
	inverted	1.410	2	3	7	1	6			0.341
	inverted	1.422	2	1	7	3	6			0.353
	inverted	1.508	4	1	6	2	7			0.439
	inverted	1.598	1	2	7	3	6			0.529
	inverted	1.683	2	3	7	4	6			0.614
TZQ			1	2	3	4	5	6	7	
Ro5-4864	standard	1.337	3	4	5	1	2	6	7	
	inverted	1.658	3	1	6	4	2	5	7	0.321
	inverted	1.945	3	4	5	2	4	1	6	0.608
	inverted	2.081	3	1	6	2	4	7	5	0.744
Diazepam			1	2	3	4	5	6	7	
QUIZ	standard	1.090	1	2	−	3	4	−	5	
TZQ	standard		4	2	−	3	5	−	7	
Ro5-4864	standard		1	3	−	6	2	−	7	
QUIZ	inverted	1.105	1	4	−	5	2	−	3	
TZQ	inverted		4	5	−	7	2	−	3	
Ro5-4864	inverted		1	2	−	7	3	−	6	

[a] For each set of two or four graphs, all *n* points of the smallest molecular graph were considered. Peaks and passes were obtained using ORCRIT for electron density maps calculated at a resolution of 3.0 Å using XTAL. The numbering of critical points is presented in Figures 6 and 9.

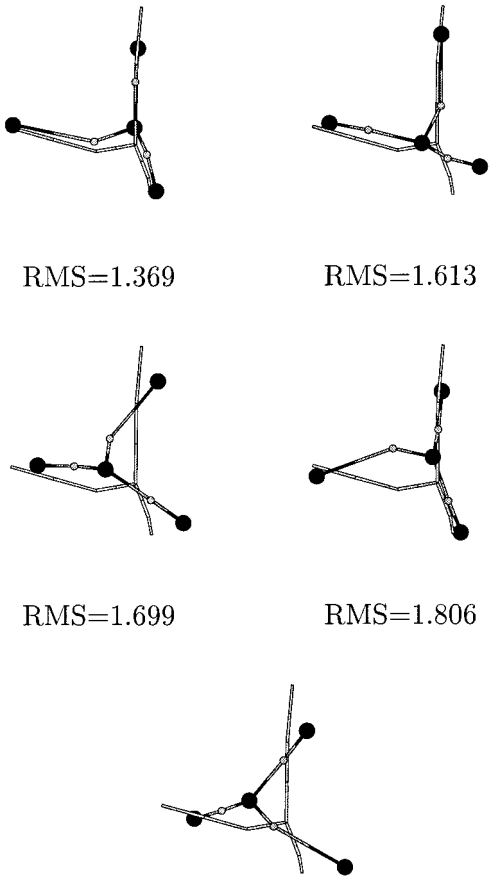


Figure 17. Selected *n*-point solutions of Monte Carlo and genetic algorithm matching procedures for Diazepam/Ro5-4864 graphs obtained using a critical point analysis of electron density distributions at a resolution of 3.0 Å. Compound Ro5-4864 is displayed using gray sticks. Diazepam is shown using black (peaks) and gray spheres (passes).

superposed over Ro5-4864 using an *n*-point approach. The superposition of planar (TZQ) and nonplanar graphs (Ro5-4864) leads to a higher rms value. There is only a slight difference in rms value between the inverted superposition (rms = 1.622) and the standard superposition (rms = 1.642); for that particular pair of compounds, the differentiation of the solutions is not as good as in the other cases, which can be justified as follows. All chlorine atoms are characterized by the highest density values, with the exception of TZQ where all peaks have about the same density value. Inverted configurations of TZQ vs Ro5-4864 are thus more likely to occur.

In standard superpositions, there is an exact mapping between the LS and δ groups (Figures 6, 10 and 7, 11). Each benzodiazepine and phenyl group is indeed associated with its counterpart in the reference molecule Ro5-4864. These solutions are not unique, as shown in Figure 14. Three classes of solutions are clearly observed for each pair of molecules. The two classes of solutions with the lowest rms values correspond to alignments with the pale that is correctly positioned.

Inverted solutions are observed in all pairwise matchings. For example, Diazepam undergoes a rotation of 180° along an axis defined by critical points 1-4-2 (Figure 4). The critical point associated with the phenyl moiety of Diazepam thus lies on top of the carbonyl group of compound Ro5-4864. Inverted solutions are characterized by a biased superimposition of the cycles. Branches of the reference and fitted molecules are pointing out in opposite directions.

A more detailed analysis of Tables 8 and 9 shows that solutions that were obtained with the leave-one-out strategy are subsets of more global solutions obtained with the *n*-point approach. Indeed, for the pair Diazepam/Ro5-4864, the (n − 1)-point solution (rms = 1.005) is a subset of the *n*-point solution (rms = 1.300). We note here that solutions with rms values equal to 1.236 and 1.300 led to identical superimposition patterns. For the pair QUIZ/Ro5-4864, the best (n − 1)-point solution (rms = 0.406) is also a subset of the best *n*-point solution (rms = 0.508), while a solution with rms = 1.267 is similar to the *n*-point solution with rms = 1.382. Solutions with rms values equal to 1.377 and 1.382 also led to identical superimposition patterns. In the case of matchings between TZQ and Ro5-4864, the standard (n − 1)-point solution with rms = 1.080 is a subset of the standard *n*-point solution with rms = 1.701. It is observed that the best solution generated using the leave-one-out strategy adopts a standard superimposition pattern, while this pattern was not the best one in the case of *n*-point superpositions. The difference comes from the elimination of the Cl peak from the matched points due to the fact that TZQ does not contain any Cl atom. Solutions characterized by rms values equal to 1.080 and 1.701 seem unphysical since the point connectivity is not preserved. For example, points 3 and 6 of the first solution, which are not connected together in structure Ro5-4864 (Figure 4), correspond to connected

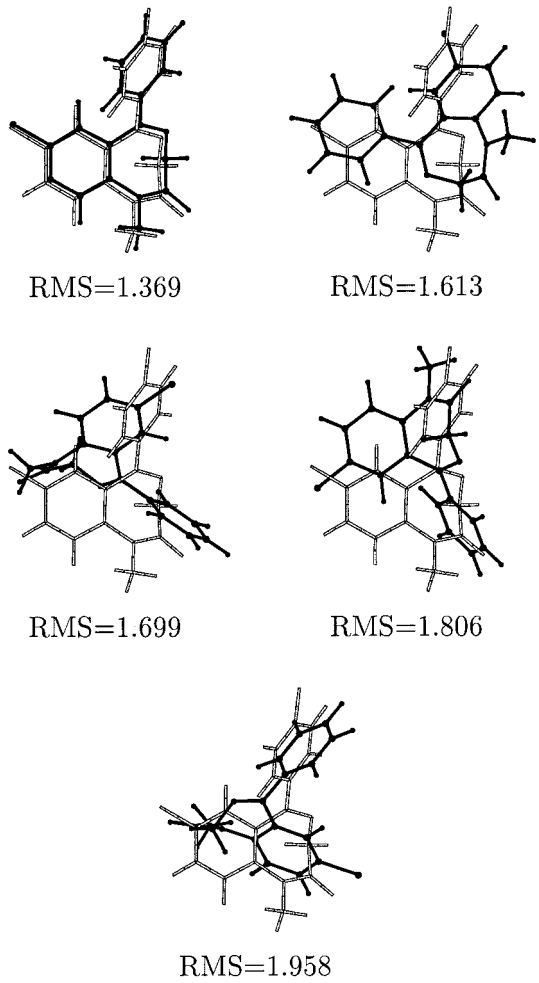


Figure 18. Selected *n*-point solutions of Monte Carlo and genetic algorithm matching procedures for Diazepam/Ro5-4864 graphs obtained using a critical point analysis of electron density distributions at a resolution of 3.0 Å. Compound Ro5-4864 is displayed using gray sticks. Diazepam is shown using black sticks and balls.

points 1 and 3 in structure TZQ. Since the original connectivity pattern was not considered in the MC and GA approaches, these solutions were not rejected. Further developments of the methods may involve a (post)processing of the MC/GA results based on the rejection of solutions which do not obey the initial connectivity pattern.

MC matchings with $(n - 1)$ points involve the elimination of the pale (point no. 7) of the Diazepam structure from the comparisons (Table 9). While this does not seem to affect the best solutions, this might correspond to the rejection of an important constraint. The same observation can be done for QUIZ where the pale is point no. 6.

Multiple alignments of the four compounds were achieved using the SA approach described previously, and with the GAGS approach. Parameters of the simulations are reported in Tables 4 and 6 for superpositions using n and $(n - 1)$ points. Corresponding molecular structure alignments are shown in Figures 15 and 16 for n and $(n - 1)$-point superimpositions, respectively. In each case the two best solutions observed with the lowest β value are presented in Tables 8 and 9, respectively. In *n*-point solutions with rms = 1.517 and 1.547, all molecules adopt a standard pattern with respect to reference Ro5-4864, except for TZQ. Diazepam is the only inverted molecule in solution with rms = 1.592. The best $(n - 1)$-point solution with rms = 1.039 is an aggregate of the already observed pair matchings: rms = 0.364 for Diazepam/Ro5-4864, rms = 0.406 for QUIZ/Ro5-4864, and rms = 1.080 for TZQ/Ro5-4864. With $(n - 1)$ points, all compounds of the solution characterized by rms = 1.039 are aligned without any change of orientation, as already observed in pairwise superimpositions. The $(n - 1)$-point solution with rms = 1.106 shows an inversion of compound QUIZ only, as displayed in the corresponding pair matching with rms = 1.377. Therefore, multiple *n*-point solutions lead to the expected standard superimposition patterns, i.e., where all critical points are matched according to their chemical nature: LS, δ, and FRA. The use of the leave-one-out approach induces a loss of information because TZQ does not cantain any Cl atom, and the standard superimpositions are not observed as the best solutions.

Table 11. Best Superposition Results Obtained Using a Monte Carlo and a Genetic Algorithm for the Comparison of Critical Point Graphs[a]

$n-1$ points		rms								Δ rms
Diazepam			1	2	3	4	5	6	7	
Ro5-4864	standard	0.784	1	4	3	5	−	6	7	
	inverted	1.234	1	−	3	7	4	6	5	0.450
	inverted	1.463	2	4	3	5	1	−	6	0.679
	inverted	1.617	2	1	3	−	4	7	6	0.833
QUIZ			1	2	3	4	5			
Ro5-4864	standard	0.690	1	3	6	−	7			
	inverted	0.976	2	1	−	3	6			0.286
	inverted	1.119	4	1	6	−	7			0.429
	inverted	1.142	1	−	6	4	5			0.452
	inverted	1.145	2	3	7	1	−			0.455
TZQ			1	2	3	4	5	6	7	
Ro5-4864	inverted	0.894	−	3	6	4	2	5	7	
	standard	1.168	−	3	5	1	2	6	7	0.274
	inverted	1.322	−	3	5	4	2	6	7	0.428
	inverted	1.581	3	4	5	−	1	7	6	0.687
	inverted	1.616	−	3	6	4	1	5	7	0.722
Diazepam			1	2	3	4	5	6	7	
QUIZ	standard	0.893	1	−	−	−	4	3	5	
TZQ	inverted		5	−	−	−	2	7	3	
Ro5-4864	inverted		2	−	−	−	3	7	6	
QUIZ	standard	0.899	1	−	2	−	−	3	5	
TZQ	standard		4	−	1	−	−	3	7	
Ro5-4864	standard		1	−	3	−	−	6	7	

[a] For each set of two or four graphs, $(n - 1)$ points of the smallest molecular graph were considered. Peaks and passes were obtained using ORCRIT for electron density maps calculated at a resolution of 3.0 Å using XTAL. The numbering of critical points is presented in Figures 6 and 9.

(c.2) Resolution of 3.0 Å. Under these conditions, the alignment of all molecules on Ro5-4864 generates several rotameric solutions with close rms values (Table 10). The comparison of the TZQ structure which is different from the others does not lead to a very different rms value. Therefore, a higher diversity in the results is observed in every case. Since critical points are characterized by their density value only, several superposition patterns are proposed. For example, the various patterns observed for the *n*-point matching of Diazepam/Ro5-4864, with the lowest rms value for each class of configurations, are presented in Figures 17 and 18. Solutions regarding other pairs of molecules and multiple matchings are shown in Tables 10 and 11. Multiple molecular structure alignments are shown in Figure 19. In each case, all best solutions correspond to the expected standard superimposition patterns. It is only when a constraint

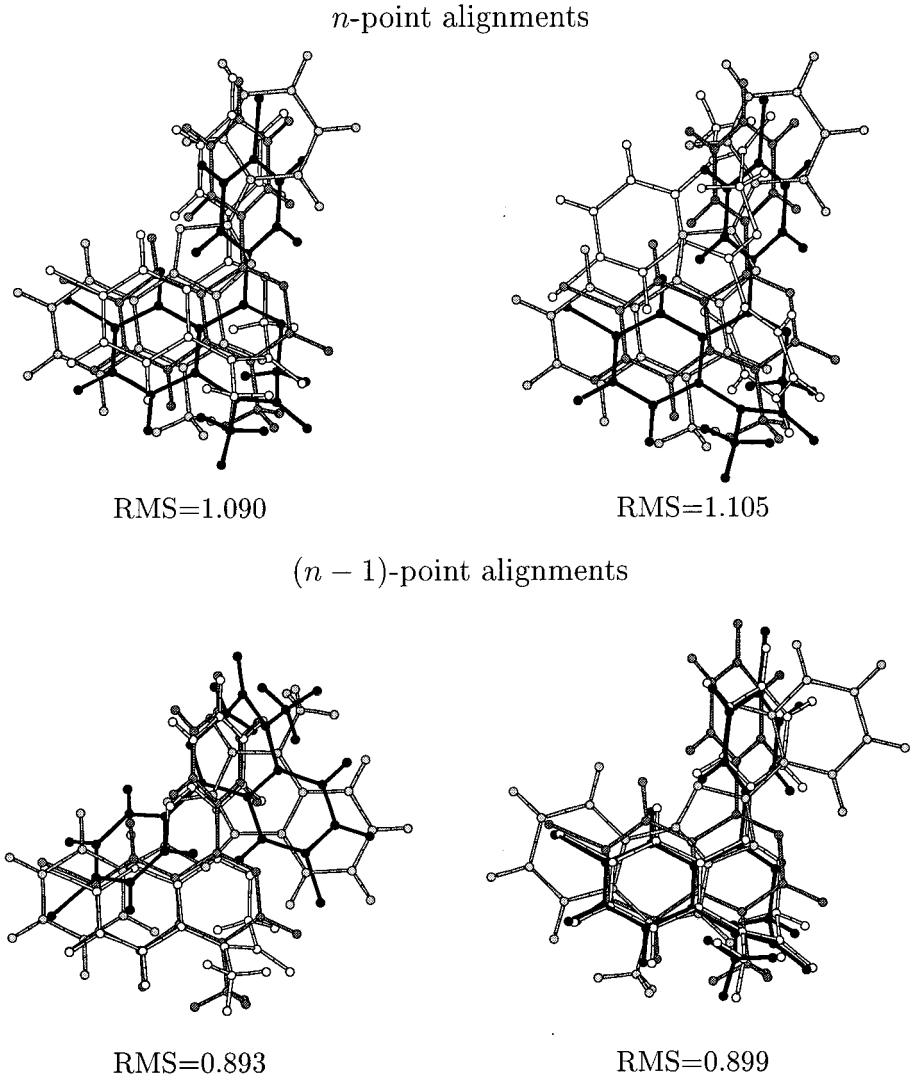


Figure 19. Selected n and $(n-1)$-point solutions of Monte Carlo/simulated annealing and genetic algorithm quaternary alignments of graphs obtained using a critical point analysis of electron density distributions at a resolution of 3.0 Å. Compounds Ro5-4864, Diazepam, QUIZ, and TZQ are shown as black, white, dark gray, and light gray sticks, respectively.

is removed, i.e., when using the leave-one-out approach, that inverted best solutions appear, e.g., in the case of pairwise TZQ/Ro5-4864 matchings. We will also note that, for the multiple Diazepam/QUIZ/TZQ/Ro5-4864 $(n-1)$-point superpositions, only Diazepam and QUIZ, and TZQ and Ro5-4864 are properly superimposed. Again, this is due to the neglect of chlorine peaks among the matched critical points.

Figures 8 and 12 show that the best n-point solutions correspond to standard superpositions. Once more, compound TZQ presents a higher similarity to Ro5-4864 when inverted, as shown by $(n-1)$-point best solutions in Figures 9 and 13. It is also interesting to note the inverted superposition of QUIZ on Ro5-4864 (rms = 0.976) which does not follow the initial connectivity pattern. Also, but with the exception of TZQ, all inverted $(n-1)$-point configurations are different from their corresponding inverted n-point configurations. This illustrates the increase of variability in the superposition results when the number of constraints, i.e., the number of critical points, diminishes.

5. CONCLUSIONS

In this work, we developed both a Monte Carlo/simulated annealing and a genetic algorithm procedure to compare various molecular structures based on their three-dimensional electron density distributions. Particularly, the procedures were applied to the superposition of benzodiazepine-related chemical structures. First, we used a topological analysis method in order to obtain a reduced dimensional representation of the information to be compared. This method generates a representation of the ED distributions in terms of graphs of critical points characterized by their local density and curvature values. Second, we applied the comparison procedures to connected graphs resulting from the topological analysis of ED maps generated at various crystallographic resolution levels, at medium 2.5 Å resolution and at a lower 3.0 Å resolution.

We observed that a representation in terms of critical points of ED maps generated at medium resolutions is efficient to rapidly compare 3D molecular properties. The analyses carried out at the various levels of resolution constitute a practical way to reduce the amount of information to be compared. From a fundamental point of view, such an approach is a way to discover new descriptive elements for the hierarchical analysis and interpretation of binding modes in terms of molecular properties. We have observed that, at

3.0 Å resolution, molecules with similar affinities adopt a similar topology of their pharmacophore elements. Extrapolation of this observation leads to the conclusion that the global recognition of a ligand by a receptor is a medium resolution phenomenon. Then, it is necessary to get more detailed (2.5 Å) representation to confirm the relative orientation of such molecules and, thus, their potential orientation within the receptor. A hierarchical way of viewing molecular interactions has already been proposed by Butzlaff et al.[51]

If the property, i.e., the ED distribution, that was used in this work is significant for graph comparison, the low affinity of TZQ with respect to mitochondrial receptors (MBR) might be explained by the multiple orientations that TZQ can adopt when interacting with the receptors.

In the present work, the ED distributions were generated through the prior calculation of structure factors. The method implies that the resulting ED function is a combination of spherical approximations of the individual atomic contributions to the function. This promolecular description of a molecule is obviously in lack of any bond polarization and/or charge transfer effects due to the chemical nature of the constituting atoms. Therefore, we have also adopted an alternative approach where ED maps are generated through quantum mechanical calculations. The resulting ED map is then smoothed using a wavelet-based procedure. Investigation of the results is now under progress.[52]

All molecules considered in this work are particularly rigid. This greatly simplifies the elaboration of a ligand comparison procedure. The treatment of flexibility is also under progress.

ACKNOWLEDGMENT

All authors wish to thank C. K. Johnson for providing the program ORCRIT and S. Fortier, J. I. Glasgow, F. H. Allen, G. M. Maggiora, and J.-J. Bourguignon for fruitful discussions. They also acknowledge the FUNDP for the use of the Namur Scientific Computing Facility (SCF) Center, the financial support of the FNRS-FRFC, the "Loterie Nationale" for Convention No 9.4595.96, and IBM-Belgium for the Academic Joint Study on "Cooperative Processing for Theoretical Physics and Chemistry". N.M. thanks the "Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture" (FRIA) for her Ph.D. fellowship

Supporting Information Available: Figures showing isocontours and connected graphs and tables of property matrices. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Robinson, D. D.; Barlow, Th. W.; Richards, W. G. Reduced Dimensional Representations of Molecular Structures. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 939−942.

(2) Robinson, D. D.; Barlow, Th. W.; Richards, W. G. The Utilization of Reduced Dimensional Representations of Molecular Structure for Rapid Molecular Similarity Calculations. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 943−950.

(3) Rohrer, D. C. In *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches;* Carbó, R., Ed.; Kluwer Academic: Dordrecht, The Netherlands, 1995; p 141.

(4) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. MIMIC: A Molecular-Field Matching Program. Exploiting Applicability of Molecular Similarity Approaches. *J. Comput. Chem.* **1997**, *18*, 934−954.

(5) Carbó, R.; Calabuig, B.; Vera, L.; Besalú, E. Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques. *Adv. Quantum Chem.* **1994**, *25*, 253−313.

(6) Good, A. C.; Hodgkin, E. E.; Richards, W. G. The Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188−191.

(7) Mezey, P. G. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990; p 321.

(8) Mezey, P. G. The Shape of Molecular Charge Distribution: Group Theory without Symmetry. *J. Comput. Chem.* **1987**, *8*, 462−469.

(9) Arteca, G. A.; Mezey, P. G. Shape Group Studies of Molecular Similarity and Regioselectivity in Chemical Reactions. *J. Comput. Chem.* **1988**, *9*, 608−619.

(10) Lin, S. L.; Nussinov, R.; Fisher, D.; Wolfson, H. J. Molecular Surface Representations by Sparse Critical Points. *PROTEINS: Struct., Funct. Genet.* **1994**, *18*, 94−101.

(11) Bader, R. W. *Atoms in Molecules−A Quantum Theory*; Clarendon Press: Oxford, U.K., 1995.

(12) Johnson, C. K. *ORCRIT. The Oak Ridge Critical Point Network Program;* Technical report. Chemistry Division, Oak Ridge National Laboratory: Oak Ridge, TN, 1977.

(13) Shirsat, R. N.; Bapat, S. V.; Gadre, S. R. Molecular Electrostatics. A Comprehensive Topographical Approach. *Chem. Phys. Lett.* **1992**, *200*, 373−378.

(14) Gadre, S. R.; Pundlik, S. S. Topographical Analysis of Electron Density and Molecular Electrostatic Potential for Cyclopropa- and Cyclobutabenzenes. *J. Am. Chem. Soc.* **1995**, *117*, 9559−9563.

(15) Wild, D. J.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures: Alignment of Molecular Electrostatic Potential Fields with a Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 159−167.

(16) Thorner, D. A.; Wild, D. J.; Willett, P.; Wright, P. M. Similarity Searching in Files of Three-Dimensional Chemical Structures: Flexible Field-Based Searching of Molecular Electrostatic Potentials. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 900−908.

(17) Moehler, H.; Okada, T. Benzodiazepine Receptor: Demonstration in the Central Nervous System. *Science* **1977**, *198*, 849−851.

(18) Braestrup, C.; Squire, R. F. Specific Benzodiazepine Receptors in Rat Brain Characterized by High Affinity [3H]-Diazepam Binding. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 3805−3809.

(19) Bernassau, J. M.; Reversat, J. L.; Ferrara, P.; Caput, D.; Lefur, G. A 3D Model of the Peripheral Benzodiazepine Receptor and its Implication in Intramitochondrial Cholesterol Transport. *J. Mol. Graphics* **1993**, *11*, 236−244.

(20) Bourguignon, J.-J. In *Peripheral Benzodiazepine Receptors*; Giesen-Crouse, E., Ed.; Academic Press: San Diego, CA, 1993; p 59.

(21) Colotta, V.; Catarzi, D.; Varano, F.; Filacchioni, G.; Cecchi, L. Synthesis and Binding Activity of Some Pyrazolo[1,5-*c*] Quinazolines as Tools To Verify an Optional Binding Site of a Benzodiazepine Receptor Ligand. *J. Med. Chem.* **1996**, *39*, 2915−2921.

(22) Loew, G. H.; Nienow, J. R.; Poulsen, M. Theoretical Structure−Activity Studies of Benzodiazepine Analogues. Requirements for Receptor Affinity and Activity. *Mol. Pharmacol.* **1984**, *26*, 19−34.

(23) Codding, P. W.; Muir, A. K. S. Molecular Structure of Ro15−1788 and a Model for the Binding of Benzodiazepine Receptor Ligands. *Mol. Pharmacol.* **1985**, *28*, 178−184.

(24) Fryer, R. I. Conformational Shifts at the Benzodiazepine Receptor Related to the Binding of Agonists, Antagonists, and Inverse Agonists. *Life Sci.* **1986**, *39*, 1947−1957.

(25) Borea, P. A.; Gilli, G.; Bertolasi, V.; Ferretti, V. Stereochemical Features Controlling Binding and Intrinsic Activity Properties of Benzodiazepine Receptor Ligands. *Mol. Pharmacol.* **1987**, *31*, 334−344.

(26) Tebib, S.; Bourguignon, J.-J.; Wermuth, C.-G. The Active Analogue Approach Applied to the Pharmacophore Identification of Benzodiazepine Receptor Ligands. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 153−170.

(27) Villar, H. O.; Davies, M. F.; Loew, G. H.; Maguire, P. A. Molecular Models for Recognition and Activation at the Benzodiazepine Receptor: A Review. *Life Sci.* **1991**, *48*, 593−602.

(28) Mátyus, P.; Barlin, G. B.; Harrison, P. W.; Wong, M. G.; Davies, L. P. Ligands for the Central Benzodiazepine Receptor: Structure-Affinity Relationship Studies on Imidazo[1,2-b]pyridazines. *Aust. J. Chem.* **1996**, *49*, 435−442.

(29) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macroe, C. F.; Mitchell, E. M.; Smith, G. F.; Watson, D. G. The Development of Versions 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187−209.

(30) Meurice, N.; Norberg, B.; Durant, F.; Didier, B.; Bourguignon, J.-J.; Vercauteren, D. P. *Acta Crystallogr.*, manuscript in preparation.

(31) *XTAL 3.0 User's Manual*; Hall, S. R.; Stewart, J. M., Eds.; Universities of Western Australia and Maryland: Nedlands, Western Australia, and College Park, MD, 1990.
(32) Leherte, L.; Allen, F. H. Shape Information from a Critical Point Analysis of Calculated Electron Density Maps: Application to DNA-Drug Systems. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 257−272.
(33) Leherte, L.; Fortier, S.; Glasgow, J.; Allen, F. H. Molecular Scene Analysis: Application of a Topological Approach to the Automated Interpretation of Protein Electron Density Maps. *Acta Crystallogr. D* **1994**, *50*, 155−166.
(34) Leherte, L.; Vercauteren, D. P. Critical Point Analysis of Calculated Electron Density Maps at Medium Resolution: Application to Shape Analysis of Zeolite-Like Systems. *J. Mol. Model.* **1997**, *3*, 156−171.
(35) Hodgkin, E. E.; Miller, A.; Whittaker, M. A Monte Carlo Pharmacophore Generation Procedure: Application to the Human PAF Receptor. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 515−534.
(36) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(37) Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Reading, MA, 1989.
(38) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms Part 1. Concepts, Properties and Context. *Chemom. Intell. Lab. Syst.* **1993**, *19*, 1−33.
(39) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms Part 2. Representation, Configuration and Hybridization. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 99−145.
(40) Paretti, M. F.; Kroemer, R. T.; Rothman, J. H.; Richards, W. G. Alignment of Molecules by the Monte Carlo Optimization of Molecular Similarity Indices. *J. Comput. Chem.* **1997**, *18*, 1344−1353.
(41) Brown, R. D.; Jones, G.; Willett, P. Matching Two-Dimensional Chemical Graphs Using Genetic Algorithms. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 63−70.
(42) Jones, G.; Willett, P.; Glen, R. C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532−549.
(43) Handschuh, S.; Wagener, M.; Gasteiger, J. Superposition of Three-Dimensional Chemical Structures Allowing for Conformational Flexibility by a Hybrid Methodol. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 220−232.
(44) Meurice, N.; Leherte, L.; Vercauteren, D. P.; Bourguignon, J.-J.; Wermuth, C. G. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; Verlag: Basel, Switzerland, 1997; p 497.
(45) Meurice, N.; Leherte, L.; Vercauteren, D. P. Comparison of Benzodiazepine-Like Compounds Using Topological Analysis and Genetic Algorithms. *SAR QSAR Environ. Res.* **1998**, *8*, 195−232.
(46) Klebe, G.; Mietzner, Th.; Weber, F. Different Approaches Toward an Automatic Structural Alignment of Drug Molecules: Applications to Sterol Mimics, Thrombin and Thermolysin Inhibitors. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 751−778.
(47) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671−680.
(48) Mundim, K. C.; Tsallis, C. Geometry Optimization and Conformational Analysis Through Generalized Simulated Annealing. *Int. J. Quantum Chem.* **1996**, *58*, 373−381.
(49) Wang, Z.; Ruth, R. Prediction of Peptide Conformation: The Adaptive Simulated Annealing Approach. *J. Comput. Chem.* **1997**, *18*, 323−329.
(50) Heisterberg, D. J. Technical report, Ohio Supercomputer Center, Columbus, OH. Translation from FORTRAN to C and input/output by Jan Labanowski, Ohio Supercomputer Center, Columbus, OH, 1990.
(51) Butzlaff, M.; Dahmen, W.; Diekmann, S.; Dress, A.; Schmitt, E.; von Kitzing, E. A Hierarchical Approach to Force Field Calculations Through Spline Approximations. *J. Math. Chem.* **1994**, *15*, 77−92.
(52) Leherte, L. Application of Multiresolution Analyses to Electron Density Maps of Small Molecules. *J. Math. Chem.*, manuscript in preparation.
(53) *IBM Visualization Data Explorer*; IBM Corp: Riverton, NJ, 1996.

CI990112D