Use of Electron Density Critical Points as Chemical Function-Based Reduced Representations of Pharmacological Ligands

John Binamé,[†] Nathalie Meurice,*[,†] Laurence Leherte,[†] Janice Glasgow,[‡] Suzanne Fortier,[§] and Daniel P. Vercauteren[†]

Laboratoire de Physico-Chimie Informatique, Facultés Universitaires Notre-Dame de la Paix, 61 Rue de Bruxelles, B-5000 Namur, Belgium, and School of Computing and Department of Chemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

Received July 29, 2003

In this paper, we propose a reduced representation of molecules of pharmacological interest based on their chemical functions. The proposed representations of the molecules are obtained by a topological analysis of their electron density maps at medium resolution, leading to graphs of critical points. The distribution of the different types of critical points are compared at various levels of resolution for a training set of 22 molecules in order to define the optimal resolution level leading to the best representation of the various chemical functions. The reduced representations can in the future be used for molecular similarity research and pharmacophore proposals.

INTRODUCTION

The development of lead compounds in biopharmacological research has always been a major subject of interest both at the experimental and theoretical levels. To design new leads, scientists need to understand as precisely as possible the human body's biological processes, including the physicochemical interactions occurring between parts of large protein targets, such as receptors or enzymes, and small organic molecules, such as ligands.[1] Therefore, the identification of the receptor and the determination of its three-dimensional (3D) structure are key steps in new lead design processes. When an identified receptor can be crystallized, with or without the ligand, its 3D structure can be solved a priori using experimental techniques such as X-ray diffraction or NMR. Alternatively, theoretical methods such as protein sequence homology or molecular mechanics type optimization can, in some cases, also help to build 3D structural models for a noncrystallized but identified receptor. Knowing the 3D structure, one can then locate and analyze the binding site and determine the physicochemical characteristics required to interact correctly with the ligand; new potential lead compounds can then be found by screening molecular structure databases[2] or by ligand−receptor docking[3] or de novo design.[4] Unfortunately, most of the time, the above-mentioned techniques still fail to provide accurate 3D models of the target proteins. Consequently, a number of different theoretical approaches, mainly based on the comparison of ligands of known affinity and/or activity, have been developed in order to bypass this difficulty.

It is well-known that ligand−receptor interactions are not only due to the shape and respective position of both partners but also to their stereoelectronic properties, such as the electron density, polarizability, electrostatic potential, lipophilicity, and/or the proton accepting/donating character of the constituting chemical functions.[5] So, if we assume that compounds with similar stereoelectronic properties interact similarly with a given biological receptor and induce the same biological response,[6,7] the comparison of known active ligands can give useful information on their equivalent functional groups as well as on the corresponding specific interactions.[8−11] The results of such a similarity search can then be gathered in a pharmacophore model,[12] defined as the ensemble of the steric and electronic features necessary to trigger (or block) the biological response of a specific receptor.[13] The pharmacophore model may then be further used for identifying new leads and proposing new potential compounds to the organic chemists and pharmacologists.

Most of the existing pharmacophore elucidation methods are based on the consideration of the atomic positions and/or the use of physicochemical properties, such as the molecular electrostatic potential, evaluated at the atomic level.[11,13] Unfortunately, these methods perform poorly when comparing molecules that present the same affinity or activity profile but have completely different chemical functions. Thus it is logical to develop pharmacophore proposition methods based at a "functional group" level rather than at an atomic level.

A major contribution of the research described in this paper is to study the basic properties of electron density (ED) calculated at medium crystallographic resolution for a training set of 22 molecules in order to define chemical function-based reduced representations and to use these to propose pharmacophore models. This idea was implicit in Bader's work, "Atom in Molecules", which considered the topological properties of the ED at atomic resolution to help characterize intermolecular interactions.[14] Later approaches, based on the study of electron density maps (EDMs) at medium and low crystallographic resolution, were proposed in the Molecular Scene Analysis project to interpret experi-

* Corresponding author phone: 32 81 72 45 58; fax: 32 81 72 45 30; e-mail: nathalie.meurice@fundp.ac.be. F.N.R.S. Scientific Research Worker.
[†] Facultés Universitaires Notre-Dame de la Paix.
[‡] School of Computing, Queen's University.
[§] Department of Chemistry, Queen's University.

10.1021/ci034157x CCC: $27.50
Published on Web 06/23/2004

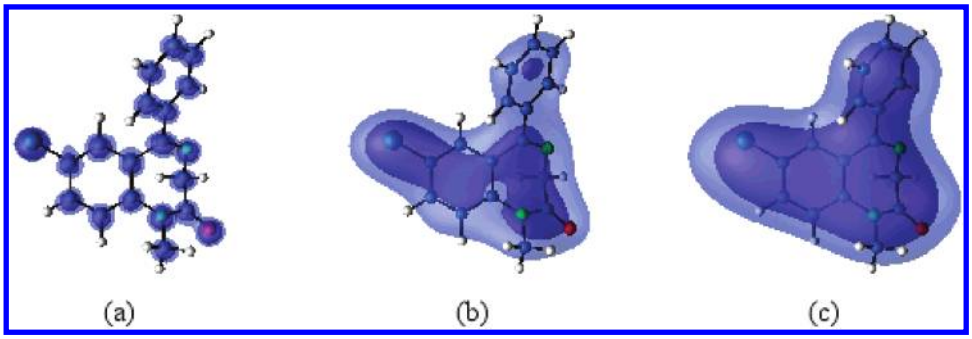

Figure 1. Electron density contours of diazepam calculated by XTAL at crystallographic resolutions of 1.0 Å (a), 3.0 Å (b), and 5.0 Å (c).

mental EDMs derived from X-ray diffraction data.[15−17] In this project, which applied techniques from artificial intelligence, EDMs are simplified into a set of critical points (CPs) using a topological analysis, according to a strategy developed by Johnson.[18] This topological approach was already successfully used in various applications including protein threading,[19] amino acid identification,[20] intermolecular interaction evaluation,[21,22] and molecular comparison.[23,24]

In the next sections, we introduce the generation of the EDMs and the topological analysis approach. We then describe how CPs can be associated with chemical functions. A detailed comparison of the CP representations of the various chemical functions at multiple levels of crystallographic resolution is presented in order to determine the optimal level of resolution, leading to the best function-based representation of biopharmacological molecules. Further steps in the total strategy are introduced in the conclusion.

ELECTRON DENSITY MAPS (EDMS)

EDMs of molecular structures can, for example, be determined experimentally by X-ray diffraction but can also be computed with the help of semiempirical or ab initio quantum mechanical methods. Because we wish to analyze and compare simultaneously a large number of EDMs at different crystallographic resolutions, we chose to generate them with the help of the XTAL program[25] [http://xtal.crystal.uwa.edu.au/]. To generate an EDM, XTAL needs the following parameters: the unit cell parameters, the space group, the fractional coordinates and the nature of the constituting atoms, a chosen crystallographic resolution and optionally, if one wants to reproduce experimental results, the atomic displacement parameters. With these data, XTAL computes the structure factors $F(\vec{h})$ of the given crystal structure at the selected resolution:

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

where f_j and B_j are the atomic scattering factor (tabulated[26]) and isotropic atomic displacement parameter of atom j, 2θ is the angle between the primary and the diffracted X-ray beam of wavelength λ, and $\vec{h}$ is a reciprocal space vector. The ED function can then be computed by taking the Fourier transform of $F(\vec{h})$:

$$\rho(\vec{r}) = \frac{1}{V} \sum_{\vec{h}=-\infty}^{+\infty} F(\vec{h}) \, e^{-2\pi i \vec{h}\cdot\vec{r}} \qquad (2)$$

where V is the volume of the unit cell.

According to Bragg's law, the crystallographic resolution level R of an EDM can be expressed as a function of the minimum distance d_{min} between two lattice planes in the reflection process:

$$R = \left(\frac{\sin\theta}{\lambda}\right)_{max} = \frac{1}{2d_{min}} \qquad (3)$$

At high or atomic resolution ($d_{min} = 1.0$ Å or $R = 0.500$), the maxima of the ED are located on the atomic positions, as one can expect from well resolved experimental EDMs (Figure 1a). As the resolution decreases to medium resolution ($d_{min} = 3.0$ Å or $R = 0.167$) or low resolution ($d_{min} = 5.0$ Å or $R = 0.100$), the ED functions are more broadly distributed, and the maxima of the ED function are less sharp and are no longer necessarily located on atomic positions (Figure 1b,c).

TOPOLOGICAL ANALYSES

Topological analyses can be applied to EDMs calculated at various crystallographic resolutions. Such analyses focus on particular features of the density distributions, such as local maxima, saddle points, and local minima. These are called "critical points" (CPs), i.e., the points where the ED gradient vanishes. The CP representation reduces the amount of data without losing significant information. The program ORCRIT,[18] used to perform these analyses, first locates the CPs and then builds a Hessian matrix $\mathbf{H}(\mathbf{r})$ based on the local second derivatives of the ED. $\mathbf{H(r)}$ can further be diagonalized:

$$\mathbf{H}(\mathbf{r}) = \begin{vmatrix} \frac{\partial^2\rho(\mathbf{r})}{\partial x^2} & \frac{\partial^2\rho(\mathbf{r})}{\partial x\partial y} & \frac{\partial^2\rho(\mathbf{r})}{\partial x\partial z} \\ \frac{\partial^2\rho(\mathbf{r})}{\partial y\partial x} & \frac{\partial^2\rho(\mathbf{r})}{\partial y^2} & \frac{\partial^2\rho(\mathbf{r})}{\partial y\partial z} \\ \frac{\partial^2\rho(\mathbf{r})}{\partial z\partial x} & \frac{\partial^2\rho(\mathbf{r})}{\partial z\partial y} & \frac{\partial^2\rho(\mathbf{r})}{\partial z^2} \end{vmatrix}$$

$$\mathbf{H}'(\mathbf{r}) = \begin{vmatrix} \frac{\partial^2\rho(\mathbf{r})}{\partial x'^2} & 0 & 0 \\ 0 & \frac{\partial^2\rho(\mathbf{r})}{\partial y'^2} & 0 \\ 0 & 0 & \frac{\partial^2\rho(\mathbf{r})}{\partial z'^2} \end{vmatrix} \qquad (4)$$

The three eigenvalues, i.e., the diagonal elements of $\mathbf{H}'(\mathbf{r})$, contain information about the shape of the CPs and

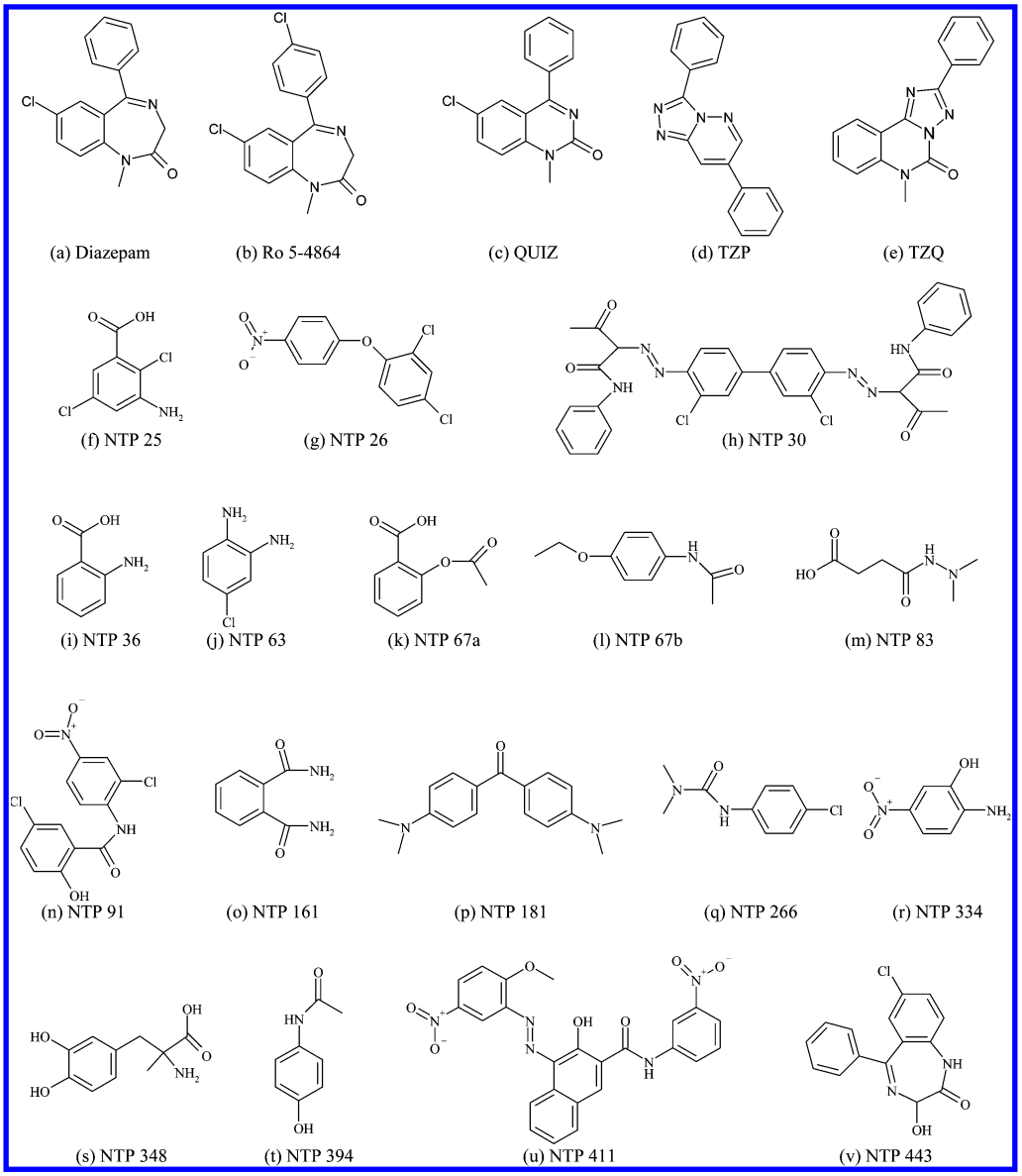

Figure 2. Training set of 22 molecules composed of 5 benzodiazepine-like molecules (a)−(e) and 17 molecules from the National Toxicity Project database (f)−(v).

can thus be used to determine the type of each CP. A point with 3 negative eigenvalues is a local maximum called a peak, 2 or 1 negative eigenvalues indicate a saddle point, called a pass or a pale, respectively, whereas a point with no negative eigenvalue is a local minimum or pit.

COMPARISON BETWEEN THE DISTRIBUTION OF THE CRITICAL POINTS OF VARIOUS FUNCTIONAL GROUPS AT DIFFERENT RESOLUTIONS

As mentioned in the Introduction, our objective is to determine a functional group-based reduced representation of biopharmalogical molecules in order to render the comparison of multiple molecules easier and faster. We based this representation on the CPs of the molecular ED function calculated at a nonatomic level of resolution. To achieve this goal, we studied the number and localization of the CPs on a training set of various molecules of pharmacological interest. This allowed us to determine which crystallographic resolution would best fit to the ideal situation where each of the different functional groups is associated with *one and only one* CP.

We first selected a series of 5 benzodiazepine-like compounds (Figure 2, molecules (a)−(e)) previously studied in our laboratory.[23,24] The structural data of molecules (a)−(b) were retrieved from the Cambridge Structural Database[27] (CSD), while the structures of molecules (c)−(e) were solved in collaboration with the "Laboratoire de Chimie Moléculaire Structurale" of the University of Namur. We then calculated their EDMs and generated the CPs at various crystallographic resolution levels ranging from 2.0 to 3.0 Å, at steps of 0.1 Å. We also considered additional resolutions above 3.0 Å, i.e., 3.2, 3.4, 3.6, 3.8, 4.0, 4.5, and 5.0 Å, to verify that the selected range was appropriate to preserve significant information. We finally enlarged the data set by considering more diversified chemical functions: we added 17 molecules from the National Toxicity Project Database (http://ntp-server.niehs.nih.gov/, Figure 2, molecules (f)−(v), 3D structures are available via the NTP database). These molecules have varying origins (e.g., pharmacologicals, pesticides, dyes, etc). By choosing this set, we avoided bias related to the lack of diversity and ensured the transferability of the CPs as well as their associated numerical descriptors,

Table 1. Electron Density Cutoff Values ($e^-/Å^3$) Used in the Topological Analysis of the Electron Density Maps Calculated at Crystallographic Resolutions Ranging from 2.0 to 2.6 Å for the Training Set of 22 Molecules

	crystallographic resolutions (Å)						
molecules	2.0	2.1	2.2	2.3	2.4	2.5	2.6
diazepam	1.5	1.3	1.2	1.1	1.1	1.5	1.0
Ro 5-4864	1.5	1.3	1.3	1.0	1.0	1.1	1.1
QUIZ	1.5	1.4	1.5	1.2	1.1	1.0	1.0
TZP	1.1	1.0	1.0	1.0	1.0	1.0	1.0
TZQ	1.2	1.2	1.0	1.0	1.0	1.0	1.0
NTP 25	1.5	1.3	1.5	1.5	1.5	1.3	1.3
NTP 26	1.3	1.2	1.3	1.0	1.0	1.0	1.0
NTP 30	1.5	1.5	1.5	1.0	1.0	1.0	1.0
NTP 36	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NTP 63	1.5	1.3	1.5	1.0	1.0	1.0	1.0
NTP 67A	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NTP 67B	1.3	1.0	1.1	1.0	1.0	1.0	1.0
NTP 83	1.3	1.1	1.3	1.3	1.0	1.3	1.0
NTP 91	1.2	1.0	1.1	1.0	1.0	1.0	1.0
NTP 161	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NTP 181	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NTP 266	1.7	1.7	1.7	1.3	1.2	1.5	1.3
NTP 334	1.2	1.2	1.2	1.0	1.0	1.0	1.0
NTP 348	1.2	1.3	1.2	1.0	1.0	1.0	1.0
NTP 394	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NTP 411	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NTP 443	1.1	1.0	1.0	1.0	1.0	1.0	1.0

to other possible data sets of organic compounds. EDMs and CPs were constructed for this set using XTAL and ORCRIT, respectively, at resolutions ranging from 2.0 to 3.0 Å at steps of 0.1 Å.

The simulated EDMs generated by XTAL contain background noise due to the Fourier transform truncation, especially for small density values. When performing the topological analysis, the background noise leads to irrelevant CPs, which are suppressed using a density cutoff based on the ED value at the CP positions. When analyzing the distribution of the ED values associated with all CPs in the 22 studied molecules, we always observed the presence of a significant gap between the highest values of the irrelevant CPs and the lowest relevant CP electron density values. As a consequence, we determined that a cutoff value of 1.0 $e^-/Å^3$ (the value in the middle of the gap) is appropriate for resolutions between 2.7 and 5.0 Å. At higher resolution (closer to the atomic level), the cutoff value was calculated case by case as presented in Table 1. We observed that higher cutoff values were set for the molecules containing chlorine atoms (e.g., diazepam, NTP 25 or NTP 266).

Before comparing the distribution of the CPs at the different resolution levels, we defined a minimum set of functional groups (Figure 3) to be assigned to each CP. To limit the number of functional groups that could be associated with more than one peak per function, we fixed their maximal size to 6 non-hydrogen atoms. As a consequence, a cyclic structure with 7 atoms or more was divided into smaller substructures, such as the imine and amide in the case of the diazepinone ring of diazepam (Figure 4). Because phenyl groups are a very frequent and important chemical function in molecular recognition phenomena, we decided to distinguish several possible substitutions. We also defined a general functional group for rings composed of 3 nitrogen and 2 carbon atoms because of the number of possible arrangements of atoms and bonds in such rings. The ethyl functional group is a priori not an "electronically" important

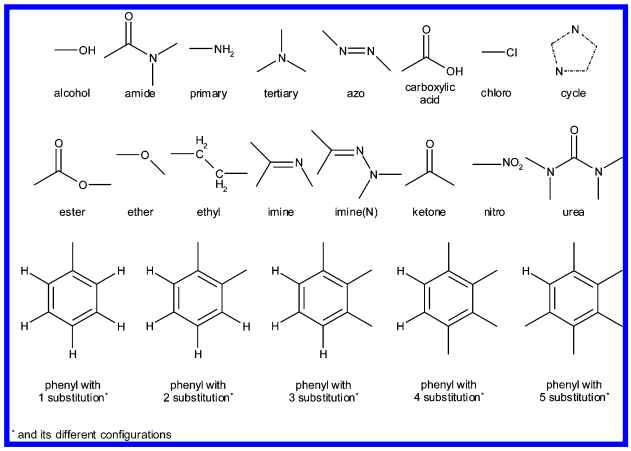

Figure 3. List of the chemical functions defined in order to study the distribution of the critical points of the electron density within the training set of 22 molecules.

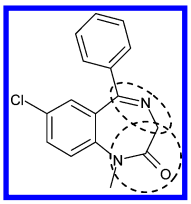

Figure 4. Division of the seven-membered atom ring of diazepam into amide (bottom circle) and imine moieties (top ellipsoid).

feature to consider, but, to completely describe all molecules, it was included in our list as a linker between the other chemical functions. On the whole, our training set of 22 molecules was divided into a total of 111 functional groups of 21 different types (Table 2).

The next step was to automate the assignment of each CP to one functional group. Therefore, we defined a centroid for each chemical moiety:

$$\mathbf{r}_{\text{centroid}} = \frac{\sum_{i=1}^{n} \mathbf{r}_i \cdot n_{\text{electron}(i)}}{\sum_{i=1}^{n} n_{\text{electron}(i)}} \qquad (5)$$

where $\mathbf{r}_i$ and $n_{\text{electron}(i)}$ are the coordinate vector and the number of electrons for atom i, respectively, and n is the number of atoms in the functional group. Then, we calculated the distance between the CPs and centroids and assigned the nearest functional group to each CP.

We plotted the mean distance separating the CPs of the 22 molecules from their nearest centroid as a function of the resolution used to calculate the EDMs and the associated peaks, passes, and pales (Table 3 and Figure 5).

First of all, we noticed that the number of pales is small in comparison to the numbers of peaks and passes. Indeed, the pales are mostly located in the cyclic moieties such as the phenyl groups and tend to be very close to the centroid of the cycle at high resolution. Because they are rare and always associated with the same functional groups, we decided not to consider them further in the reduced representations of the molecules. Second, we observed that the mean distance separating the passes from the centroid at all resolutions is around 1.17 ± 0.08 Å, which is comparable to the radius of a phenyl cycle; the peaks better match the

Table 2. Chemical Moieties Repartition in the Set of 22 Molecules[a]

	molecules																				
chemical moieties	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
alcohol														1				1	2	1	1	1
amide	1	1						2				1	1	1	2					1	1	1
amine (primary)						1			1	2								1	1			
amine (tertiary)													1			2						
azo								2													1	
carboxylic acid						1			1		1		1						1			
chloride	1	2	1			2	2	2		1				2			1					1
cycle (2C-3N)				1	1																	
ester											1											
ether							1					1									1	
ethyl												1	1						1			
imine	1	1																				1
imine (N)				1																		
ketone								2								1						
nitro							1							1				1			2	
phenyl − 1 substitution	1		1	2	1			2														1
phenyl − 2 substitutions		1			1		2		1		1	1			1	2	1			1	2	
phenyl − 3 substitutions	1	1	1				1	2		1				2				1	1		1	1
phenyl − 4 substitutions						1																
phenyl − 5 substitutions																					1	
urea			1		1												1					

[a] Molecules are as follows: (a) diazepam, (b) Ro-5(4864), (c) QUIZ, (d) TZP, (e) TZQ, (f) NTP 25, (g) NTP 26, (h) NTP 30, (i) NTP 36, (j) NTP 63, (k) NTP 67a, (l) NTP 67b, (m) NTP 83, (n) NTP 91, (o) NTP 161, (p) NTP 181, (q) NTP 266, (r) NTP 334, (s) NTP 348, (t) NTP 394, (u) NTP 411, and (v) NTP 443.

Table 3. Mean Distance (Å) between the Critical Points of the Electron Density and the Centroids of the Chemical Functions and Number of Peaks, Passes, and Pales in the 22 Molecules of the Training Set at Crystallographic Resolutions from 2.0 to 3.0 Å

	mean distance from nearest centroid (Å)			number of critical points		
resolution (Å)	peak	pass	pale	peak	pass	pale
2.0	1.04	1.21	0.24	201	175	16
2.1	1.02	1.25	0.24	194	170	27
2.2	0.95	1.19	0.43	165	143	31
2.3	0.87	1.14	0.58	148	128	28
2.4	0.78	1.11	0.71	142	121	21
2.5	0.66	1.10	0.86	124	98	11
2.6	0.56	1.16	1.18	110	82	9
2.7	0.47	1.16	1.34	98	77	9
2.8	0.49	1.09	1.23	95	70	8
2.9	0.47	1.17	1.47	91	67	4
3.0	0.48	1.10	1.53	80	53	2

centroids of the corresponding functions as the resolution decreases. We finally noted, as shown in Figure 6, that the peaks tend to be located on the functional groups, with the passes logically located between them. With these observations in mind, we decided to focus our study on the distribution of CPs which best depicted the functional groups, i.e., the peaks.

The distribution of the peaks over all molecules was analyzed and compared to identify the optimal crystallographic resolution for a chemical function-based reduced representation. For the set of 22 molecules, we calculated the percentage of functional groups associated with exactly one peak for each function type. This was done at resolution levels ranging from 2.0 to 3.0 Å for the 111 chemical moieties (Table 4). The same analysis was carried out at resolution levels ranging from 3.2 to 5.0 Å for the 23 functions constituting the 5 benzodiazepine-type ligands (Table 5). We observed that the best results (85% and higher) were obtained in the 2.6−2.8 Å range. As the resolution becomes lower than 2.6 Å or greater than 2.8 Å, the percentage of functions that correspond to single CPs decreases dramatically (Figure 7).

At higher resolutions (∼2.0 Å), the ED of the large functional groups (e.g., phenyl ring) and those with more than one oxygen or nitrogen atom (e.g., carboxylic acid, ester, or urea) is associated with more than one peak. Figure 8, wherein we compare the distribution of peaks on the phenyl rings of TZP at crystallographic resolutions ranging from 2.0 to 2.9 Å, illustrates this situation. As the resolution increases, the number of peaks corresponding to the two phenyl groups logically increases, and their distributions tend to the results of Bader's approach, that is to say, one peak at each atom position.[14] As noted already in Figure 5, the peaks tend to be closer to the centroid of the functional group as the resolution decreases (Figure 8). At lower resolution (>3.0 Å), the problem is reversed: with the loss of detailed information, some functional groups are no longer associated with any peak. This trend is illustrated in Figure 9 for Ro 5-4865 and TZQ at resolutions of 2.6 and 3.2 Å. Here, the imine moiety and the benzo ring in Ro 5-4864 lose their associated peak, as does the urea in TZQ.

Finding the optimal crystallographic resolution required determining the appropriate compromise between the two phenomena. Best results were obtained at resolution levels of 2.6 and 2.7 Å, and if we neglect the CP on the ethyl groups (the linker moieties), both levels led to ∼90% of the functional groups being associated with a single CP. When examining the incorrect 10%, we noted differences between the two resolutions. At 2.6 Å, the deviations come from the urea and the phenyl ring with 3 substitutions which were associated with two peaks, while these same functional groups correspond to one peak at 2.7 Å. At this resolution, however, some of the alcohols, as well as the phenyl rings with 1 and 5 substitutions, which were correctly associated with one peak at 2.6 Å, are no longer associated with a single peak. Given that it is preferable for a functional group to be associated with two peaks rather than no peaks, we selected

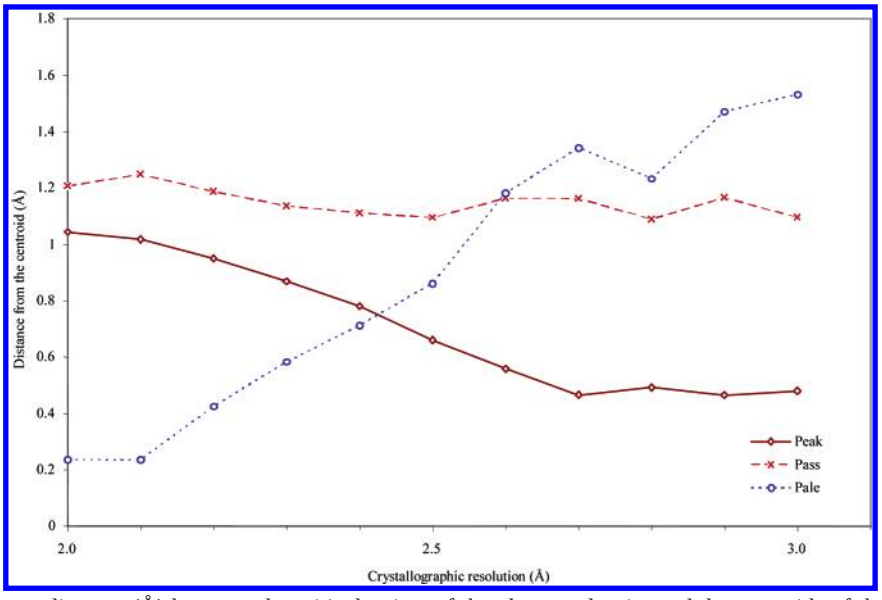

Figure 5. Plot of the mean distance (Å) between the critical points of the electron density and the centroids of the chemical functions for the peaks, passes, and pales in the 22 molecules of the training set, as a function of the crystallographic resolution (Å).

Table 4. Proportion (in %) of the Functional Groups Associated with One and Only One Peak of the Electron Density by Chemical Moiety Type for the 22 Molecules of the Training Set at Varying Crystallographic Resolutions

		crystallographic resolution (Å)										
functional group	occurrence	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
alcohol	7	100	100	100	100	100	100	100	86	43	57	14
amide	12	25	33	50	50	75	100	100	100	100	100	100
amine (primary)	6	100	100	83	100	100	83	83	83	67	67	
amine (tertiary)	3	33	33	100	100	100	100	100	100	100	100	100
azo	3	67	100	100	100	100	100	100	100	100	100	33
carboxylic acid	5	40	40	100	100	100	100	100	100	100	100	100
chloride	15	100	100	100	100	100	100	100	100	100	100	100
cycle (2C-3N)	2	100	100	100	100	100	100	100	100	100	100	100
ester	1						100	100	100	100	100	100
ether	3	100	100	100	100	100	100	100	100	100	100	67
ethyl	3	67	33	67		33	33	33				
imine	3	100	100	100	100	100	100	100	100	100	100	100
imine (N)	1				100	100	100	100	100	100		
ketone	3	33	33	67	100	100	100	100	100	100	67	33
nitro	5	80	100	100	100	100	100	100	100	100	100	100
phenyl − 1 substitution	8					13	25	100	75	88	88	100
phenyl − 2 substitutions	13				23	38	62	92	85	77	69	77
phenyl − 3 substitutions	13			23	38	38	62	69	85	69	77	46
phenyl − 4 substitutions	1			100	100	100	100					
phenyl − 5 substitutions	1					100	100	100		100		
urea	3				33		33	33	100	100	100	100
total	111	46	48	59	66	72	82	91	89	85	83	70

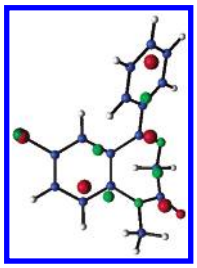

Figure 6. Critical points graph of diazepam at a crystallographic resolution of 2.6 Å. Peaks are in red and passes in green.

the crystallographic resolution of 2.6 Å as optimal for generating the CP graphs.

CONCLUSIONS

The primary objective of our work was to design a reduced representation for molecules that could be applied in pharmacophore proposition methods. Our approach was based on the topological properties of the electron density (ED) at medium or low crystallographic resolution. We selected a training set of 22 molecules of pharmacological interest and calculated their ED maps at crystallographic resolutions ranging from 2.0 to 5.0 Å using the program XTAL. Critical point (CP) representations were then obtained through a topological analysis using the ORCRIT algorithm. The distributions of the CPs were analyzed and compared, revealing that the CPs of the ED at a resolution of 2.6 Å present the highest number of functional groups associated with one and only one peak of the ED. Consequently, we confirmed that a representation in terms of CPs at 2.6 Å is a valuable "functional group" based representation of pharmacological molecules. This study constitutes the initial step in the development of a novel pharmacophore elucidation strategy.

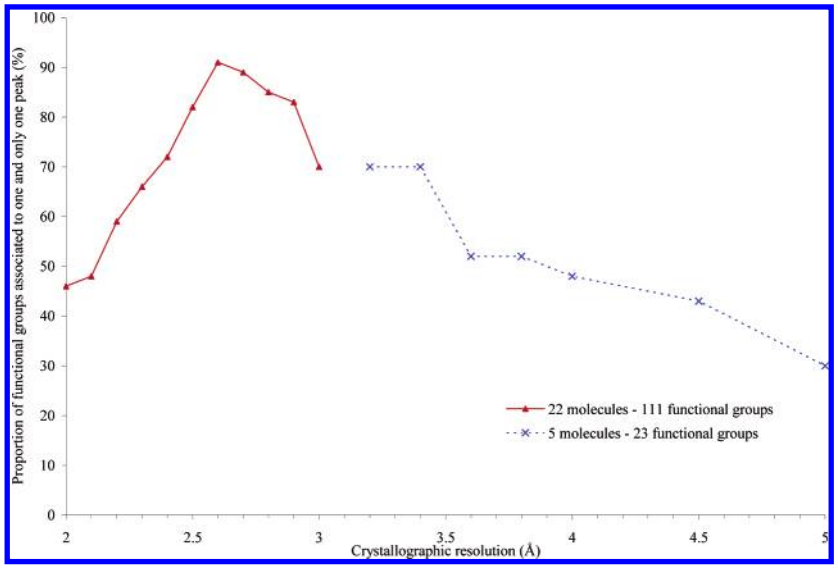

Figure 7. Plot of the proportion (in %) of the functional groups associated with one and only one peak of the electron density by chemical moiety type as a function of the crystallographic resolution (Å) for the 22 molecules of the training set.

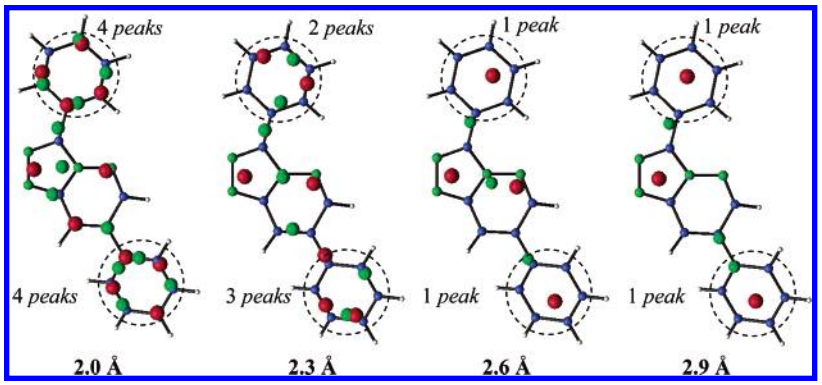

Figure 8. Critical point graphs of TZP at crystallographic resolutions of 2.0, 2.3, 2.6, and 2.9 Å. Peaks are in red and passes in green; circles mark phenyl groups.

Table 5. Proportion (in %) of the Functional Groups Associated with One and Only One Peak of the Electron Density by Chemical Moiety Type for the 5 Benzodiazepine-like Molecules at Varying Crystallographic Resolutions

		crystallographic resolution (Å)						
functional group	occurrence	3.2	3.4	3.6	3.8	4.0	4.5	5.0
amide	2	100	100			100		
chloride	4	100	100	100	100			
cycle (2C-3N)	2	100	100	100	100	50	50	50
imine	2	50	50				50	50
imine (N)	1							
phenyl − 1 substitution	5	100	100	80	80	100	100	60
phenyl − 2 substitutions	2	50	50	50	50	50		
phenyl − 3 substitutions	3						67	33
urea	2	50	50	50	50	100		50
total	23	70	70	52	52	48	43	30

Presently, we are associating each CP with several numerical properties to provide a more detailed representation of the molecules. These descriptors include stereoelectronic information on the local shape of the CPs as well as on their neighborhood. Selected local descriptors for a CP include the following: (i) the value of the ED at the point position, (ii) a measurement of the volume of the CP,[22] and (iii) the 3 eigenvalues of the Hessian matrix for the CP. Selected neighborhood descriptors include the distance and the local properties of the nearest CPs. At present, a statistical analysis of the values of each descriptor is carried out for a large set of molecules selected from the Cambridge Structural Database.[27] An objective of our ongoing research is to carry out

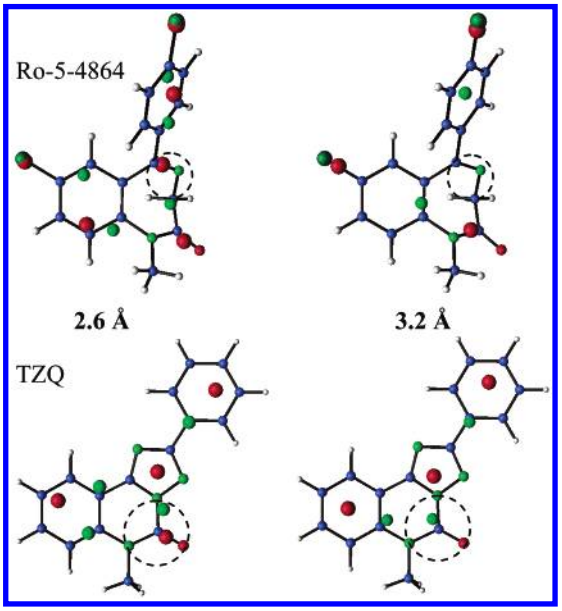

Figure 9. Critical point graphs of Ro-5-4864 and TZQ at crystallographic resolutions of 2.6 and 3.2 Å. Peaks are in red and passes in green; circles mark imine groups for Ro-5-4864 and urea for TZQ.

a classification analysis to regroup the chemical functions associated with the CPs with similar descriptors and compare the results to isosteric fragments, i.e., structurally different groups that can be replaced in pharmacological molecules without changing their activity.[28] A second objective is to define dependencies between the descriptors and the chemical function types. To discover and, if possible, improve such relationships, we plan to use self-organizing maps and case-based reasoning.[29,30] Finally, we will combine the numerical relations with conceptual chemical relations to implement a pharmacophore proposition procedure using Inductive Logic Programming[20,31,32] (ILP). Such an approach has previously been used for pharmacophore proposition at an atomic level of representation.[33]

ACKNOWLEDGMENT

All authors thank C.K. Johnson for providing the program ORCRIT. The calculations have been performed on the Interuniversity Scientific Computing Facility (ISCF), installed at the "Facultés Universitaires Notre-Dame de la Paix", for which the authors gratefully acknowledge the financial support of the FNRS-FRFC and the Loterie Nationale for the convention n*o*. 2.4578.02. J.B. wants to thank the "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture" for his Ph.D. fellowship. N.M. thanks the "Fonds National de la Recherche Scientifique" for her Scientific Research Worker position.

REFERENCES AND NOTES

(1) Boyd, D. B. Drug Design. In *Encyclopedia of Computational Chemistry*; von Ragué Schleyer, P., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; pp 795−804.

(2) Willett, P. Searching for Pharmacophoric Patterns in Databases of Three-Dimensional Chemical Structures. *J. Mol. Recognit*. **1995**, *8*, 290−303.

(3) Oshiro, C. M.; Kuntz, I. D.; Knegtel, R. M. A. Molecular Docking and Structure-Based Design. In *Encyclopedia of Computational Chemistry*; von Ragué Schleyer, P., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; pp 1606−1613.

(4) Murcko, M. A. Recent Advances in Ligand Design Methods. In *Reviews in Computational Chemistry − Volume 11*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, U.S.A., 1997; pp 1−66.

(5) Rohrer, D. C. 3D Molecular Similarity Modelling in Computational Drug Design. In *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*; Carbò-Dorca, R., Ed.; Kluwer Academic: Dordrecht, The Netherlands, 1995; pp 141−161.

(6) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078−1082.

(7) Itai, A.; Tomoïka, N.; Yamada, M.; Inoue, A.; Kato, Y. *Molecular Superposition for Rational Drug Design. In 3D QSAR in Drug Design;* Kubinyi H., Ed.; Kluwer/Escom: Dordrecht, The Netherlands, 1993; pp 200−225.

(8) Frühbeis, H.; Klein, R.; Wallmeier, H. Computer-Aided Molecular Design (CAMD) − An Overview. *Angew. Chem., Int. Ed. Engl.* **1987**, *26*, 403−418.

(9) Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; Wiley: New York, U.S.A., 1990.

(10) Dean, P. M. Defining Molecular Similarity and Complementarity for Drug Design. In *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Blackie Academic & Professional: London, Great Britain, 1995; pp 1−23.

(11) Lemmen, C.; Lengauer, T. Computational Methods for the Structural Alignment of Molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.

(12) Milne, G. W. A. Pharmacophore and Drug Discovery. In *Encyclopedia of Computational Chemistry*; von Ragué Schleyer, P., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; Wiley: Chichester, 1997; pp 2046−2056.

(13) Wermuth, C. G.; Langer, T. Pharmacophore Identification. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; Escom: Leiden, The Netherlands, 1993; pp 117−136.

(14) Bader, R. F. W. *Atoms in Molecules − A Quantum Theory*; Clarendon Press: Oxford, Great Britain, 1995.

(15) Fortier, S.; Castleden, I.; Glasgow, J.; Conklin, D.; Walmsley, C.; Leherte, L.; Allen, F. H. Molecular Scene Analysis: the Integration of Direct-Methods and Artificial-Intelligence Strategies for Solving Protein Crystal Structures. *Acta Crystallogr. D* **1993**, *49*, 168−178.

(16) 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.

(17) Leherte, L.; Glasgow, J.; Baxter, K.; Steeg, E.; Fortier, S. Analysis of Three-Dimensional Protein Images. *J. Artif. Intelligence Res.* **1997**, *7*, 125−159.

(18) Johnson, C. K. *The Oak Ridge Critical Point Network Program*; Oak Ridge National Laboratory: Oak Ridge, 1977.

(19) Ableson, A.; Glasgow, J. Crystallographic Threading. In *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology (ISMB99)*; Lengauer, T., Schneider, R., Bork, P., Brutlag, D., Glasgow, J., Mewes, H.-W., Zimmer, R, Eds.; AAAI Press: New York, U.S.A., 1999; pp 2−9.

(20) Whelan, K.; Glasgow, J. Identifying Amino Acid Residues in Medium Resolution Critical Point Graphs Using Instance Based Query Generation. *Pac. Symp. Biocomputing* **2000**, *5*, 401−412.

(21) Leherte, L.; Latour, T.; Vercauteren, D. P. Topological Analysis of Electron Density Maps of Chiral Cyclodextrin-Guest Complexes: A Steric Interaction Evaluation. *Supramol. Sci.* **1995**, *2*, 209−217.

(22) 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.

(23) 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.

(24) Leherte, L.; Meurice, N.; Vercauteren, D. P. Application of Multiresolution Analyses to Electron Density Maps: A Critical Point Analysis Approach for the Comparison of Molecules. In *Proceedings of the WSES/MIUE/HNA International Conference on Mathematics and Computers in Biology and Chemistry (MCBC 2000)*; Mastorakis, N. E., Ed.; WSES Press: Jamaica, 2000; pp 158−164.

(25) Hall, S. R.; du Boulay, D. J.; Olthof-Hazekamp, R. *Xtal3.7 System;* University of Western Australia: Crawley, 2000.

(26) Wilson, A. J. C. *International Table for Crystallography, Volume C: Mathematical, Physical, and Chemical Tables*; Kluwer Academic: Dordrecht, The Netherlands, 1992.

(27) Allen, F. H. The Cambridge Structural Database: a Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr. B* **2002**, *58*, 380−388.

(28) Sheridan, R. P. The Most Common Chemical Replacements in Drug-Like Coumpounds. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 103−108.

(29) Leake, D. *Case-Based Reasoning: Experiences, Lessons, and Future Directions*; AAAI Press: Menlo Park, CA, 1996.

(30) Jurisica, I.; Rogers, P.; Glasgow, J.; Collins, R. J.; Wolfley, J. R.;. Luft, J. R.; DeTitta G. T. Improving Objectivity Scalability in Protein Crystallization: Integrating Image Analysis with Knowledge Discovery. *IEEE Intelligent Syst.* **2001**, *16*, 26−34.

(31) King, R. D.; Muggleton, S.; Lewis, R. A.; Sternberg, M. J. E. Drug Design by Machine Learning: the Use of Inductive Logic Programming to Model the Structure-Activity Relationships of Trimethoprim Analogues Binding to Dihydrofolate Reductase. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 11322−11326.

(32) Dehaspe, L.; Toivonen, H.; King, R. D. Finding Frequent Substructures in Chemical Compounds. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD-98)*; Agrawal, R., Stolorz, P., Piatetsky-Shapiro, G., Eds.; AAAI Press: New York, 1998; pp 30−36.

(33) Finn, P.; Muggleton, S.; Page, D.; Srinivasan, A. Pharmacophore Discovery Using the Inductive Logic Programming System Progol. *Machine Learning* **1998**, *30*, 241−271.

CI034157X