J Mol Model (2005) 11: 237–247
DOI 10.1007/s00894-005-0264-7

ORIGINAL PAPER

Nathalie Meurice · Gerald M. Maggiora
Daniel P. Vercauteren

Evaluating molecular similarity using reduced representations of the electron density

Received: 23 December 2004 / Accepted: 3 March 2005 / Published online: 12 May 2005

Abstract A model system of four benzodiazepine-like ligands for the central benzodiazepine receptors (CBRs) and peripheral benzodiazepine receptors (PBRs)is examined using a genetic algorithm procedure (GAGS) designed for evaluating molecular similarity. The method is based on the alignment of reduced representations generated from the critical points of the electron density computed at medium crystallographic resolution. The results are further characterized by a comparison with alignments produced by MIMIC, a field-based superimposition method that matches both steric and electrostatic molecular fields. The alignments produced by the two methods are generally seen to be consistent. The relationships of the compounds' binding affinities for both CBRs and PBRs to the alignments determined by GAGS yield a set of structural features required for significant binding to benzodiazepine receptors. Benefits of using reduced representations for evaluating molecular similarities and for constructing pharmacophore models are discussed.

N. Meurice · G. M. Maggiora
Department of Pharmacology and Toxicology,
University of Arizona, College of Pharmacy,
1703 E. Mable, Tucson,
AZ 85721, USA

N. Meurice (✉) · D. P. Vercauteren
Laboratoire de Physico-Chimie Informatique,
Facultés Universitaires Notre-Dame de la Paix,
61 Rue de Bruxelles, B-5000 Namur,
Belgium
E-mail: meurice@pharmacy.arizona.edu
Tel.: +1-520-626-0974
Fax: +1-520-626-2466

G. M. Maggiora
BIO5 Institute, University of Arizona,
Tucson, AZ 85721, USA

N. Meurice
T.N.R.S. Scientific Research Worker

Keywords Molecular similarity · Molecular alignment · Critical point graphs · Genetic algorithms · Peripheral benzodiazepine receptors

Introduction

When the three-dimensional (3D) structure of a receptor is unavailable or is not precisely known, molecular alignments based on the similarities of ligands active for the receptor can provide insight into possible binding modes of the ligands. Moreover, the alignments can provide a basis for the construction of pharmacophore models that group specific stereoelectronic features of the aligned molecules corresponding to important ligand–receptor interactions. In this context, simultaneous alignment of several ligands based on reduced representations of 3D molecular-property distributions may be particularly appropriate, especially when considering structurally dissimilar ligands. Therefore, a genetic algorithm procedure, called GAGS, was developed based on the use of reduced representations of the electron density to simultaneously align sets of molecules. The reduced representations are generated from the critical points (CPs) of the electron density computed at medium crystallographic resolution [1–5]. To validate and assess the usefulness of using reduced representations to evaluate molecular similarities, a "model system" of four benzodiazepine-like molecules is first investigated using the GAGS approach. The results obtained by GAGS are then compared with the molecular alignments obtained with MIMIC [6–12], a field-based method that aligns both the steric and electrostatic fields of molecules.

Molecules of the benzodiazepine family have been of pharmacological interest since scientists discovered their sedative, anxiolytic, myorelaxant, and anti-convulsive effects. In particular, compounds containing the 1,4-benzodiazepin-2-one moiety, which are widely

available on the pharmaceutical market, are used for the treatment of anxiety and sleep disorders and for moderation of depression. Numerous studies have thus been undertaken since the 1970s in an effort to gain a better understanding of the molecular mechanism by which these drugs interact with their receptors and how they produce their therapeutic effects.

Early studies have shown the existence of specific binding sites on $GABA_A$ receptors, which are a part of the GABA/chloride ionophore complex, found in the central nervous system; thus they are called central benzodiazepine receptors (CBRs) [13–20]. Later studies revealed the existence of an additional benzodiazepine binding site, which is mainly located in the periphery, more precisely in the mitochondrial membranes of steroidogenic cells [21–24]. Due to their tissue location, these receptors are called peripheral benzodiazepine receptors (PBRs) or mitochondrial benzodiazepine receptors. PBR proteins are highly hydrophobic and possess five membrane-spanning helices located in the outer membrane of mitochondria. It is part of the mitochondrial permeability transition pore, a trimeric complex located on both the outer and inner mitochondrial membranes [23, 25]. Even though their physiological function remains somewhat unclear, PBRs have been considered as possible targets, for example for apoptosis, inflammation, immunomodulation, and cell proliferation [26, 27]. In particular, PBRs are involved in intra-mitochondrial cholesterol transport towards the inner mitochondrial membrane, where the conversion of cholesterol into pregnenolone is catalyzed by cytochrome P450 [23, 25, 28, 29].

This paper focuses on a computational study of four benzodiazepine-like compounds (BZs) [30], which have been shown to bind to both CBRs and PBRs by Bourguignon and co-workers [31]. As the 3D structure of their binding sites is still unknown, direct modeling of the interactions involved in binding is not possible. However, an alternative approach is possible using similarity-based molecular alignment of the selected compounds, but the alignment method must be up to the task. The present study involves structurally different molecules that can be difficult to align using only atomic and geometric criteria. In contrast, the GAGS procedure used here is based on the entire electron density, albeit a reduced representation of it, and thus is not constrained to atoms. This provides a more comprehensive stereoelectronic picture of the molecules that should be suitable for the proposed work.

In the following sections, the alignment results of a study of selected BZs using the GAGS approach are presented and compared to alignment results obtained using the field-based similarity method MIMIC. As the comparison will show, both methods yield very similar results, providing a measure of validation of the reduced-representation GAGS procedure and, in addition, showing the suitability of this strategy for determining molecular similarities. Lastly, the molecular alignments produced by GAGS are used as a basis for a structure–activity study that highlights the salient stereo-electronic features responsible for binding to both CBRs and PBRs.

Materials and methods

Benzodiazepine ligands

We selected four benzodiazepine-like molecules that bind, in varying degrees, to both CBRs and PBRs. Among the chosen molecules are two benzodiazepines, diazepam [1] and Ro 5-4864, [2] as well as two structurally related compounds, a quinazolinone (Quiz [3]) and a triazoloquinazolinone (Tzq [4]), whose planar structural formulae and binding affinities are given in Fig. 1 and Table 1, respectively. At first glance, it is clear that the presence or absence of a chlorine atom on the phenyl moiety has a direct influence on the affinity of the compounds, Ro 5-4864 being more selective for PBRs. Second, removing the methylene group from a benzodiazepine leads to quinazolinone derivatives such as Quiz, which possesses a totally planar heterocyclic ring and has a binding affinity in the micromolar range. Third, replacement of the Quiz imine with a triazole function leads to triazoloquinazolinone derivatives, such as Tzq, that lose significant affinity for CBR and become completely insensitive for PBR. The non-charged, quasi-planar, and nearly rigid structures of these molecules make them ideal candidates for study.

Because the molecules are considered to be rigid and because the GAGS alignment procedure deals with critical point graphs computed at medium crystallographic resolution, which is somewhat insensitive to slight conformational changes, a unique conformation is chosen for each compound, namely its crystalline structure. Atomic coordinates of both diazepam and Ro 5-4864 are available in the Cambridge Structural Database (CSD) [32], with refcodes DIZPAM10 [33] and FULWUE [34], respectively. Since the 3D crystalline structures of Quiz and Tzq are unavailable in the CSD, they were synthesized and crystallized by our colleagues in the Laboratoire de Pharmacochimie de la Communication Cellulaire of the Louis Pasteur University in Illkirch (France). The structures were subsequently solved by X-ray diffraction in collaboration with the Laboratoire de Chimie

[1] 7-Chloro-1,3-dihydro-1-methyl-5-(phenyl)-2H-1,4-benzodiazepine-2-one.
[2] 7-Chloro-1,3-dihydro-1-methyl-5-(p-chlorophenyl)-2H-1,4-benzodiazepine-2-one.
[3] 6-Chloro-N1-methyl-4-phenylquinazoline-2-one.
[4] [1, 2,4]Triazolo[1,5-c]quinazoline-5(6H)-one.

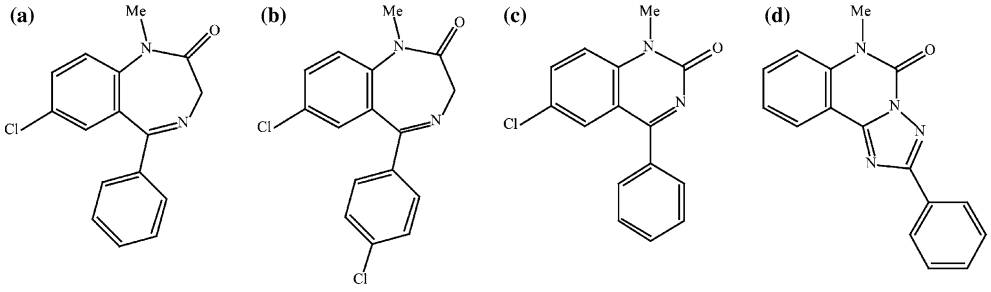

Fig. 1 Structural formulae of: (**a**) diazepam, (**b**) Ro 5-4864, (**c**) Quiz, and (**d**) Tzq

Moléculaire Structurale at the University of Namur (FUNDP) [35].

Reduced representations of the electron density

Electron-density (ED) distributions contain all of the stereoelectronic information in molecules and thus can be used as the basis of many shape-matching methods. The method described in this work uses a reduced representation based on critical points (CPs) of the ED distribution as a means of capturing much of the salient information in the distribution in simpler form [36]. Approximate ED distributions are obtained by computing the ED on 3D grids at medium crystallographic resolution using the program XTAL [3, 37]. The resulting grids are further simplified by a topological analysis of their CPs—points where the first derivatives of the ED distributions are zero—carried out with an improved version of the ORCRIT program [38–40] developed in the ''Laboratoire de Physico-Chimie Informatique'' of the FUNDP.

The topological analysis is implemented in two steps. First, the location of the CPs is determined, followed by a determination, using second partial derivatives, of their nature, namely, whether they correspond to local maxima (peaks), local minima (pits), or saddle points (passes or pales). Such an approach considerably reduces the amount of data in an ED distribution while retaining information on its invariant geometric features. In typical applications a subset of n CPs, computed at a given resolution, is used. These CPs are linked together in a 3D graph-like structure called a CP graph that is akin to a molecular graph, except that all of its nodes (i.e., CPs) are fully connected—it is a complete graph.

The nodal and connectivity properties of each molecule being investigated are summarized into a *symmetric property matrix* P_L:

Table 1 Binding affinity values of diazepam, Ro 5-4864, Quiz, and Tzq for CBR and PBR [31]

Molecule	IC_{50}(CBR) (nM)	IC_{50}(PBR) (nM)
Diazepam (D)	6	79
Ro 5-4864 (R)	5 000	6
Quiz (Q)	~1 000	~1 000
Tzq (T)	Not significant	2 800

$$P_L = \begin{pmatrix} \rho_1 & \cdots & d_{1i} & \cdots & d_{1n} \\ & \ddots & & & \vdots \\ & & \rho_i & & d_{in} \\ & & & \ddots & \vdots \\ & & & & \rho_n \end{pmatrix} \quad (1)$$

where ρ_i corresponds to the ED at the ith CP and d_{ij} corresponds to the distance of separation between the ith and jth CPs. Simultaneous alignment of the molecules then requires that their CP graphs be superimposed to ensure that comparable CPs, their ED values, and their pairwise distances over the set of molecules match in some optimal way (*vide infra*).

Construction of 3D molecular graphs

Each of the four compounds under study is centered in a fictitious rectangular unit cell (P1 symmetry) of dimensions 16 Å× 8 Å× 16 Å, which is large enough to avoid boundary effects, with a grid spacing of 0.25 Å. The ED maps are calculated with XTAL at a medium crystallographic resolution of 2.5 Å in order to describe the stereo-electronic properties of the molecules at the functional group level [36]. Topological analysis of the resulting ED maps then yields 3D molecular graphs whose nodes are given by the CPs of the ED distribution (Fig. 2). Only CPs with ED values above 1.5 $e^-/Å^3$ are considered, generating molecular graphs with 13, 17, 15, and 13 nodes for diazepam, Ro 5-4864, Quiz, and Tzq, respectively. Among these nodes, CPs corresponding to local maxima (peaks) are not strictly located at atom positions, except for chlorine atoms, but are associated with ''chemical functions'' such as phenyl rings, which are characterized by one or two peaks. The benzodiazepine heterocycle or its equivalent is represented by four peaks located on the imine, carbonyl groups, and benzo ring, respectively, with passes (second-order saddle points) connecting each pair of peaks and a low-density region corresponding to a pale (first-order saddle point) lying approximately at the center of the four peaks. As each pass is systematically located between two peaks, we have considered only peaks and pales in the molecular alignments. Thus, the final number of CPs in each of the four molecular graphs is 7, 9, 8, and 7, with respect to diazepam, Ro 5-4864, Quiz, and Tzq. As will be

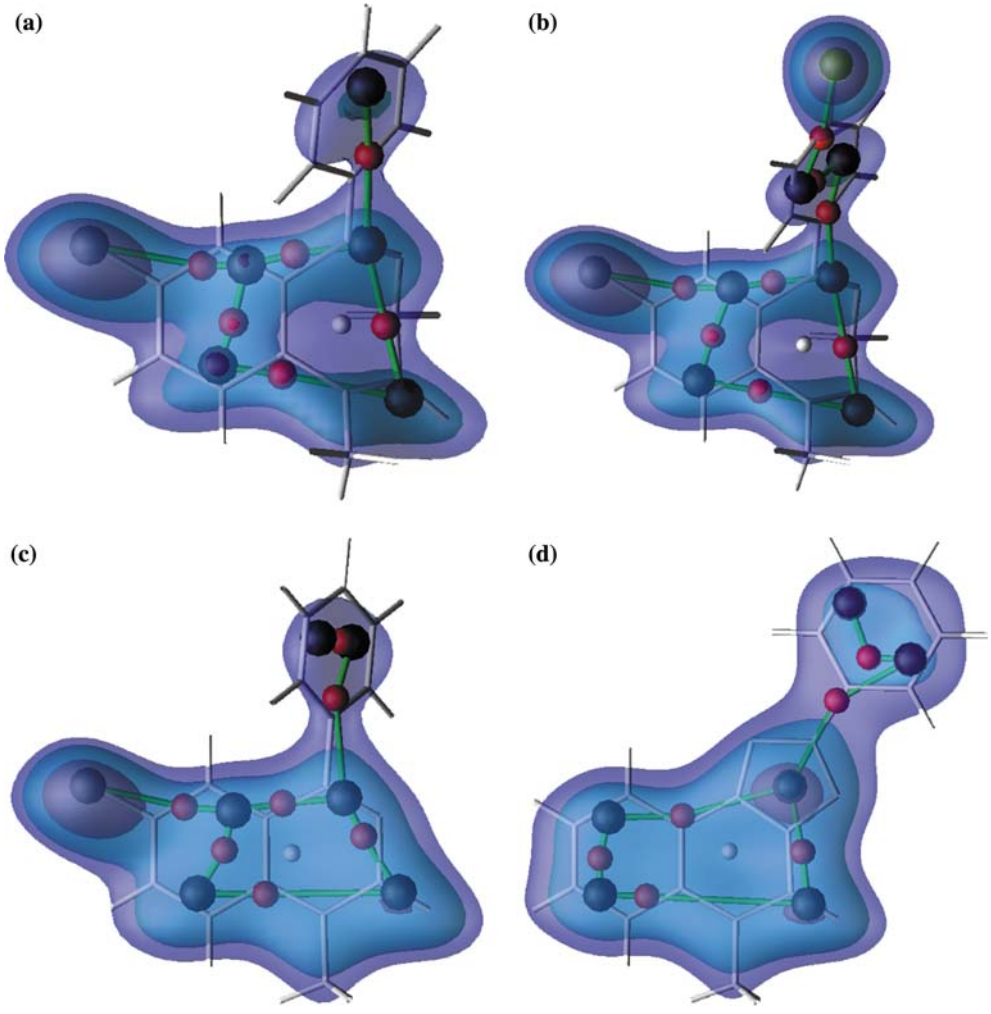

Fig. 2 Peaks (*big black spheres*), passes (*medium red spheres*), and pales (*small white spheres*) of the electron density calculated with XTAL at a crystallographic resolution level of 2.5 Å for: (**a**) diazepam, (**b**) Ro 5-4864, (**c**) Quiz, and (**d**) Tzq, superimposed to the heavy atoms skeleton and to the isodensity surfaces of 1.5, 2.0, and 3.0 $e^{-}/Å^{3}$ (*blue surfaces*)

seen in the sequel, the peaks may be seen as molecular features, namely hydrophobic moieties and H-bond acceptors, that are important for binding to CBRs and PBRs.

Graph similarity searching—the GAGS approach

Genetic algorithms (GAs) are artificial intelligence optimization techniques that find their origin in the Darwinian theory of evolution, combining the principles of ''struggle for life'' and ''survival of the fittest chromosomes'' [41–43]. Basically, GAs require an appropriate coding of the problem in the form of a chromosome, a fitness function to evaluate the suitability of a given chromosome, and several types of ''genetic operators'' that carry out ''crossovers'' between pairs of chromosomes and ''mutations'' of individual chromosomes. Initially, a random population of chromosomes is generated, each of whose fitness is evaluated, and a subset of the fittest is retained. Chromosomal crossover and mutations are then carried out, the fitness of the modified chromosomes is evaluated and the fittest chromosomes are again retained. This process is continued until the process is considered to be converged.

Using these concepts, a novel GA-based strategy for matching the CP graphs was developed [1–5]. More particularly, the procedure, called Genetic Algorithm for Graph Similarity search (GAGS), was developed for the purpose of carrying out simultaneous similarity-based alignments of multiple molecules based on their CP graphs derived from reduced representations of their ED distributions. Coding of the chromosomes was thus designed specifically for matching CP graphs—each chromosome is represented by an $n \times m$ table, where m is the number of molecules being aligned and n is the smallest number of CPs found in any of those m molecules (Fig. 3). Each cell of the table corresponds to a specific CP of a given molecule.

The fitness function (*Fit*) is the sum of two weighted contributions: one associated with the CP densities (T_R) and the other associated with the CP inter-distances (T_D). Their respective normalization factors are N_R and N_D, the corresponding weights being W_R and W_D. The global form of *Fit* used in GAGS calculations is given in Eq. 2:

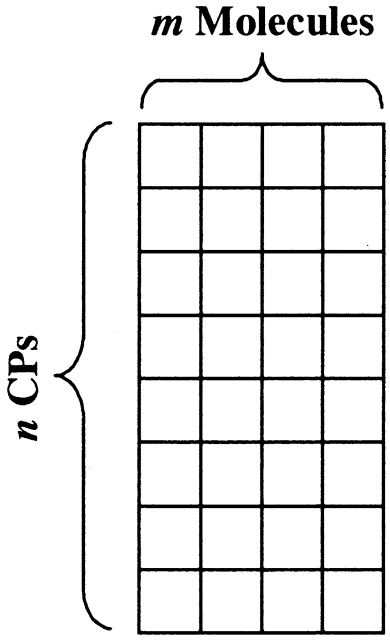

Fig. 3 Schematic representation of a $n \times m$ chromosome in GAGS, coding for the simultaneous superimposition of m molecules based on n critical points

$$Fit = W_R\left(\frac{T_R}{N_R}\right) + W_D\left(\frac{T_D}{N_D}\right) \quad (2)$$

with

$$T_R = \left[\frac{2}{m(m-1)}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m-1}\sum_{l=k+1}^{m}\left(\rho_i^k - \rho_i^l\right)^2\right]^{\frac{1}{2}}$$

$$N_R = \left[\frac{1}{m}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\left(\rho_i^k\right)^2\right]^{\frac{1}{2}}$$

$$T_D = \left[\frac{2}{m(m-1)}\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{m-1}\sum_{l=k+1}^{m}\left(d_{ij}^k - d_{ij}^l\right)^2\right]^{\frac{1}{2}}$$

$$N_D = \left[\frac{1}{m}\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{m}\left(d_{ij}^k\right)^2\right]^{\frac{1}{2}}$$

$$W_R + W_D = 1$$

where ρ_i^k corresponds to the ED at the ith CP of the kth molecule being aligned and d_{ij}^k corresponds to the inter-distance between the ith and jth CPs of the kth compound. The lower the value of *Fit* for a given chromosome the better its performance. After each run GAGS provides several multiple alignment solutions of the CP graphs, which are "decoded" one at a time. This is accomplished by first fitting the Cartesian coordinates of the CPs involved in the solutions, and second the molecules are aligned by applying the rotations and translations obtained during the fitting stage to the atomic coordinates. GAGS is thus able to provide a set of molecular overlays using only a small number of CPs to describe each molecule. Note also that the CPs clearly provide a reduced representation of the molecules based upon their medium-resolution electron density properties [36].

Multiple alignments were carried out employing a "leave-one-out" procedure using $n - 1$ of the n CPs determined as described in "Results and discussion". Although GAGS is designed to perform simultaneous alignments of multiple molecules in a single run, a two-step strategy was adopted because of the relatively weak binding affinity of Tzq (T) for PBR and its insignificant binding affinity for CBRs, as shown in Table 1. First, a ternary alignment of diazepam (D), Ro 5-4864 (R), and Quiz (Q), denoted by "D/R/Q" was carried out, where diazepam is taken to be the fixed reference compound. Second, a quaternary alignment "D/R/Q/T" was performed in which Tzq is treated as a perturbation of the ternary alignments of the more potent compounds. The sizes of the aligned molecular graphs and the number of nodes used in the alignments are summarized in Table 2, and the parameters used in the GAGS computer experiments are given in Table 3. The parameters have been set to their usual default values, except in the case of quaternary alignments, where the mutation and crossover rates have been modified to maximize the diversity of the solutions. An elitist strategy was also chosen for this reason [2]. Ten independent experiments, with different starting populations of chromosomes, were carried out in each case. The molecular alignments obtained with GAGS, expressed in terms of CPs, are then automatically converted into molecular alignments as described in the previous paragraph.

Field-based molecular similarity—MIMIC

Field-based similarity methods are based upon the assumption that it is the nature of a molecule's fields that determines its ability to bind to a given receptor; multiple molecules binding to the same receptor should

Table 2 Characteristics of the GAGS comparisons of molecular graphs based on the critical points obtained by topological analysis of the electron density at 2.5 Å, for diazepam (D), Ro 5-4864 (R), Quiz (Q), and Tzq (T)

Initial graph	Size of the graph
D	7
R	9
Q	8
T	7

Comparison	Sizes	$(n-1)$
D/R/Q	7/9/8	6
D/R/Q/T	7/9/8/7	6

Table 3 Parameters used for the GAGS comparisons of molecular graphs based on the critical points obtained by topological analysis of the electron density at 2.5 Å, for diazepam (D), Ro 5-4864 (R), Quiz (Q), and Tzq (T)

Comparison	Number of generations	Population size	Mutation Rate (%)	Crossover rate (%)
D/R/Q	1 000	500	0.001	0.60
D/R/Q/T	10 000	100, elitism	0.10	0.55

possess similar fields, at least in their binding regions. Thus, the degree of alignment of the molecular fields of multiple molecules provides a measure of their intermolecular similarities. The program MIMIC, used here to provide a "field-based perspective" for the reduced representation GAGS results, aligns both the molecular steric volume (MSV) and the molecular electrostatic potential (MEP) fields of molecules. Inclusion of the MEP field in addition to the MSV field ensures that the electronic aspects of molecular shapes, which may play a role in ligand–protein binding, are also accounted for. A 2:1 ratio of MSV and MEP field values used in this work provides a suitable balance of the two fields and ensures that the MEP field does not dominate. A detailed description of the method and several applications illustrating its use have already been published [6–12]. For consistency, ternary matches "D/R/Q" were considered first, with diazepam again being chosen as the fixed reference compound. The weaker binding Tzq molecule being added subsequently to give the quaternary alignment "D/R/Q/T", as was the case in the GAGS procedure described in the previous section.

Results and discussion

GAGS-based comparison of the BZs

In the ternary alignments, three major classes appear in the best solutions obtained with the GAGS procedure, as shown in Fig. 4. The best ternary solution, with a *Fit* value of 0.115, corresponds to the "standard orientation", named ST(3), of the aligned molecules with the benzo and phenyl rings as well as the carbonyl oxygens in near-perfect registration. The two other alignment classes correspond to "inverted orientations", I1(3) and I2(3), of the adapting molecules relative to the reference molecule. In the case of alignment solutions in the I1(3) class, the best *Fit* value of 0.188 corresponds to an

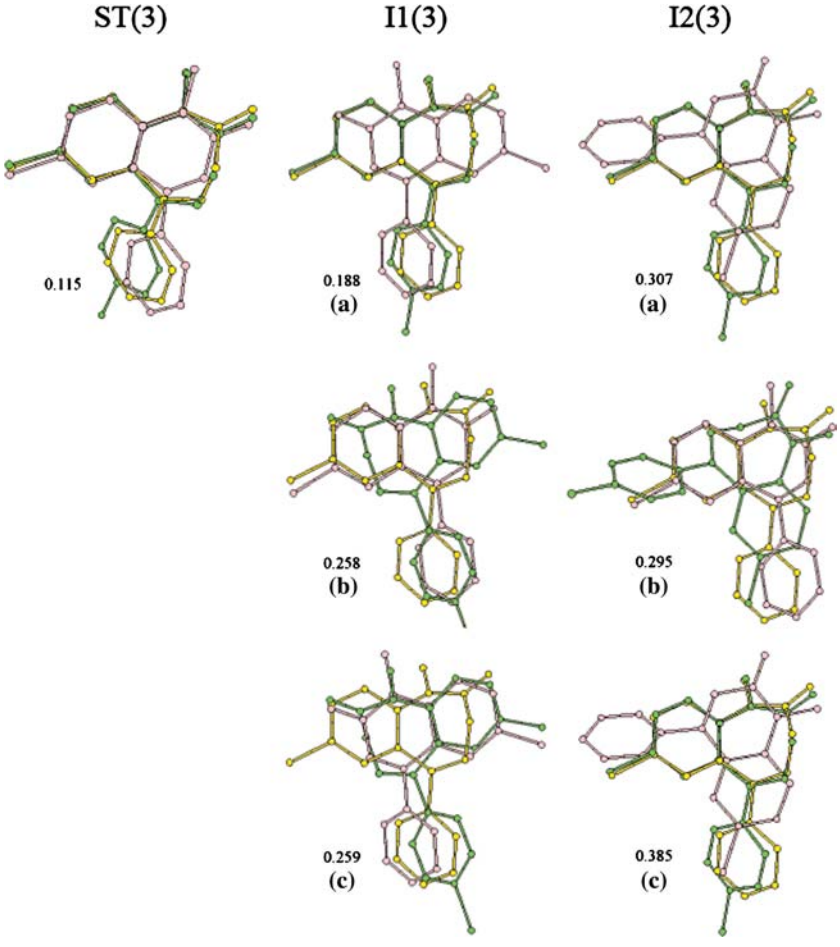

Fig. 4 Best ternary alignments, obtained with GAGS, of compounds D (*yellow*), R (*green*), and Q (*purple*). Solutions corresponding to standard (ST) and inverted (I1, I2) orientations are illustrated; numbers indicate the *Fit* value associated to each superimposition

approximate 180° rotation of Quiz (in purple) about the vertical axis. The remaining two solutions with *Fit* values of 0.258 and 0.259 correspond to similar rotations, respectively, for Ro 5-4864 (in green) and for the Quiz and Ro 5-4864 pair. Alignment solutions in the I2(3) class correspond to approximate 180° rotations with respect to a diagonal axis that lies approximately along the carbonyl bond of the adapting molecules. The three alignment solutions in this class have *Fit* values of 0.307, 0.295, and 0.385 corresponding to a rotation of Quiz (in purple), I2a(3), to a rotation of Ro 5-4864 (in green), I2b(3), and to a rotation of both molecules, I2c(3).

The best quaternary alignment solutions also fall into the same three classes as shown in Fig. 5. Again the best *Fit* value of 0.313 corresponds to the standard orientation, ST(4), with excellent alignment of the benzodiazepine and quinazolinone moieties and of the carbonyl oxygens of all four compounds but significant mismatching of the phenyl and triazolo rings of Tzq, suggesting that this may be responsible for its weak activity. In the I1 class of quaternary orientations, I1a(4) with a *Fit* value of 0.334 corresponds to I1a(3) for the ternary alignment, I1b(4) with a *Fit* value of 0.337 corresponds to an approximate 180° rotation of Tzq (in blue) about the vertical axis, and I1c(4) with a *Fit* value of 0.374 corresponds to I1b(3) for the ternary alignment, Tzq being again rotated about the vertical axis. We also observed an I2(4) class of orientations, corresponding to a 180° rotation of Tzq about the diagonal axis (*Fit* = 0.372).

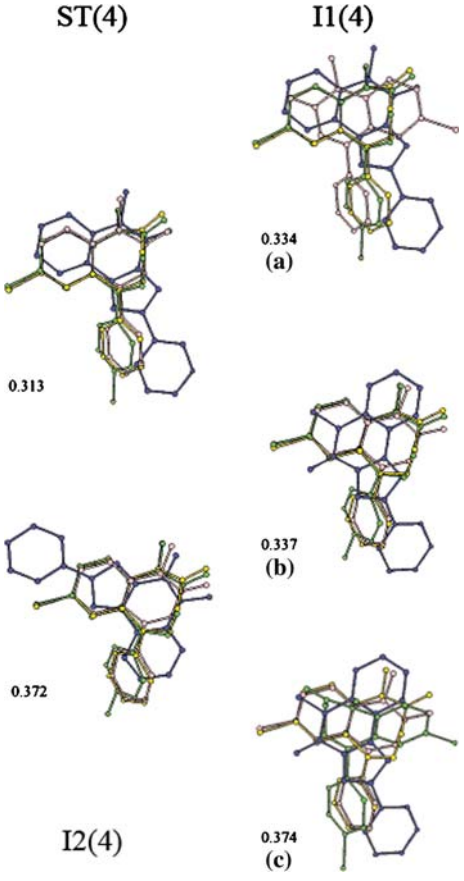

Fig. 5 Best quaternary alignments, obtained with GAGS, of compounds D (*yellow*), R (*green*), Q (*purple*), and T (*blue*). Solutions corresponding to standard (ST) and inverted (I1, I2) orientations are illustrated; numbers indicate the *Fit* value associated to each superimposition

MIMIC-based comparison of the BZs

In order to further evaluate the molecular superimpositions obtained with the GAGS procedure, molecular alignments were also carried out using the field-based method MIMIC described in the "Field-based molecular similarity" section of "Materials and methods". Figure 6 depicts the ternary molecular alignments computed by MIMIC. From the figure it is clear that the standard orientation, ST(3), which has the highest similarity, is in agreement with the best ternary alignment produced by the GAGS procedure. The second best MIMIC solution, I1(3), is approximately in agreement with the best GAGS solution but is rotated slightly in a counter-clockwise manner about an axis that is normal to the plane of the paper. The third best MIMIC solution, I2(3), is in agreement with the corresponding GAGS-based alignment. The situation differs, however, in the case of quaternary alignments as shown in Fig. 7. For example, in the standard orientation, ST(4), with a similarity value of 0.711, the quinazolinone ring in Tzq is displaced significantly to the left in comparison to its position in the corresponding GAGS-based ST(4) shown in Fig. 5. In the I1 class of quaternary orientations, the I1(4) solution presented in Fig. 7, with a similarity value of 0.671, corresponds to the I1b(4) GAGS-based alignment. In the I2 class of orientations, an I2(4) superimposition is obtained with a similarity value of 0.647, totally consistent with the GAGS-based I2(4) solution shown in Fig. 5. In addition, a third class of inverted orientation, I3(4), is observed with a similarity value of 0.674, corresponding to the 180° rotation of Tzq about a diagonal axis (i.e., I2 orientation) followed by a 180° rotation of the same molecule about an horizontal axis approximately located at the middle of the bicyclic moiety of the three other ligands.

Structure – activity relationships

During the last 20 years, several research groups have attempted to develop pharmacophore models that describe the geometric and electronic structural features of BZs such as diazepam, and the relationship of these features to their binding affinity for CBR. In the early 1980s, Crippen proposed the first model, which included atoms of the benzo and phenyl aromatic rings, as well as both nitrogen atoms of the diazepinone cycle [44, 45]. Later, Loew and co-workers [46–49], Codding *et al*. [50], Skolnick, Cook, and co-workers [51–53], as well as Bo-

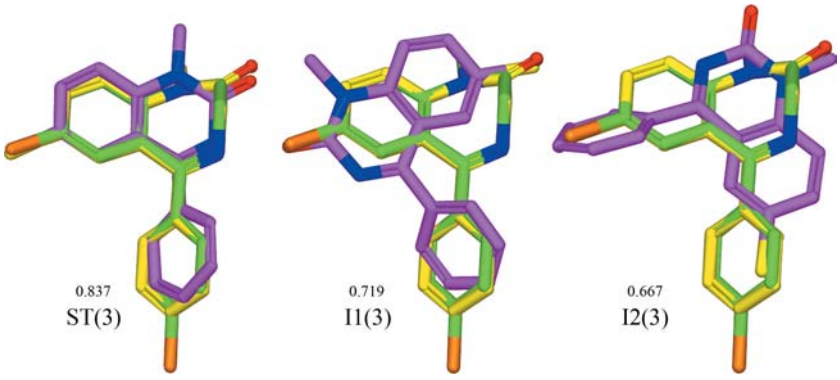

Fig. 6 Best ternary alignments, obtained with MIMIC, of compounds D (*yellow*), R (*green*), and Q (*purple*). Solutions corresponding to standard (ST) and inverted (I1, I2) orientations are illustrated; numbers indicate the similarity value associated to each superimposition

rea, Bertolasi, and co-workers [54, 55] showed, on the one hand, the importance of the π-aromatic system (benzo ring) and its chlorine substitution, and on the other hand, the existence of two proton acceptor sites: the carbonyl dipole and the imine nitrogen. These conclusions were summarized by the four-point pharmacophore model of Tebib *et al*. [56] as illustrated in Fig. 8: two electronegative regions δ1 and δ2 containing the carbonyl and imine dipoles or equivalent functions, surrounded by two hydrophobic regions containing the aromatic cycles denoted as PAR (''π-aromatic region'') and FRA (''freely rotating aromatic''). Additionally, according to Bourguignon *et al*. [33] and Didier [57], the selectivity of the benzodiazepine-type ligands for either CBR or PBR would be based on very local structural-modulation effects. Therefore, the pharmacophore model describing their binding to PBR is expected to be nearly the same as that for the CBR model.

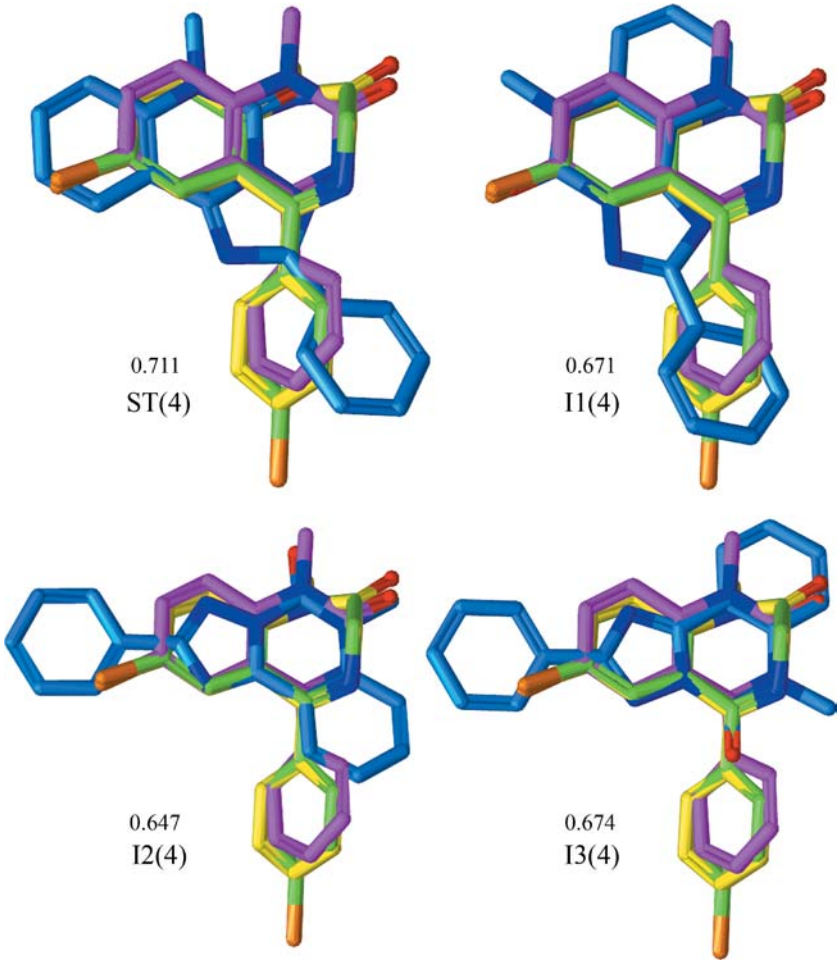

Fig. 7 Best quaternary alignments, obtained with MIMIC, of compounds D (*yellow*), R (*green*), Q (*purple*), and T (*blue*). Solutions corresponding to standard (ST) and inverted (I1, I2, I3) orientations are illustrated; numbers indicate the similarity value associated to each superimposition

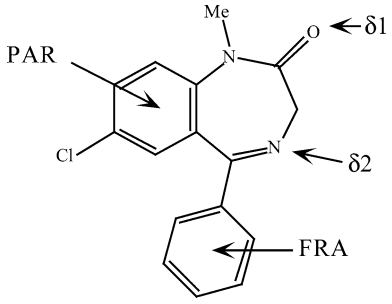

Fig. 8 Pharmacophore model proposed by Tebib *et al.* [56]

From the data in Table 1 and the structures depicted in Fig. 1 it appears that the *para*-chloro substituent on the phenyl ring of Ro 5-4864 plays a major role in determining the differential binding affinities of diazepam, Ro 5-4864, and Quiz towards CBR and PBR. This is supported by the ST(3) molecular alignments produced by both the GAGS and MIMIC procedures as shown in Figs. 4 and 6. In both procedures the *para*-chloro-benzo and N-methyl amide moieties are well aligned and the ''envelopes'' containing the phenyl groups of the three molecules are also in reasonable agreement suggesting some structural latitude in this region of the corresponding receptors. The imine nitrogens are less well aligned, which may account for some of the smaller differences in binding to the two receptors. However, the outstanding structural feature is the *para*-chloro substituent on the phenyl ring of Ro 5-4864 that extends outside of the phenyl ring envelope of all three molecules. Two pieces of evidence support this conclusion (cf. Table 1): (1) Ro 5-4864 binds approximately 1000× weaker than diazepam to the CBR, and (2) Ro 5-4864 binds very strongly to PBR and approximately 1000× less to CBR. These results suggest that, everything else being equal, the PBR binding site has additional space to accommodate the *para*-chloro substituent not present in the corresponding CBR binding site. Noting that diazepam binds approximately 10× weaker than Ro 5-4864 to the PBR, and noting the extremely similar structures of the two molecules suggests that the *para*-chloro substituent contributes about an order of magnitude to the affinity constant for PBR binding. This also suggests that the *para*-chloro derivative of Quiz might exhibit increased binding affinity towards the PBR with decreased affinity for the CBR. Unfortunately, although synthesis of the compound was carried out, it was found to be too insoluble for accurate measurement of its binding affinity to the two receptors [58].

None of the ''inverted'' alignments, I1 and I2, depicted in Figs. 4 and 6 provide a consistent explanation of the data. For example, in superimposition I1a(3) of Fig. 4, the Quiz molecule is rotated about the vertical axis of the figure, which moves the chloro-substituent to the opposite side of the figure in the region of the polar carbonyl and imine nitrogen groups. A similar argument can be made for the other two I1 alignments. The other ''inverted'' form, I2, also involves a reflection or approximate 180° rotation about an axis that lies approximately along the carbonyl bond (cf. Figs. 4 and 6). This results in an unfavorable placement of the phenyl and benzo rings that is difficult to rationalize with the experimental binding data for the two receptors. However, Anzini *et al.* [59] have presented a docking study of various PBR ligands including a theoretical model of the receptor, suggesting the possible existence of multiple binding modes among the inverted orientations of the molecules. Such inverted classes of orientations are thus not completely excluded.

Figures 5 and 7 depict the GAGS- and MIMIC-based quaternary alignments for diazepam, Ro 5-4864, Quiz, and Tzq. As described in the ''Gaph similarity searching'' and ''Field-based molecular similarity'' sections of ''Materials and methods'', because of its weak binding to both CBR and PBR, the alignment of Tzq is treated as a perturbation to the ternary ST(3) alignments of Diazepam, Ro 5-4864, and Quiz. The quaternary standard ST(4) alignments derived from both GAGS and MIMIC provide a reasonable structural explanation of the binding data. The quinazolinone ring of Tzq matches nicely with the corresponding rings of diazepam, Ro 5-4864, and Quiz such that the N-methyl and carbonyl groups are in reasonable alignment. The basic nitrogen of the triazole ring lies in the same region as the imine nitrogen of the other three molecules but is somewhat displaced, a feature that would most likely reduced binding affinity. The most significant structural deviation of Tzq is, however, the important displacement of the phenyl ring outside of the envelope of the phenyl rings of the other three molecules. Since, as noted above, there appears to be more room in this region of the PBR binding site it is not surprising that the binding affinity for CBR is insignificant while that for PBR lies in a range which is comparable to that of Quiz but considerably less than those of diazepam and Ro 5-4864. A significant difference exists between the ST(4) alignments obtained from the GAGS and MIMIC methods. In the latter, the alignment of the Tzq molecule is shifted to the left so that the quinazolinone ring now lies in the region of the chloro-substituent of the benzo ring of the other three molecules. While this alignment cannot be ruled absolutely out based on the available data, it does not seem to be a likely position for Tzq relative to the other three molecules. Thus, it appears, at least in this case, that the GAGS-based alignment makes more sense than that produced by MIMIC. Lastly, none of the GAGS- or MIMIC-based ''inverted'' alignments appears to provide a better rationalization of the existing data. However, the I1, I2, and I3 types of relative orientations, supported by the work of Anzini *et al.* [59] involving bigger ligands, strongly suggests that Tzq might bind to the PBR in an alternative orientation.

All of this suggests, as noted by Bourguignon and coworkers [31, 57, 59] that relatively small structural differences in between CBR and PBR are responsible for the differential binding affinities of diazepam, Ro 5-4864, Quiz, and Tzq. In this regard, it should be noted

that in assessing the differences in binding affinities due to the structural differences among diazepam, Ro 5-4864, and Quiz, small differences in the free energy of binding of only about 1.3 kcal mol^{-1} yield a tenfold change in affinity constant. Thus, even something as small as a ''misaligned'' or ''non-ideal'' hydrogen bond can be responsible for an order of magnitude change in binding.

Summary and conclusions

In this contribution, the stereo electronic features of four benzodiazepine-type ligands, Diazepam, Ro 5-4864, Quiz, and Tzq, were analyzed in terms of their binding affinities (IC_{50}'s) to two benzodiazepine receptors, CBR and PBR, using two similarity-based molecular alignment methods, namely GAGS and MIMIC. In the former, reduced representations are generated by constructing 3D molecular graphs linking the critical points of the electron density computed at medium crystallographic resolution (2.5 Å). The molecular graphs are then matched using GAGS, a novel GA-based algorithm. In the latter, the steric and electrostatic fields of the molecules are matched. Both methods can align multiple molecules simultaneously, leading to solutions in terms of molecular superimpositions with high performance (GAGS) or similarity (MIMIC). Because Tzq binds weakly to both receptors, ternary overlays of Diazepam, Ro 5-4864, and Quiz were first carried out, followed by quaternary overlays of Tzq with the diazepam, Ro 5-4864, and Quiz molecules previously used to determine ternary alignments.

Comparison of the results produced by the two methods reveals considerable similarities. Three major classes of alignments possessing high fitness/similarity values are observed with both methods. The best alignment corresponds to a standard orientation (ST) that is consistent with available experimental data. Two ''inverted'' alignments that are less probable due to inconsistencies with experimental data are also observed. By relating the various ligand alignments to their binding affinities for both CBR and PBR a structure–activity relationship is developed that is consistent with the available binding data. It confirms that three structural features, a hydrogen bond-acceptor site and two hydrophobic regions, are required for significant binding to both types of receptors. Additional space in the phenyl-group binding region of the PBR is suggested by the enhanced binding affinity of Ro 5-4864 due to the *para*-chloro substituent on the phenyl moiety.

Other studies involving compounds with a higher level of structural diversity, for example including HIV reverse transcriptase inhibitors, have also been carried out with GAGS [60] and led to molecular alignments consistent with the published MIMIC results [11]. Future work will include the analysis of an expanded set of molecules and inclusion of conformational flexibility using the GAGS method [61]. This will facilitate the study of more flexible molecules such as peptides and expand the potential scope of the studies.

Acknowledgments The authors would like to thank the FUNDP and the Scientific Computing Facility (SCF) centre for the computing resources. The authors would also like to acknowledge Dr. D.C. Rohrer (Pfizer Global Research, St. Louis, MO, USA) and Prof. J.-J. Bourguignon (Louis Pasteur University, Illkirch, France) for numerous fruitful discussions. N.M. thanks the ''Fonds National de la Recherche Scientifique'' for her Scientific Research Worker position.

References

1. Meurice N, Leherte L, Vercauteren DP, Bourguignon J-J, Wermuth CG (1997) Development of a genetic algorithm method especially designed for the comparison of molecular models: application to the elucidation of the benzodiazepine receptor pharmacophore. In: van de Waterbeemd H, Testa B, Folkers G (eds) Computer-assisted lead finding and optimization. Verlag Helvetica Chimica Acta, Basel, pp 497–510
2. Meurice N, Leherte L, Vercauteren DP (1998) SAR QSAR Environ Res 8:195–232
3. Leherte L, Meurice N, Vercauteren DP (2000) J Chem Inf Comput Sci 40:816–832
4. Meurice N, Maggiora GM, Vercauteren DP (2000) Study of the molecular similarity among three HIV reverse transcriptase inhibitors in order to validate GAGS, a genetic algorithm for graph similarity search. In: Gundertofte C, Jorgensen FS (eds) Molecular modeling and prediction of bioactivity. Kluwer, New York, pp 427–428
5. Leherte L, Meurice N, Vercauteren DP (2000) Applications of multiresolution analyses to electron density maps: a critical point analysis approach for the comparison of molecules. In: Mastorakis NE (ed) Mathematics and computers in modern science. World Scientific and Engineering Society, Danvers, MA, pp 158–164
6. Rohrer DC (1995) 3D molecular similarity modeling in computational drug design. In: Carbó R (ed) Molecular similarity and reactivity: from quantum to phenomenological approaches. Kluwer, Amsterdam, pp 141–161
7. Mestres J, Rohrer DC, Maggiora GM (1997) J Comput Chem 18:934–954
8. Mestres J, Rohrer DC, Maggiora GM (1997) J Mol Graph Model 15:114–121
9. Rohrer DC (1998) 3D molecular similarity methods: in search of a pharmacophore. In: Codding PW (ed) Structure-based drug design. Kluwer, Dordrecht, pp 65–76
10. Blinn JR, Rohrer DC, Maggiora GM (1999) Field-based similarity forcing in energy minimization and molecular matching. In: Altman RB, Dunker AK, Hunter L, Klein TE (eds) Proceedings of the Pacific Symposium on Biocomputing '99. World Scientific, Singapore, pp 415–425
11. Mestres J, Rohrer DC, Maggiora GM (1999) J Comput Aided Mol Des 13:79–93
12. Mestres J, Rohrer DC, Maggiora GM (2000) J Comput Aided Mol Des 14:39–51
13. Möhler H, Okada T (1977) Science 198:849–851
14. Braestrup C, Squires RF (1977) Proc Natl Acad Sci USA 74:3805–3809
15. Squires RF, Braestrup C (1977) Nature 266:732–734
16. Tallman JF, Thomas JW, Gallager DW (1978) Nature 274:383–385
17. Costa E, Guidotti A, Toffano G (1978) Br J Psychiatry 133:239–248
18. Haefely WE (1987) Chimia 41:389–396
19. Sieghart W (1992) Trends Pharmacol Sci 13:446–450
20. Jordan AD, Kordik CP, Reitz AB, Sanfilippo PJ (1996) Expert Opin Ther Patients 6:1047–1060

21. Schoemaker H, Boles RG, Horst WD, Yamamura HI (1983) J Pharmacol Exp Ther 225:61–69
22. Wang JKT, Taniguchi T, Spector S (1984) Mol Pharmacol 25:349–351
23. Anholt RRH, Pedersen PL, De Souza EB, Snyder SH (1986) J Biol Chem 261:576–583
24. Gavish M, Bachman I, Shoukrun R, Katz Y, Veenman L, Weisinger G, Weizman A (1999) Pharmacol Rev 51:629–650
25. Bernassau JM, Reversat JL, Ferrara P, Caput D, Lefur G (1993) J Mol Graph 11:236–244
26. Casellas P, Galiegue S, Basile AS (2002) Neurochem Int 40:475–486
27. Galiegue S, Tinel R, Casellas P (2003) Curr Med Chem 10:1563–1572
28. Anholt RRH (1986) Trends Pharmacol Sci 7:506–511
29. Farges R, Joseph-Liauzun E, Shire D, Caput D, Le Fur G, Loison G, Ferrara P (1993) FEBS Lett 335:305–308
30. Meurice N (2001) Part of Ph.D. thesis work. Développement d'un algorithme génétique original pour l'alignement de graphes moléculaires issus de la densité électronique—Application à l'étude comparative de composes d'intérêt pharmacologique. Laboratoire de Physico-Chimie Informatique, University of Namur, Belgium
31. Bourguignon J-J (1993) Endogenous and synthetic ligands of mitochondrial benzodiazepine receptors: structure–affinity relationships. In: Giesen-Crouse E (ed) Peripheral benzodiazepine receptors. Academic, London, pp 59–85
32. Allen FH (2002) Acta Crystallogr B 58:380–388
33. Camerman A, Camerman N (1972) J Am Chem Soc 94:268–272
34. Hempel A, Camerman N, Camerman A (1987) Can J Chem 65:1608–1612
35. Meurice N, Norberg B, Durant F, Didier B, Bourguignon J-J, Vercauteren DP (2005) Acta Cryst C (in preparation)
36. Binamé J, Meurice N, Leherte L, Glasgow J, Fortier S, Vercauteren DP (2004) J Chem Inf Comput Sci 44:1394–1401
37. Hall SR, du Boulay DJ, Olthof-Hazekamp R (2000) Xtal3.7 System. University of Western Australia, Crawley
38. Johnson CK (1977) ORCRIT—the Oak Ridge Critical point network program. Chemistry Division Oak Ridge National Laboratory, Oak Ridge TN
39. Johnson CK (1993) ORCRIT—the Oak Ridge Critical point network program. User's guide, Oak Ridge, TN
40. Leherte L, Fortier S, Glasgow J, Allen FH (1994) Acta Crystallogr D 50:155–166
41. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. University of Alabama (ed) Addison-Wesley, Reading MA
42. Cartwright HM (1993) Applications of artificial intelligence in chemistry. In: Compton RG, Davies SG, Evans J (eds) Oxford University Press, New York
43. Lucasius CB, Kateman G (1993) Chemom Intell Lab Syst 19:1–33
44. Crippen GM (1979) J Med Chem 22:988–997
45. Crippen GM (1982) Mol Pharmacol 22:11–19
46. Loew GH, Nienow JR, Poulsen M (1984) Mol Pharmacol 26:19–34
47. Loew GH, Nienow JR, Lawson JA, Toll L, Uyeno ET (1985) Mol Pharmacol 28:17–31
48. Villar HO, Uyeno ET, Toll L, Polgar W, Davies MF, Loew GH (1989) Mol Pharmacol 36:589–600
49. Villar HO, Davies MF, Loew GH, Maguire PA (1991) Life Sci 48:593–602
50. Codding PW, Muir AKS (1985) Mol Pharmacol 28:178–184
51. Allen MS, Hagen TJ, Trudell ML, Codding PW, Skolnick P, Cook JM (1988) J Med Chem 31:1854–1861
52. Hollinshead SP, Trudell ML, Skolnick P, Cook JM (1990) J Med Chem 33:1062–1069
53. Wong G, Koehler KF, Skolnick P, Gu Z-Q, Ananthan S, Schönholzer P, Hunkeler W, Zhang W, Cook JM (1993) J Med Chem 36:1820–1830
54. Borea PA, Gilli G, Bertolasi V, Ferretti V (1987) Mol Pharmacol 31:334–344
55. Bertolasi V, Ferretti V, Gilli G, Borea PA (1990) J Chem Soc Perkin Trans 2:283–289
56. Tebib S, Bourguignon J-J, Wermuth CG (1987) J Comput Aided Mol Des 1:153–170
57. Didier B (1998) Ph.D. thesis. Laboratoire de Pharmacochimie de la Communication Cellulaire, Université Louis Pasteur, Illkirch
58. Didier B, Bourguignon J-J, Personal communication
59. Anzini M, Cappelli A, Vomero S, Seeber M, Menziani MC, Langer T, Hagen B, Manzoni C, Bourguignon J-J (2001) J Med Chem 44:1134–1150
60. Meurice N, Maggiora GM, Vercauteren DP (2000) Study of the molecular similarity among three HIV reverse transcriptase inhibitors in order to validate GAGS, a genetic algorithm for graph similarity search. In: Gundertofte C, Jorgensen FS (eds) Molecular modeling and prediction of bioactivity. Kluwer, New York, pp 427–428
61. Meurice N, Vercauteren DP (2005) J Chem Inf Model (in preparation)