# Shape group studies of molecular similarity: relative shapes of Van der Waals and electrostatic potential surfaces of nicotinic agonists

**Gustavo A. Arteca, Victoria B. Jammal, and Paul G. Mezey**
Department of Chemistry, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

**Janardan S. Yadav, Mark A. Hermsmeier, and Tamara M. Gund**
Division of Chemistry, Department of Chemical Engineering, Chemistry, and Environmental Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102, USA

*In drug design, the usual strategy involves characterizing and comparing the shapes of molecules. We apply a simple method to accomplish this goal: determining the symmetry-independent shape groups (homology groups of algebraic topology) of a molecular surface.*

*In this paper, we have adapted the method to describing the interrelation between Van der Waals and electrostatic potential surfaces. We describe rigorously the shape features in a series of molecules by using specific ranges of electrostatic potential over a Van der Waals surface. We consider a series of four nicotinic agonists as an example and discuss their expected activities as potential drugs on the basis of the shape similarities found.*

*Keywords: molecular similarity, Van der Waals and electrostatic potential surfaces, shape groups, nicotinic agonists*

Received 23 September 1987
Accepted 3 November 1987

## INTRODUCTION

Computer-aided drug design is a topic of major interdisciplinary interest. Its growing importance in the pharmaceutical industry has stimulated research in medicinal as well as theoretical chemistry, biochemistry, theoretical biology, and applied mathematics, including computer science (for reviews, see References 1–3).

The task of designing drugs involves a wide variety of problems of experimental as well as theoretical nature. From the computational viewpoint, the methods of quantum chemistry allow us to study, in an efficient way, the basic properties of a large set of molecules. Recent progress in computational chemistry and in our understanding of the significance of steric and electronic factors has enabled us to modify the practice of expensive, and often somewhat random, testing of potential drugs. Accordingly, we can reduce the actual experimental screening to a smaller family of compounds, which are predicted to be the most promising ones, based on a prescreening by the computer.

In this context, new developments in computer hardware and software play a key role. First, faster computers (with larger memories) reduce the computation time required and make possible the analysis of larger molecules. Second, the development of computer graphics facilities helps to display in a clear and more easily understandable way one important result of the theoretical approach: a collection of some physically meaningful functions that are chosen to describe the "molecular shape."

Visual inspection, however powerful, is not always rigorous, and it can be cumbersome if applied, for example, to a family of several hundred molecules. We need nonvisual, computational methods for describing and analyzing the shapes of molecules. These techniques will allow the computer to compare a large number of molecules without human interaction. This paper concerns the use of an efficient *nonvisual* method of shape characterization and shape comparison of molecules.

In order to appreciate the relevance of certain shape studies of drugs to the general problem of drug design, we recall the strategy usually followed:

(1) Several potential drugs are tested experimentally under the hypothesis that they produce similar biochemical effects (though usually not with the same potency) due to their assumed reaction with the same receptor.
(2) The physical, electronic and structural features of the effective drugs found in Step 1 are analyzed. Some physically meaningful properties are used to compare the molecules and to determine what struc-

tural features they have in common. Usually, this structural feature (the whole molecule or a part of it, the *pharmacophore*) is believed to be important in interactions at the receptor site and is believed to be responsible in part for the observed medicinal action.

(3) According to the information obtained in Step 2, new drugs can be designed that show similarities with the above molecules.

In the case of interactions involving *biomolecules* (for example, the drug and its protein receptor), we know that they are in part due to electrostatic and steric effects. This fact renders the problem of studying the structure of drug molecules well suited to employ molecular electrostatic potential (MEP) maps for their characterization. The MEP has long been recognized as an important tool of characterization,[4-6] and it is widely studied (see, for instance, References 7–12). Consequently, the analysis of the shape of MEP maps (for isolated molecules) becomes an important tool to describe a molecule as seen by the receptor. It is also well known, however, that the evaluation of the MEP, based on a model of point charges or fractional charges on atoms, fails to describe the *interaction* between large molecules. Nevertheless, the MEP is an accurate property of the isolated molecule. As long as the medicinal action can be correlated with structural properties of the isolated molecules, the analysis of MEPs is a powerful tool.

Although the MEPs can provide a description for the "shape of the molecule," they are not ideally suited to describe that part of the interaction between the drug and its receptor that is due to size or volume effects. By contrast, a Van der Waals surface (VDWS) can describe the latter features in a rather straightforward way. It seems reasonable that a combined analysis of electrostatic potential and VDWSs would provide a more complete description. In fact, mapping the MEP over the VDWS is a simple, pictorial alternative. The usefulness of this combined representation was first realized when comparing the Lewis basicity in a family of molecules.[13] Such a description is also important in many other problems, including studies on chemical reactivity, polarizability effects, and general solvent-solute interactions, among others.[14-16]

We will focus our attention on the two properties mentioned above. However, we should add that other methods could also be useful to study the relation between molecular shape and chemical reactivity. These methods include the study of isodensity contours[17,18] and the recent application of fractal dimension to analyze molecular surfaces.[19]

Steps 2 and 3 in the strategy for drug design involve a common and fundamental problem: shape comparison. In our case, this boils down to the following question: Given two MEPs or their mappings on VDWSs, what criteria should we use to decide whether asymmetric objects have different or similar shapes? Even though a visual inspection on a graphics computer terminal might be revealing, it is somewhat short of a mathematically rigorous method of shape description. Furthermore, it is not easy to express the results of a visual "shape" characterization in a concise manner, suitable for further algebraic analysis by a computer. These problems are well recognized, and useful techniques to accomplish such an analysis already have been proposed.[20-23]

Recently, a simple method, based on general results from algebraic topology,[24] was developed that can approach the above problem from a different perspective.[25,26] The method is based on the determination of the homology (and cohomology) groups of a simplicial cell complex derived from the molecular surface. It can be easily applied to characterize the shape of isopotential or isodensity contours associated with a given molecular configuration,[25,26] as well as with their change during a conformational rearrangement in a configurational space.[27] Furthermore, it can be generalized to *intersecting* surfaces.[28] This particular possibility, adapted appropriately, is the technique we apply in this paper to characterize a family of biochemically active molecules: a series of four molecules that are nicotinic agonists. These molecules represent a nontrivial example of drugs that are supposed to play a significant role with relation to the nicotinic receptor. This receptor is important because of its action in controlling essential processes in the nervous system (see, for instance, references quoted in Ref. 1). Many diseases, including Alzheimer's[30] and Parkinson's[31] diseases, are believed to be linked to disorders involving the corresponding membrane-bound protein receptor. Accordingly, the compounds chosen as examples for demonstrating the use of the new shape analysis are of medicinal significance.

In Section 2, we will describe the molecules of interest and their importance. In addition, we will discuss the techniques used to evaluate the MEP and its map over the VDWS. We have included a pictorial representation for the shape of a drug molecule as an example of the general technique. In Section 3 we will apply the shape group method for the description of the shape of these molecules (in particular, their pharmacophores) by computing the set of Betti numbers characterizing their homology groups. Finally, in Section 4 we will comment on the conclusions derived from the shape analysis in relation to the expected activity of these potential drugs.

## METHOD FOR DETERMINING THE MOLECULAR ELECTROSTATIC POTENTIAL OVER A VAN DER WAALS SURFACE

The series of compounds we have chosen are the isoarecolone (I), isoarecolol (II), and the corresponding derivatives obtained by saturating the double bond on the ring (compounds III and IV, respectively).[32,33] The schematic structural formulas of these compounds are shown in Figure 1.

These molecules are of interest because they act as nicotinic agonists. That is, the potential drugs are believed to have a selective effect at the level of ganglia and striated muscle junction. This action mimics that of nicotine or acetylcholine on the corresponding receptor. As it is well known, the nicotinic activity is relevant to many essential processes involving the central nervous system (for an introductory review, see Reference 1 and others quoted therein). Accordingly, it is important to study the structure of these potential drugs and to obtain a detailed comparative description of their shapes. The standard computation of the MEP,

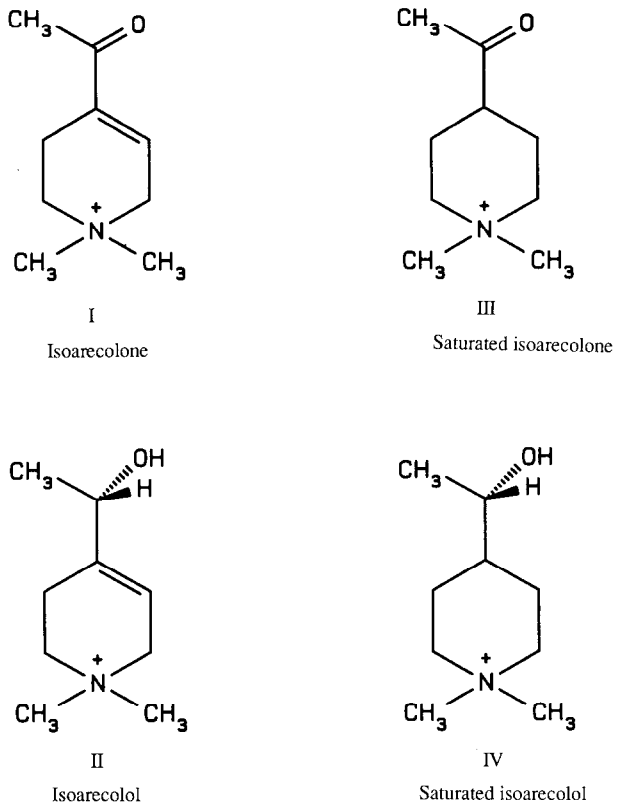


*Figure 1. Structural formulas for isoarecolone, isoarecolol, and their corresponding saturated derivatives*

denoted by $V(\mathbf{r})$, requires in principle the knowledge of the electronic charge density function $\rho(\mathbf{r})$:

$$V(\mathbf{r}) = \sum_A Z_A/\|\mathbf{r} - \mathbf{R}_A\| - \int_v \{\rho(\mathbf{r}')/\|\mathbf{r} - \mathbf{r}'\|\}d\mathbf{r}' \qquad (1)$$

where $V(\mathbf{r})$ represents the electrostatic potential produced by the molecule at point $\mathbf{r}$ and where $Z_A$, and $\mathbf{R}_A$ represent the nuclear charge and position vector of nucleus A, respectively. (Atomic units are used throughout.)

The integral in Equation 1 is evaluated over all space. The potential and the density are related by the Poisson equation: $\Delta V(\mathbf{r}) = 4\pi\rho(\mathbf{r})$ (see Reference 29 for a description of the actual computation).

In practice, you can use the electronic wavefunction $\psi(\mathbf{r})$ at *ab initio* or semiempirical levels to obtain the density. With a larger system, you can resort to computing the MEP by means of the fractional charges on the atoms and a multipolar expansion (see, for instance, References 14 and 15). The latter can be determined in the context of an MO population analysis. Mulliken charges are known to provide a deficient description of the MEP.[15] Other fractional charges, such as those in Kollman's potential, are often advantageous.[16]

We computed both *ab initio* and semiempirical wavefunctions for all the molecules in order to compare their electronic properties. We also analyzed the basis set dependence of the results. The procedure can be considered now as standard and has been described in detail in the literature.[7,34,35]

To compute the MEP, we considered a grid of points on the VDWS for each molecule. After computing the MEP value for each point in the grid, we determined a sequence of characteristic ranges of the MEP on the VDWS values.

Figures 3–8 display a typical example of these surface domains, considering the case of molecule I (isoarecolone). Each figure shows a different view of molecule I along each direction of the coordinate axes (Figure 2 shows the corresponding structural, stereographic model). The so-called front view shows the oxygen of the carbonyl group at the top and pointing to the viewer. One of the hydrogen atoms, corresponding to one of the methyl groups bonded to the nitrogen, is placed at the bottom of the figure, whereas two hydrogen atoms of the other methyl groups are pointing to the front. The remaining five views for this molecule (right, left, bottom, top and back) are derived by the appropriate rotation.

In these figures you can see the atomic spheres from which the VDWS is made up, as well as the ranges of potential, indicated with a different pattern, according to the convention explained in Figure 3.

The six views in Figure 2 give a detailed visual description of the VDWS associated with the whole molecule, as well as its interrelation with the MEP. We are interested in describing the shape of MEP along the molecular VDWS by using the information provided by the distribution of ranges of values of MEP on it. These figures depict the intersection of the *level sets* of two molecular functions: the *fixed* VDWS and the *level sets* of MEP, defined as the collection of points where the MEP is equal to a constant value. Our task is to characterize the shape of the surface domains obtained by the *intersection* of the VDWS and the MEP level sets. Our main goal is to provide a rigorous criterion to establish whether the patterns of these surface domains, for a sequence of different molecules, do show similarities or not. In particular, we are interested in describing how the shape of the pharmacophore is affected by the various substitutions and structural alterations carried out in the series of molecules. In the next section, we discuss the method proposed to accomplish the above goal.

## COMPUTING THE SHAPE GROUPS

In our case, the relevant shape groups of a molecule are the symmetry-independent homology groups of a simplicial cell complex obtained from the molecular surface.[24] The method to compute these groups has been described elsewhere;[25-27] here we shall confine the discussion to those aspects of the method that are relevant to our present specific problem. In fact, for an intuitive understanding of the shape group technique, there is no need for detailed mathematical background, and the discussion provided in this paper should be sufficient.

The central question we must clarify in order to apply the method is the nature of the three-dimensional (3D) object whose shape we are interested in determining. In our present case, the objects are those determined by the interpenetration of the isopotential surfaces with the VDWS. The generalization of the method to interpenetrating surfaces has been presented elsewhere.[28] The molecules we consider in this paper are of importance as prototypes for potential drugs, and they also provide an excellent example of the utility of the shape group method.

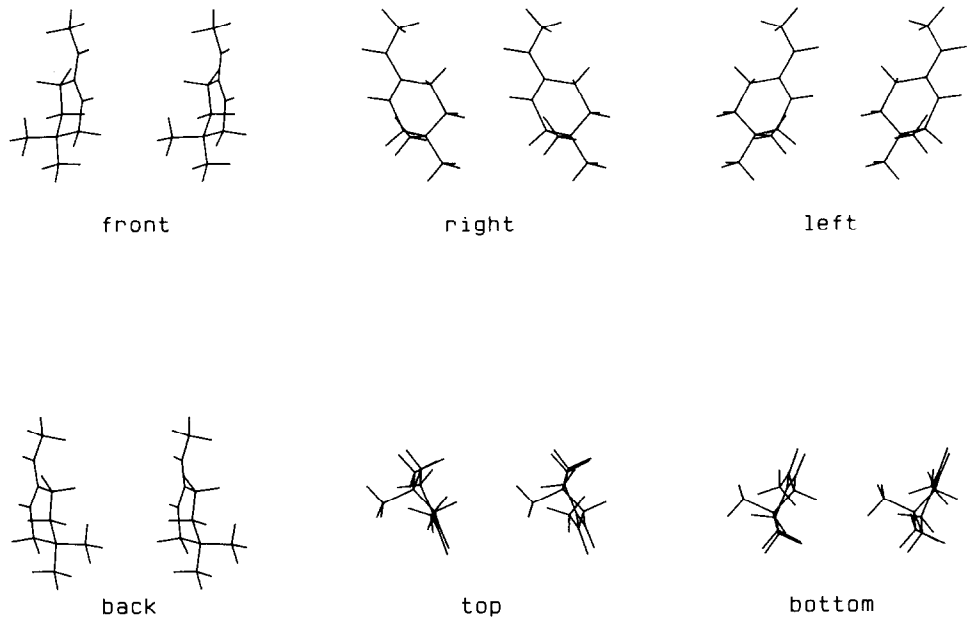


*Figure 2. Stereographic views of the bond skeleton of the isoarecolone molecule from different orientations*

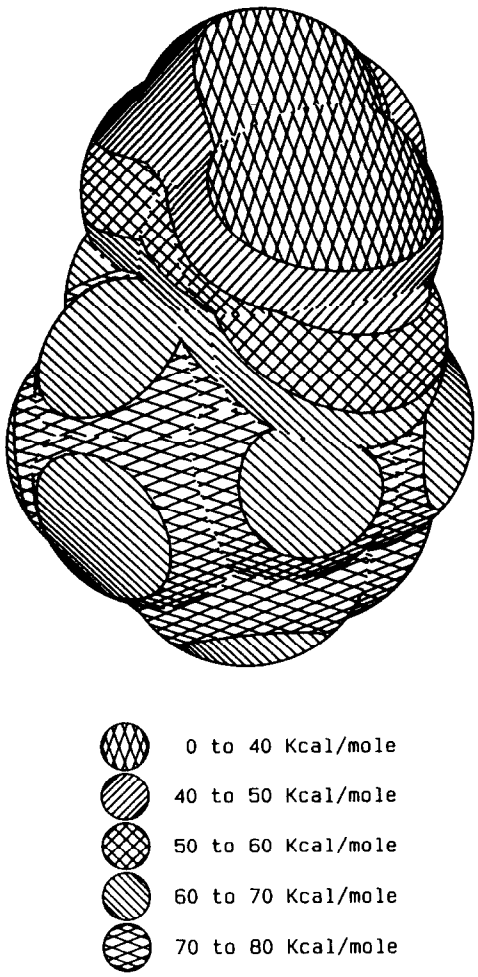


*Figure 3. Front view of the Van der Waals surface of isoarecolone (the different patterns on the surface specify the ranges of electrostatic potential)*

In order to construct the surface on which the analysis is performed, let us consider the two physical properties of interest, $f_1(\mathbf{r})$ and $f_2(\mathbf{r})$, regarded as functions of position vector $\mathbf{r}$ in the three-dimensional space ${}^3\mathbb{R}$, $\mathbf{r} \in {}^3\mathbb{R}$. In our case, $f_1$ is the MEP and $f_2$ is a function that has the value 1 on the VDWS, and it is zero elsewhere. Of course, the method can be applied by a straightforward extension to the interrelations between any other pair of 3D functions relevant to molecular properties (total charge density, HOMO or LUMO contours, etc).[28]

In what follows we will take $G_1(a_1)$ as the 3D closed contour surface corresponding to the value $f_1(\mathbf{r}) = a_1$ of the electrostatic potential. This surface can be changed continuously upon variation of the parameter $a_1$. In other words, there exist infinitely many contours $G_1(a_1)$, one for each value of $a_1$. On the other hand, there exists a unique VDWS, which will be denoted by $G_2(1)$.[28] (See the Appendix for a mathematically more rigorous discussion of these contours and their relationships.)

Suppose we have selected a set of increasing MEP values, $a = a_1^{(1)}, \ldots, a_1^{(i)}, \ldots$, and that we can identify all the points $\mathbf{r}$ *on the* VDWS $G_2(1)$, for which the MEP $f_1(\mathbf{r})$ satisfies the condition $a_1^{(i)} \leqslant f_1(\mathbf{r}) < a_1^{(i+1)}$, with $a_1^{(i)}$ and $a_1^{(i+1)}$ two accessible values of the electrostatic potential. This region on $G_2(1)$ can be viewed as the one formed by all the points of the VDWS that are enclosed *between* the two closed contours $G_1(a_1^{(i)})$ and $G_1(a_1^{(i+1)})$. In this sense, we can define "regions" or "domains" of electrostatic potential on the VDWS by determining how the latter is interpenetrated by the two MEP contours (see the Appendix for the general case). In particular, we will denote as $d_i$ the set formed by all the points on the VDWS for which the corresponding MEPs take values in the range $d_i = [a_1^{(i)}, a_1^{(i+1)})$. Note that here we consider closed-open parameter intervals. All these domains or regions on the VDWS appear as "caps" or "belts" or, possibly, as objects with more "holes," as shown in Figures 3–8.

As discussed in previous papers,[25-28] you can

characterize the shape of a 3D object once a partitioning into some well-defined class of shape domains of its surface is given. In fact, you can perform this analysis in a number of ways (see Appendix). For our present purposes, it is sufficient to determine the shape groups of the objects obtained upon *truncating* a family of domains from the VDWS $G_2(1)$. We will simply truncate *all* the domains $d_i$ on the VDWS containing points at which the MEP *is smaller than a specified value* $a_1^{(i)}$. The object left after this operation will be denoted by $G^{(i)}$, and it constitutes the object for which we shall determine the shape groups. The shape groups are the *homology* groups of algebraic topology,[24-28] as applied to truncated contour surfaces. In order to describe all the essential shape features of our surfaces, it is sufficient to consider the one- and two-dimensional homology groups $H^1(G^{(i)})$ and $H^2(G^{(i)})$ for different values of the index $i$. The zero-dimensional groups will be isomorphic to the additive group of integers in all the actual cases, and they need not be considered explicitly. The groups $H^p(G^{(i)})$ arising in the present shape analysis are topological invariants. We shall focus our analysis on a simple set of numbers used for their characterization: their ranks $b_p(H^p(G^{(i)}))$, also known as the $p$-th Betti numbers.

Let us consider the example of molecule I (isoarecolone). First, we focus our attention on the MEP value $a_1^{(1)} = 0$ kcal/mole, ($i = 1$). As it is evident from Figures 3–8, no point on $G_2(1)$ has a value of the MEP below 0 kcal mol$^{-1}$. Consequently, the corresponding truncation leaves the VDWS intact; that is, the whole VDWS is left as the object of analysis, hence $G_2(1)$ coincides with $G^{(1)}$. Since the VDWS is a topological two-sphere, we know that:[24]

$$b_1(H^1(G^{(1)})) = 0, b_2(H^2(G^{(1)})) = 1 \quad (2)$$

Considering now a different value, $a_1^{(2)} = 40$ kcal mol$^{-1}$, we notice that the truncation implies that a cap at the "top" of the molecule is removed. This leaves us with a topological two-sphere with a hole, which has the following Betti numbers:

$$b_1(H^1(G^{(2)})) = 0, b_2(H^2(G^{(2)})) = 0 \quad (3)$$

A similar analysis can be performed easily for all the other selected values of $a_1$. Particularly interesting is the case of $a_1^{(5)} = 70$ kcal mol$^{-1}$. From Figures 3–8 we realize that the VDWS possesses a belt and several cap regions in the range of potential between 60 and 70 kcal mol$^{-1}$. A comparable situation appears for all the remaining molecules. In the particular case of the isoarecolone, we notice that some of the caps are neatly separated from the belt, but others are very close. Figure 5 (right view of the VDWS) shows that, in fact, one of the domains that appears as separate cap in the other molecules has *merged* into the belt; hence, it must be considered part of the belt. Nevertheless, it is clear that this situation must be close to a *bifurcation point* in the following sense: A small change in the choice of the parameter $a_1^{(5)}$ would lead to the separation of this region as a new cap (disjoint from the belt). All these modifications lead to changes in the homology groups and allow us to distinguish different shape characteristics among the molecules.

In the particular case of molecule I, we notice that the truncation for the index $i = 5$ gives rise to an object that is a topological two-sphere with seven holes. This

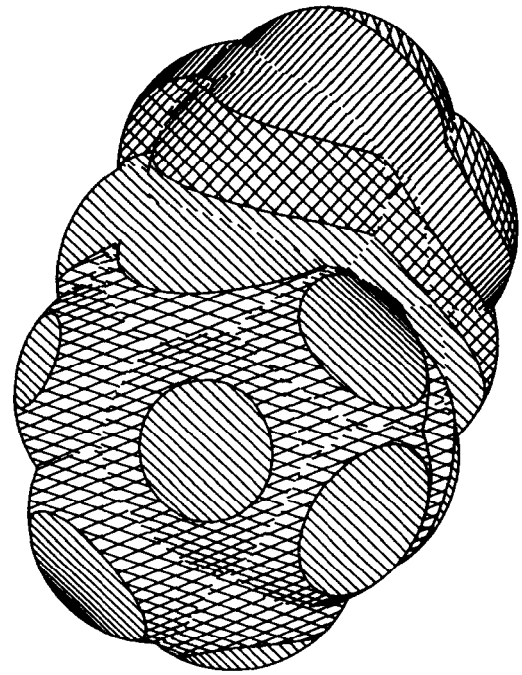

*Figure 4. Left view of the isoarecolone Van der Waals surface*

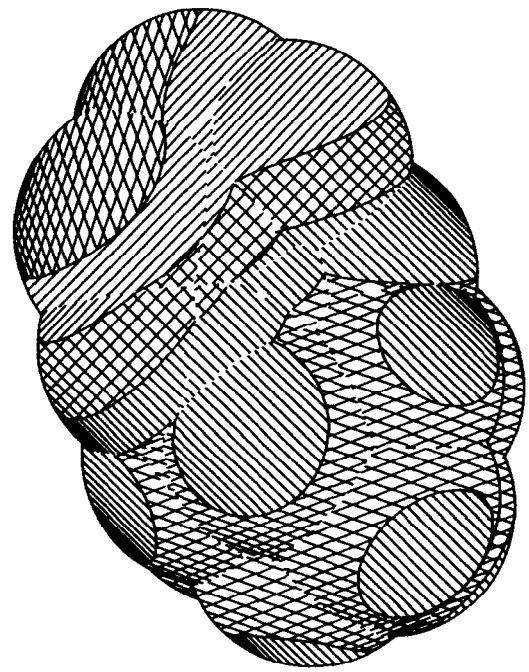

*Figure 5. Right view of the isoarecolone Van der Waals surface*

object has a one-dimensional homology group isomorphic to a free Abelian group with six generators. For such an object, the relevant Betti numbers are:

$$b_1(H^1(G^{(5)})) = 6, b_2(H^2(G^{(5)})) = 0 \quad (4)$$

The case of $a_1^{(5)} = 80$ kcal mol$^{-1}$ is also of special interest. The truncation for the index $i = 6$ eliminates all the points of the VDWS (i.e., the whole object is removed, and nothing is left for analysis). This represents a "no group" situation. Due to the fact that in the

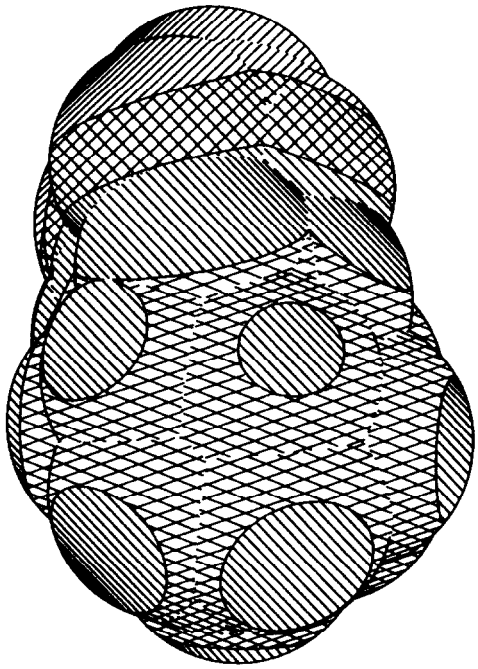

*Figure 6. Back view of the isoarecolone Van der Waals surface*

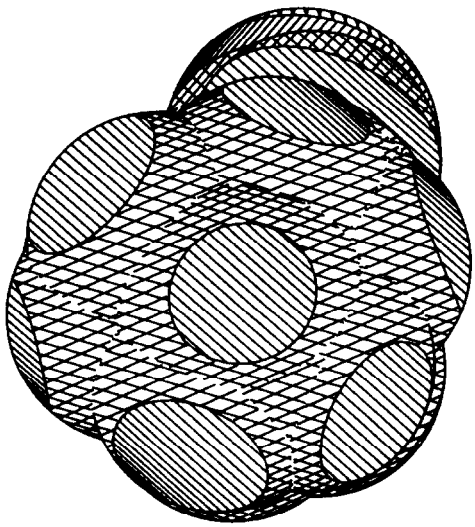

*Figure 7. Bottom view of the isoarecolone Van der Waals surface*

case of $i = 6$ one obtains the same result for all the molecules I to IV, the choice of $a_1^{(6)}$ does not provide a discriminating shape description, and it can be omitted.

As mentioned above, the range of potential between 60 and 70 kcal $mol^{-1}$ is a most interesting one for all four molecules. In order to illustrate the differences found in other cases, we can consider molecule II (isoarecolol). Figure 9 displays the right hand side view of the molecule. In this particular example, the cap that is included in the belt for molecule I is now a separate entity. The diagrams for molecule II are also close to the bifurcation point for the given range 60–70 kcal $mol^{-1}$, but in an opposite sense (with respect to the case of molecule I). As a result, the truncation with

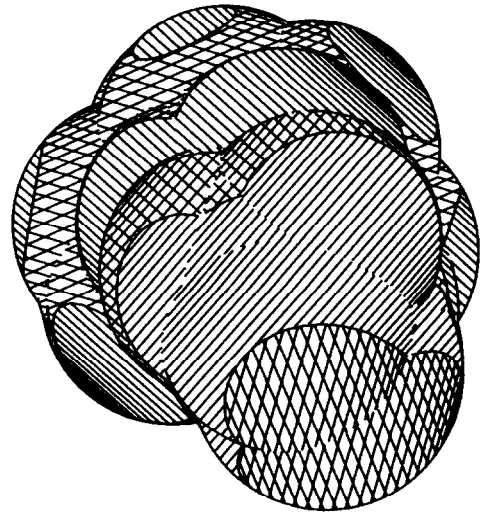

*Figure 8. Top view of the isoarecolone Van der Waals Surface*

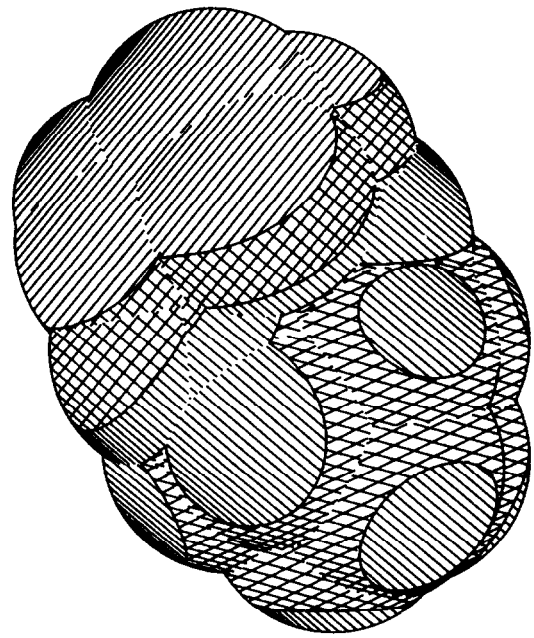

*Figure 9. Right view of the Van der Waals surface of isoarecolol (the same pattern code is used for ranges of MEP as in Figure 3)*

index $i = 5$ for isoarecolol gives us an object whose relevant Betti numbers are:

$$b_1(H^1(G^{(5)})) = 7,\ b_2(H^2(G^{(5)})) = 0 \qquad (5)$$

Tables 1 through 4 display the results of the Betti numbers $b_p(p = 1$ and $2)$ of the relevant homology groups of the four molecules. The analysis, similar to that described above for the case of isoarecolone, was performed for the four molecules using six molecular views for each. The results for these molecules show many common features, as well as some apparently small differences. In the next section, we discuss briefly the meaning of these differences and their relevance to the similarity and dissimilarity between the shapes of different molecules.

## DISCUSSION AND CONCLUSIONS

Previous studies have suggested that a cationic quaternary nitrogen is a minimum requirement expected in a compound with cholinergic activity.[37] (For references about previous models of the nicotinic pharmacophore, see, for instance, Refs. 16, 32, 33, 36, 38 and others quoted therein). In our case, all four molecules considered (related to isoarecolone) possess this feature, with two methyl groups and the ring as formal substituents. A relatively small neighborhood of the nitrogen is supposed to carry a large fraction of the positive charge.[39] This neighborhood is believed to have a major role in the biochemical activity of these compounds, and it is often regarded as the pharmacophore. We could assume hereafter that this neighborhood of the quaternary nitrogen is, indeed, the moiety primarily responsible for the medicinal activity. However, one must mention that the biological activity of the drug may be modified by replacing the substituent at the other extreme of the molecule with one of larger electronic density (for instance, a ketone with an alcohol).

As mentioned in Section 3, it appears significant that the likely pharmacophore group is located on a region of the VDWS, where the accessible values of MEP range from 60 to 80 kcal $mol^{-1}$. This corresponds to the $d_4$ and $d_5$ domains on the surface $G_2(1)$ characterized by the constants $a_1^{(4)}$ and $a_1^{(5)}$. As a result, we can assume that the essential differences in the shapes of the active moiety of the molecules should be well described by the domains $d_4$ or $d_5$. Following this observation, we can use the results in Section 3 to analyze which chemical changes affect more the shape of the active region in the VDWS — the hydrogenation of the carbon-carbon double bond or the ketone-alcohol substitution.

In the case of unsaturated molecules, Tables 1 and 2 show clearly that the change from a ketone to an alcohol affects the groups $H^p(G^{(4)})$, $H^p(G^{(5)})$, and their Betti numbers $b_p(H^p(G^{(4)}))$ and $b_p(H^p(G^{(5)}))$, indicating significant shape changes. However, within the same potential range, the same substituent replacement does not produce any change in the shape groups of the saturated compounds (Tables 3 and 4). This difference in the sensitivity of shapes to substituent replacement is caused by the caps merging into the belt on the VDWS, that occur in the potential range of from 60 to 70 kcal $mol^{-1}$ in isoarecolone (see Section 3). Within the given potential range, isoarecolol and its saturated analog have identical shape groups.

A change in the groups and their Betti numbers occurs in the case of hydrogenation (from $b_1(H^1(G^{(5)})) = 6$ in molecule I to $b_1(H^1(G^{(5)})) = 7$, in molecule III). In this case, the saturation of the double bond in isoarecolone modifies the shape of the pharmacophore region. After saturation, the pharmacophore seems to possess features very similar to those present in isoarecolol and its corresponding saturated derivative. Indeed, the homology groups $H^p(G^{(i)})$, for MEP ranges $i = 3$, 4, and 5, are exactly the same for molecules II, III and IV.

The above analysis suggests that there are major differences in the roles of the ketone-alcohol substitution and the double bond hydrogenation in causing shape changes. If we focus our attention on that part of the VDWS where the electrostatic potential takes values similar to those best characterizing the pharmacophore region, we might conclude that the molecules — particularly I, III and IV — are likely to show similar activities.

## SUMMARY

The experimental information presently available about the activity of these compounds (as potential nicotinic agonist drugs) can be complemented with the results of the MEP and VDWS calculations. To illustrate this, we have provided a detailed description of the similarity of the shape of four molecules (with regard to the neighborhood of a quarternary nitrogen). The somewhat vague, intuitive concept of molecular similarity has been

**Table 1. Betti numbers for the one- and two-dimensional homology groups of sets $G^{(i)}$ for the isoarecolone (molecule I)**

| $i$ | $b_1(H^1(G^{(i)}))$ | $b_2(H^2(G^{(i)}))$ |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 6 | 0 |

**Table 2. Betti numbers for the one- and two-dimensional homology groups of sets $G^{(i)}$ for the isoarecolol (molecule II)**

| $i$ | $b_1(H^1(G^{(i)}))$ | $b_2(H^2(G^{(i)}))$ |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 0 | 1 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 7 | 0 |

**Table 3. Betti numbers for the one- and two-dimensional homology groups of sets $G^{(i)}$ for the saturated isoarecolone (molecule III)**

| $i$ | $b_1(H^1(G^{(i)}))$ | $b_2(H^2(G^{(i)}))$ |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 7 | 0 |

**Table 4. Betti numbers for the one- and two-dimensional homology groups of sets $G^{(i)}$ for the saturated isoarecolol (molecule IV)**

| $i$ | $b_1(H^1(G^{(i)}))$ | $b_2(H^2(G^{(i)}))$ |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 0 | 1 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 7 | 0 |

replaced by a more rigorous mathematical model: a symmetry-independent group theoretical model.

Our study is based on the interpenetration of contour surfaces of two molecular functions of interest: Van der Waals surfaces and molecular electrostatic potentials. There are two ways to apply the shape group model. First, the shape analysis and the analysis of molecular similarity can be performed automatically by the computer for an entire family of molecules; it has been shown that the method of shape characterization is rather straightforward. Alternatively, if fairly complete pictorial information on the VDWS and MEP surfaces is available, as in our present examples, then the shape group method makes it possible to present the usual intuitive results, extracted from a graphic display of a molecular surface by visual inspection, in a concise and rigorous way.

## ACKNOWLEDGEMENTS

This work was partly supported by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. We wish to thank the New Jersey Governor's Commission on Science and Technology; the US Army Medical Research Division, Fort Detrick, MD; Merck, Sharp and Dohme Labs, Rahway, NJ; Schering Plough Corp., Bloomfield, NJ; Control Data Corporation and the John von Neumann Supercomputer Center for supporting this work. We also wish to thank Tripos Associates for a grant of the Sybyl software and Chemical Design for a generous academic discount of the chemX software.

## REFERENCES

1 Richards, W. G. *Quantum Pharmacology* (2nd Ed.). Butterworths, London, 1983
2 Taylor, J. L., and Durant, J. C. *J. Mol. Graph.*, 1985, **3,** 158–165
3 Smith, G. J., Macrae, C. F., and King, P. M. *J. Mol. Graph.*, 1986, **4,** 238–244
4 Scrocco, E., and Tomasi, J. *Topics Current Chem.*, 1973, **42,** 95–170
5 Scrocco, E., and Tomasi, J. *Adv. Quantum Chem.*, 1978, **11,** 115–193
6 Tomasi, J. On the use of electrostatic molecular potentials in theoretical investigations on chemical reactivity. *Quantum Theory of Chemical Reactions,* Daudel, R., *et al.*, Eds. Reidel, Dordrecht, Holland, 1979, Vol. 1, 191–228
7 Weinstein, H., *et al.* Electrostatic potentials as descriptors of molecular reactivity. *Chemical Applications of Atomic and Molecular Electrostatic Potentials,* Politzer, P., and Truhlar, D. G., Eds. Plenum, New York, 1981, 309–334
8 Kaufmann, J. J., *et al.* Electrostatic molecular potential contour maps from *Ab Initio* Calculations, same volume as in Ref. 7, 335–380
9 Pullman, A., and Pullman, B. Electrostatic molecular potential of the nucleic acids, same volume as in Ref. 7, 381–405
10 Rabinowitz, J. R., and Little, S. B. *Int. J. Quantum Chem., Quantum Biol. Symp.*, 1986, **13,** 9–18
11 Culberson, J. C., *et al. Int. J. Quantum Chem., Quantum Biol. Symp.*, 1986, **13,** 267–276
12 Náray-Szabó, G., and Surján, P. R. Computational methods for biological systems. *Theoretical Chemistry of Biological Systems,* Náray-Szabó, G., Ed., *Studies in Physical and Theoretical Chemistry,* **41.** Elsevier, Amsterdam, 1986, 1–100
13 Kollman, P. A. *Acc. Chem. Res.*, 1977, **10,** 365–371
14 Hayes, D. M., and Kollman, P. A. *J. Am. Chem. Soc., 1976,* **98,** 3335–3345
15 Cox, S., and Williams, D., *J. Comput. Chem.* 1981, **2,** 304–323
16 Kollman, P. A. *J. Am. Chem. Soc.*, 1978, **100,** 2974–2984
17 Coppens, P., and Hall, M. B., Eds. *Electron Distributions and the Chemical Bond.* Plenum, New York and London, 1982
18 Purvis III, G. D., and Culberson, C. *Int. J. Quantum Chem., Quantum Biol. Symp.*, 1986, **13,** 261–265
19 Åqvist, J., and Tapia, O. *J. Mol. Graph.*, 1987, **5,** 30–34
20 Richard, A. M., and Rabinowitz, J. R. *Int. J. Quantum Chem.*, 1987, **31,** 309–323
21 Carbó, R., Leyda, L., and Arnau, M. *Int. J. Quantum Chem.*, 1980, **17,** 1185–1189
22 Bowen-Jenkins, P. E., Cooper, D. L., and Richards, W. G. *J. Phys. Chem.*, 1985, **89,** 2195–2197
23 Hodgkin, E. E., and Richards, W. G. *J. Chem. Soc., Chem. Commun.*, 1980, 1342–1344
24 See, for instance:
Spanier, E. H. *Algebraic Topology.* McGraw-Hill, New York, 1966
Massey, W. S. *Homology and Cohomology Theory.* Dekker, New York, 1978
25 Mezey, P. G., *Int. J. Quantum Chem., Quantum Biol. Symp.*, 1986, **12,** 113–122
26 Mezey, P. G. *J. Comput. Chem.*, 1987, **8,** 462–469
27 Arteca, G. A., and Mezey, P. G. *Int. J. Quantum Chem., Quantum Biol. Symp.* 1987, **14,** 133–147
28 Mezey, P. G. *Int. J. Quantum Chem., Quantum Biol. Symp.* 1987, **14,** 127–132
29 Srebrenik, S., Weinstein, H., and Pauncz, R. *Chem. Phys. Letters,* 1973, **20,** 419–423
30 Sitaram, N., Weingartner, H., and Gill, J. C. in *Nutr. Brain,* 1979, **5,** 367–375
31 Ross, M. J., *et al. J. Mol. Biol.*, 1977, **116,** 635–659
32 Spivack, C. E., *et al. J. Pharmacol.*, 1986, **120,** 127–131
33 Waters, J. A., *et al. J. Med. Chem.* (in press)
34 Singh, U. C., and Kollman, P. A. *J. Comp. Chem.*, 1984, **5,** 129–145
35 Yadav, J. S., Hermsmeier, M., and Gund, T. M. *Int. J. Quantum Chem.* (in press)
36 Sheridan, R. P., *et al. J. Med. Chem.*, 1986, **29,** 899–906
37 Barlow, R. B. *Introduction to Chemical Pharmacology.* John Wiley and Sons, New York, 1955
38 Beers, W. H., and Reich, E. *Nature* (London), 1970, **228,** 917–922
39 Pullman, B., Courriere, P. H., and Coubeils, J. L. *Mol. Pharmacol.*, 1971, **7,** 397–405

## APPENDIX

In this appendix we discuss briefly the formal definition of the domains on the VDWS that we use to determine the relevant shape groups.

Let $f_1(\mathbf{r})$ be a molecular function of interest (in our case, the MEP), and let $G_1(a_1)$ be a contour surface of this function, associated with a value $a_1$:

$$G_1(a_1) = \{\mathbf{r} \in {}^3\mathbb{R} : f_1(\mathbf{r}) = a_1\} \tag{A.1}$$

Let us now introduce a second function $f_2(\mathbf{r})$ that will allow us to define the second surface $G_2(a_2)$, that in the typical case has common points with $G_1(a_1)$. We refer to such a situation as interpenetration. In order to describe a VDWS, we assume, without any loss of generality, that $f_2(\mathbf{r})$ is equal to unity on the surface and zero elsewhere.[28] As a consequence, the VDWS becomes the following contour (a formal level set):

$$G_2(1) = \{\mathbf{r} \in {}^3\mathbb{R} : f_2(\mathbf{r}) = 1\} \tag{A.2}$$

There exists, of course, an infinite number of contours $G_1(a_1)$ interpenetrating $G_2(1)$. However, we can arrange those infinitely many contours into *ranges*. As a result, we can define *domains* of the function $f_1$ (molecular electrostatic potential) over the contour of the function $f_2$ (Van der Waals surface). To that purpose, we proceed in the following way. Let $A$ be a set of real numbers:

$$A = \{a_1^{(1)}, a_1^{(2)}, \ldots, a_1^{(N)}\},$$
$$a_1^{(i)} < a_1^{(i+1)},\ i \leq N,\ N < \infty \tag{A.3}$$

representing some arbitrary, but accessible values for the function $f_1(\mathrm{r})$. Accordingly, we can define various domains $d_i$ on $G_2(1)$ as follows:

$$d_i = \{\mathbf{r} \in G_2(1) : f_1(\mathbf{r}) \in [a_i, a_{i+1})\} \tag{A.4}$$

These domains $d_i$ are the ranges of MEP indicated with different patterns in Figures 3–8. In the typical case, these domains resemble "belts" and "caps." One might study the shape groups of the 3D objects obtained when a given domain $d_i$ ($i = 1, 2, \ldots, N$) is eliminated from $G_2(1)$. Of course, this might lead to a single closed object (topological sphere) or to an object with a number of holes (punctured spheres) or to a number of similar but disjoint objects. This analysis is general for all molecules, and it provides a characterization of their shape. As an alternative, we can define a collection of domains $D_i$ as:

$$D_i = \bigcup_{j=1}^{i} d_j \tag{A.5}$$

and study the homology groups of the simplicial cell complex related to the set $G^{(i)}$ obtained as:

$$G^{(i)} = G_2(1)\, D_i \tag{A.6}$$

The set $G^{(i)}$ is the remainder of the VDWS, when all the points on it for which the MEP is smaller than the value $a_1^{(i)}$ are eliminated. In Section 3 the results for the actual shape groups of $G^{(i)}$ of the examples are discussed, for several values of $i$, and for a series of VDWSs.