BUTTERWORTH HEINEMANN

# Interrelation between electrostatic and lipophilicity potentials on molecular surfaces

**Isabel Rozas,*† Qishi Du,‡ and Gustavo A. Arteca‡**

**Department of Chemistry, Queen's University, Kingston, Ontario, Canada †Permanent address: Instituto de Química Médica, CSIC, Madrid, Spain ‡Département de Chimie et Biochimie, Laurentian University, Sudbury, Ontario, Canada*

*Molecular electrostatics and lipophilicity are two important properties included in quantitative structure–activity relationships (QSARs) employed for rational drug design. The molecular electrostatic potential (MEP) provides information on the position, distribution, and extent of electrophilic and nucleophilic regions around a molecule. Similarly, the solvent affinity can be represented by a local phenomenological potential of semiempirical nature: the molecular lipophilicity potential (MLP). Although a simultaneous, three-dimensional display of MEP and MLP is possible, it may not provide a practical tool for comparing molecules. In this work, we deal with the simpler two-dimensional maps of the entire molecular surface projected onto an MEP–MLP plane. We analyze how these maps change with the following factors: (1) composition and molecular geometry, (2) the quality of the computation of MEP, and (3) the parameter set used for evaluating the lipophilicity potential. The approach is used to compare series of pyrazole derivatives. The methodology is useful in assessing molecular similarity, as well as in establishing the nature of differences between compounds.*

## INTRODUCTION

Within the context of rational drug design, the simultaneous analysis of various molecular properties is a useful approach by which to compare molecules, and possibly highlight similarities. Selecting those properties that allow greater discrimination is a key step toward the development of quantitative structure–activity relationships (QSARs).

Some biological actions are known to correlate with microscopic properties such as molecular volumes, surface areas, and molecular electrostatic features. Important electrostatic characteristics include nucleophilic and electrophilic reactive positions of a molecule. These can be conveyed by the molecular electrostatic potential (MEP) on a single, isolated molecule. In contrast, other electrostatic properties are the result of averaged interactions of the molecule with its surroundings (e.g., the solvent). This average is commonly represented by a "phenomenological interaction" such as the lipophilicity (or hydrophobicity) "force."[1] Whereas there is no simple satisfactory approach to represent the hydrophobic interaction in terms of local solvent–solute electrostatics, there are a number of methods for estimating the degree of lipophilicity at global and local levels.

In this work, we employ an approach in which electrostatic properties and lipophilicity are mapped simultaneously onto a surface. As a result, we present the two-dimensional (2D) electrostatics–lipophilicity maps of a molecular surface. These maps are used as a tool with which to recognize reactivity features and assess molecular similarity. Our approach to the study of these two properties is as follows:

**1.** Size, volume, surface area, and overall three-dimensional (3D) shape can all be derived from a molecular surface. In our case, we use the van der Waals surface, as defined by the intersection of fused atomic spheres. Other alternative model surfaces can also be used.[2–4]

**2.** Molecular electrostatics is usually described by the MEP function $V(\mathbf{r})$. This function is actually the electric potential energy associated with a positive point charge brought from infinity to a point $\mathbf{r}$ on the molecular surface. The function $V(\mathbf{r})$ is derived from the one-electron density function at a given molecular conformation. The minima and maxima of $V(\mathbf{r})$ on the molecular surface serve to locate positions open to electrophilic and nucleophilic attacks, respectively.[5]

**3.** Molecular lipophilicity can be described by a single number using a macroscopic property such as the partition coefficient $P$ for a solute and a pair of "immiscible" sol-

Address reprint requests to Dr. Arteca at the Département de Chimie et Biochimie, Laurentian University, Ramsey Lake Road, Sudbury, Ontario, Canada P3E 2C6.

Received 1 June 1994; revised 1 October 1994; accepted 1 December 1994

Journal of Molecular Graphics 13:98–108, 1995

655 Avenue of the Americas, New York, NY 10010

0263-7855/95/$10.00
SSDI 0263-7855(94)00017-M

vents, for example, *n*-octanol–water. Several techniques have been developed to fit experimental log *P* values by using atomic or functional group contributions.[6–10]

Other lipophilicity indices have been proposed.[11–19] Among these, we shall deal here with the so-called molecular lipophilicity potential (MLP) function $L(\mathbf{r})$, which measures lipophilicity at a given point on a molecular surface.[11] This function is evaluated by using atomic fragmental contributions to the lipophilicity.[6,10] In this work, we analyze in detail the qualitative interrelations between the functions $V(\mathbf{r})$ and $L(\mathbf{r})$ for a number of molecules.

The work is organized as follows. In the next section we discuss the strategies for the computation of the 2D electrostatics–lipophilicity maps of molecular surfaces. The section "Electrostatics–lipophilicity maps for small molecules" shows the application of the methodology to the study of small molecules. We use small molecules to illustrate the changes in the maps due to the atomic basis set, and the set of lipophilicity parameters. As well, we analyze the effect of substituents on the distribution of MEP and MLP. In "Properties of (*V*, *L*) maps for substituted pyrazoles," we use the method to study several pyrazole derivatives. The electrostatic and lipophilicity properties of these compounds are analyzed on the total molecular surface, as well as on some selected groups of atoms (e.g., the five-member ring atoms and the substituent side chain atoms). Strategies using this methodology to assess molecular similarity are proposed, including the use of difference maps.

## COMPUTATION OF ELECTROSTATIC–LIPOPHILICITY MAPS

A number of molecules have been characterized by their corresponding MEP function $V(\mathbf{r})$ and MLP function $L(\mathbf{r})$, at each point $\mathbf{r}$ of a van der Waals surface. We have used the familiar distribution of points on the van der Waals surface developed by Connolly.[20] (That is, we do not use a solvent probe to "smooth" the surface.) The electrostatic potential is computed at each point by integrating over the ab initio electron density as follows (atomic units are used throughout):

$$V(\mathbf{r}) = \sum_{\alpha=1}^{n} \frac{Z_\alpha}{\|\mathbf{R}_\alpha - \mathbf{r}\|} - \sum_{i,j} A_{ij} \int \varphi_i^*(\mathbf{r}') \, \varphi_j(\mathbf{r}') \frac{d\mathbf{r}'}{\|\mathbf{r}' - \mathbf{r}\|} \qquad (1)$$

for a molecule with $n$ atoms, where $\{Z\alpha\}$ and $\{\mathbf{R}_\alpha\}$ are the set of nuclear charges and position vectors, respectively. The marginal density function in Eq. (1) is expressed in terms of the elements of the density matrix $\{A_{ij}\}$, and a basis set of atomic orbitals $\{\varphi_i\}$. The actual computations of MEP are performed with a version of the program Gaussian 80,[21] adapted by Singh and Kollman[22] to study electrostatics on points defined on molecular surfaces.[20] The van der Waals radii are taken from Ref. 23. The local contribution to the lipophilicity at each point $\mathbf{r}$ can be estimated by a comparable (although semiempirical) potential. The function we use in this work is the "lipophilicity potential" $L(\mathbf{r})$ proposed by Croizet et al.[15]:

$$L(\mathbf{r}) = \sum_{\alpha=1}^{n} \frac{f_\alpha}{1 + \|\mathbf{R}_\alpha - \mathbf{r}\|} \qquad (2)$$

where $\|\mathbf{R}_\alpha - \mathbf{r}\|$ is the distance between the atomic nucleus $\alpha$ and the point $\mathbf{r}$ on the van der Waals surface. The $L(\mathbf{r})$ value can be either positive or negative. The contribution of each atom to the lipophilicity, represented by the atomic lipophilicity parameters $f_\alpha$, is evaluated from empirical fitting to log *P* data.[24] A comprehensive database of these parameters is given in Ref. 6. Function (2) incorporates an analytical dependence on the distance that mimics the long-range behavior of electrostatic interactions (a nonempirical dependence). Function $L(\mathbf{r})$ measures lipophilicity in a direct fashion: the larger the $L(\mathbf{r})$ value, the more lipophilic the position $\mathbf{r}$. The lower $L(\mathbf{r})$ values signal the hydrophilic regions of the molecule. Similarly, the low $V(\mathbf{r})$ values indicate nucleophilic positions (i.e., positions susceptible to electrophilic attack), whereas high $V(\mathbf{r})$ values correspond to electrophilic positions (i.e., positions susceptible to nucleophilic attack). Electron-rich positions are thus expected to appear at low, possibly negative, electrostatic potential energy values. In many cases, one would expect that positions with low MEP values correspond to atoms that can act as acceptors in forming hydrogen bonds. However, this is not always the case. A proper description must explore hydrophilic and nucleophilic positions with separate, independent criteria. In this work, our approach is to analyze such interrelations in two-dimensional diagrams of electrostatics and lipophilicity, referred to in what follows as (*V*, *L*) maps.

In the above-described maps, a point on the molecular surface is represented by its *V* and *L* values, independently of its position on the surface. Points that are close to each other on a molecular surface may appear close in a (*V*, *L*) map. However, a constant density of points on the surface will not be reflected by a uniform distribution of points on the (*V*, *L*) map. In the most extreme case, the entire surface for an isolated atom will be represented by a single point on the (*V*, *L*) map. In the case of molecules, the actual distribution of points in the latter diagrams will depend in part on the shape of the molecular surface, in particular its changes in local curvature. In the next sections we present examples of such diagrams for a number of molecules and discuss their properties.

## ELECTROSTATICS–LIPOPHILICITY MAPS FOR SMALL MOLECULES

Some properties of (*V*, *L*) maps are illustrated using methane and chloromethane derivatives. Figure 1 shows the results for methane at its experimental equilibrium geometry. The van der Waals surface has been constructed with a density of 25 points per Å$^2$ ($\delta$ = 25 pt/Å$^2$), using Connolly's algorithm with no solvent probe.[20] The MEP values are derived from Eq. (1) using ab initio wavefunctions evaluated with two different atomic basis sets. The MLP [Eq. (2)] is computed with the lipophilicity parameters for H and

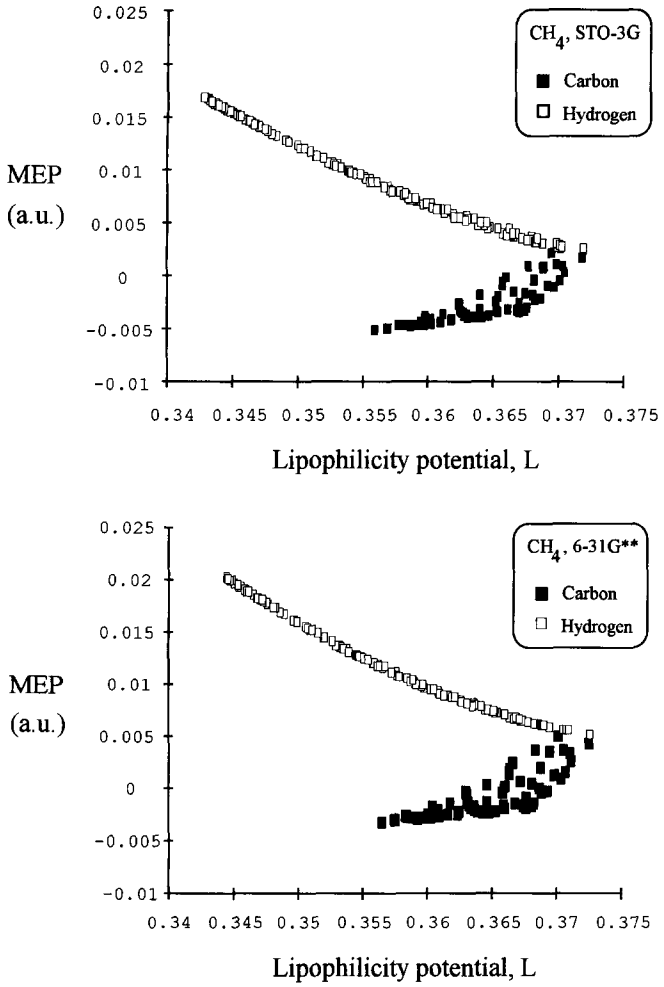


*Figure 1. Two-dimensional electrostatics–lipophilicity maps [2D (V, L) maps] of the van der Waals surface of $CH_4$. The diagrams show the electrostatic potential computed at RHF ab initio level using the STO-3G (top) and 6-31G** (bottom) basis sets. The molecular surface is represented with a density of 25 pt/Å$^2$. The lipophilicity contributions per atom are taken from Ref. 10 [cf. Eq. (2)].*

C atoms given in Ref. 10. Each point in Figure 1 represents one point on the molecular surface.

For methane, an extreme change in the basis set from STO-3G to 6-31G** produces a systematic raising of the electrostatic potential throughout the surface by an amount ranging from 0.005 to 0.002 au. (Note that 0.001 au for MEP is equivalent to 0.6275 kcal/mol.) Despite this difference in values, Figure 1 indicates that, with a fixed nuclear geometry, the essential structure of the (*V, L*) map depends little on the basis set. The results in Figure 1 involve two different electron density functions associated with the same (experimental, nonoptimized) molecular geometry. We have also verified that the basic features of the maps do not change when the molecular geometry is optimized at the ab initio level within each corresponding basis set.

Because all hydrogen atoms are equivalent, there are only two distinctive regions in the map in Figure 1, one corresponding to C atoms and another to H atoms. These regions would appear continuous if a sufficiently high density of points were used. The density of points inside the C and H regions is not constant. As mentioned before, this nonuniform density reflects not only the distribution of potential on the surface, but also the shape features of the surface.

A simple example illustrates this notion. Consider, for instance, a given atomic sphere, and select only those points on the sphere lying farther away from all other spheres. At these points, the surface properties depend least on all other atoms ("the environment"). Therefore, these points should share similar electrostatic potential and lipophilicity. Such a region on a van der Waals surface with constant density of points will be translated into a region on the (*V, L*) map with a higher density of points. In contrast, for points on the van der Waals surface lying near the boundary of two atomic spheres, electrostatic and lipophilicity properties will be strongly influenced by neighboring atoms. In this case, these points will be more separated on a (*V, L*) map, because their *V* and *L* values will change rapidly over the molecular surface. In conclusion, one can view the changing density of points in a (*V, L*) map as providing information on the surface derivatives of MEP and MLP. The higher the density of points on a (*V, L*) map, the smaller the local change of MEP and MLP over the region of the molecular surface where the points are located.

The dependence of MEP with the MLP is almost linear for H but shows a marked dispersion for C. The C and H regions in the diagram meet at maximum values of lipophilicity and intermediate values of MEP. Because at these points the distinction between the atoms disappears, this section of the map must correspond to the arc where the C and H van der Waals atomic spheres intersect. Consequently, the map indicates that the less lipophilic (i.e., the "less hydrophobic") positions correspond to the region of the H sphere farthest from the carbon nucleus. This region also presents the maximum MEP. In other words, the positions for nucleophilic attack and maximum affinity for water coincide on the hydrogen sphere. This result is not surprising because the hydrogen atoms are depleted of charge by the carbon atom. In contrast, the positions for electrophilic attack and minimum affinity for water do not coincide. The lower MEP values are found on C, at the points farther away from the H atomic spheres. The minimum water affinity would be expected, as mentioned above, at the boundary between two atomic spheres.

In conclusion, a map such as the one in Figure 1 indicates the location of regions on the van der Waals surface that are relevant to reactivity and solvent affinity. The interrelation between these two properties does not appear to be much affected by the change of basis set. Some features in the map are unexpected, such as an almost linear relation between MEP and MLP for some atoms. Moreover, the extension of the regions in the map illustrates the available surface area.

Figure 2 illustrates the change in (*V, L*) maps caused by a different choice of lipophilicity parameters $\{f_\alpha\}$. Two sets of atomic parameters are considered here. The set given by Ghose and Crippen[10] contains the atomic contributions to hydrophobicity of 90 "atom types," derived from the octanol–water partition coefficients of 494 molecules. The second set, given by Viswanadhan et al.,[6] improves on the latter by using the experimental data on 893 compounds to derive the 120 atom type contributions. With these two data sets, we can assess the qualitative dependence of our results

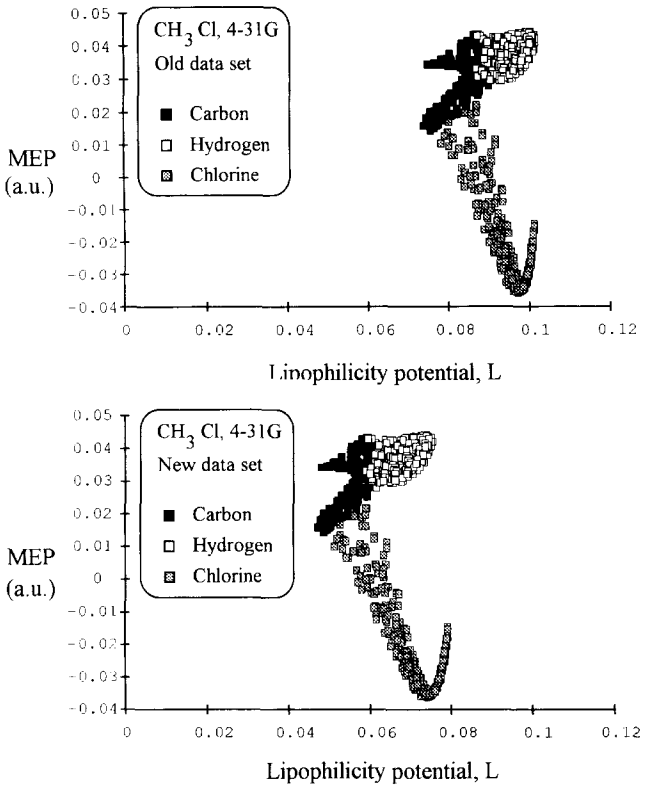


*Figure 2. Differences in 2D (V, L) maps due to the set of lipophilicity parameters $\{f_\alpha\}$. The diagrams show the computation of lipophilicity for $CH_3Cl$ by using the set of $\{f_\alpha\}$ parameters from Ref. 10 ("old data set"), at the top, and the parameters from Ref. 6 ("new data set"), at the bottom.*

on the quality of the experimental calibration for the atomic hydrophobicities.

The molecule of methyl chloride is taken as an example. The top map in Figure 2 represents the MLP computed with the parameter set in Ref. 10 ("old data set"), whereas the bottom map has been computed with the parameter set in Ref. 6 ("new data set"). In both cases, the MEP has been calculated at the ab initio level with the 4-31G basis set. The nuclear geometry was minimized at the same computational level. The molecular (van der Waals) surface is represented with a density of $\delta = 25$ pt/Å$^2$. As Figure 2 indicates, the change in parameter set causes few local changes in the $(V, L)$ map. A small modification in the lipophilicity of the Cl atom is noticeable near the minimum of potential energy. Otherwise, the new data set causes a global change in the map: the entire molecular surface is displaced systematically toward lower lipophilicity. A similar behavior is found in other compounds, although the actual value of the displacement in $L$ depends on the molecule. For $CH_3Cl$, the absolute displacement is approximately $0.025 \pm 0.005$. In conclusion, the essential shape of the $(V, L)$ map is not changed by the choice of lipophilicity parameters $\{f_\alpha\}$, except for a shift along the $L$ axis.

A comparison of Figures 1 and 2 indicates that the more radical changes in the electrostatics–lipophilicity maps are caused by chemical substitution. Figure 2 shows the presence of three distinct regions. Because the Cl atom has the largest atomic radius, it spans the largest of these regions. The regions corresponding to C and H atoms are utterly changed with respect to the results in Figure 1. Even though the H atoms are the most susceptible to nucleophilic attack in $CH_3Cl$ and $CH_4$, their lipophilicity properties are reversed. The molecular surface about the H atoms (and farthest from the C atom) corresponds to a lipophobic region in $CH_4$ and to a lipophilic one in $CH_3Cl$. The region about the Cl atom also presents interesting properties: despite being electron rich, some points on the Cl sphere are as lipophilic as points on the H sphere. In contrast, the region where the C and Cl spheres meet is the most lipophobic in $CH_3Cl$. As a final observation, note that the substitution of one H for one Cl causes a large change in the span of MEP values, whereas the span of $L$ values remains almost the same.

Figure 3 completes the analysis of the role of substituents by comparing chloromethane, dichloromethane, and trichloromethane. (Note that the $L$ axis has the same span, $\Delta L = 0.05$, for all three compounds). The same basis set (4-31G) has been employed to compute the MEP. Molecular geometries have been optimized at the same level. The computation of lipophilicity has been carried out with the $\{f_\alpha\}$ parameters in Ref. 6.

Figure 3 indicates that the number of Cl atoms has a large effect on $(V, L)$ maps. Despite their comparable nucleophilicity, the Cl atoms in $CH_3Cl$ and $CH_2Cl_2$ have reversed lipophilicity behavior. In the case of $CHCl_3$, we find a virtually constant MLP value (negative) throughout the molecular surface. (Owing to the larger atomic radius, the molecular surface area is contributed mostly by Cl.) This latter molecule presents the largest affinity for water.

In summary, despite their similar compositions, the chloroderivatives of methane give rise to distinct $(V, L)$ maps. The interrelation between MEP and MLP indicates a relative independence between electrophilic and lipophobic positions on the molecular surface.

## PROPERTIES OF $(V, L)$ MAPS FOR SUBSTITUTED PYRAZOLES

A series of four derivatives of pyrazole has been considered to analyze the discriminating power of the electrostatics–lipophilicity maps. Pyrazole derivatives are relevant to medicinal chemistry because they may act as inhibitors of liver alcohol dehydrogenase. Molecular electrostatics can be used to explain the binding of pyrazoles to the enzyme active site. In addition, it is known that the inhibitory power of these compounds depends on the position and nature of any substituents on the pyrazole ring.[25,26] In the case of 4-substituted pyrazole derivatives, the inhibitory power increases with the lipophilicity of the substituent group. As a rule, substitution in positions other than position 4 decreases the inhibitory power. (A more detailed discussion can be found in Ref. 27). From this background information, it is clear that both electrostatics and lipophilicity are relevant to the biochemical characterization of these compounds.

The derivatives considered are 4-methylpyrazole, 4-propylpyrazole, 4-hydroxycarbonylpyrazole, and 3-methylpyrazole. With these molecules, we can test the following

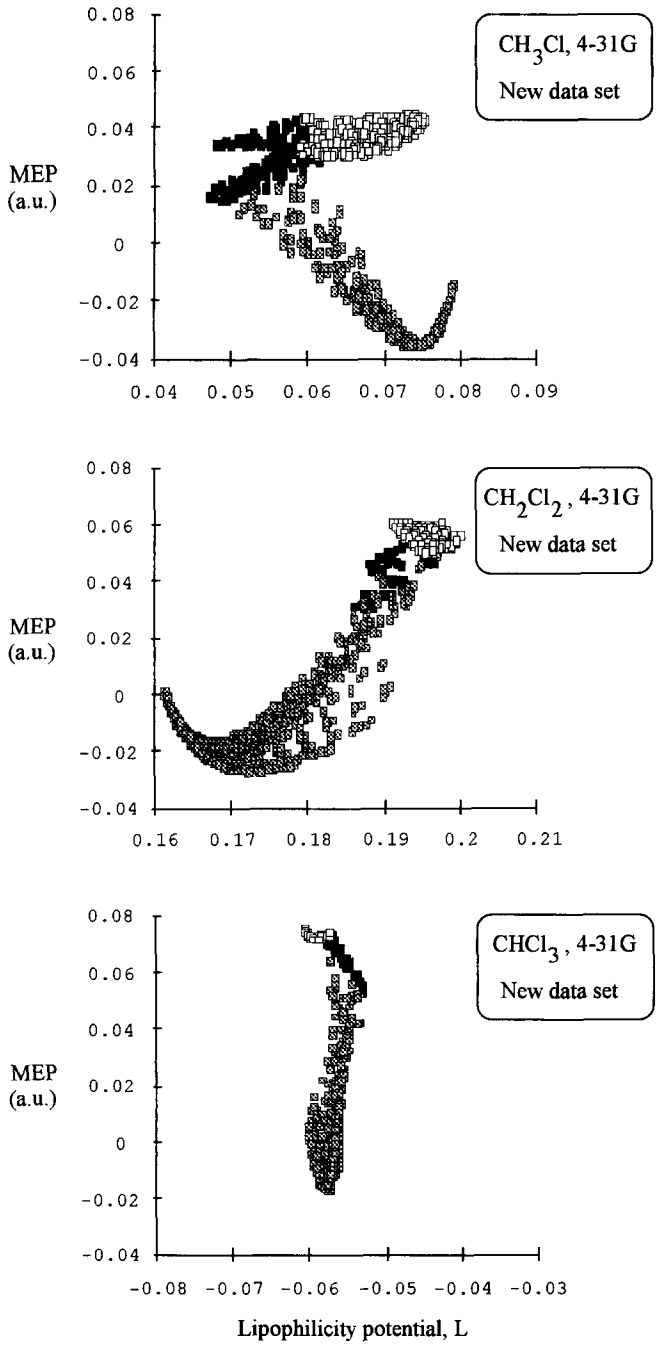


*Figure 3. Role of substituents on 2D (V, L) maps. The diagrams compare the series of molecules $CH_3Cl$, $CH_2Cl_2$, and $CHCl_3$. The MEP has been computed at the ab initio level using the basis set 4-31G. The nuclear geometries are optimized at the same level. The lipophilicity potential is evaluated with the parameter set from Ref. 6, with a surface density of $\delta = 25$ pt/Å$^2$. The convention for the shading of points according to the atom type is the same as in Figures 1 and 2.*

effects on (*V, L*) maps: (1) lipophilic substituents of variable length in position 4, (2) a lipophobic substituent, and (3) a substituent in a position other than position 4.

In this section we describe only essential qualitative features of the (*V, L*) maps for these compounds. For this reason, we have adopted an approach in which both the nuclear geometry and the electron density are only roughly approximated. For the actual computation, we have proceeded as follows:

**1.** For each compound, we have searched the conformational space using the semiempirical method AM1.[28] Conformers were generated by sequential changes of 30° in all the torsions (dihedral angles) involving atoms in the substituent side chains. After minimizing the conformers over all internal coordinates (including those for the pyrazole ring), the lowest potential energy minimum was retained.

**2.** The MEP has been computed at the ab initio level using a minimal basis set at the AM1-optimized geometry of the lowest energy conformer. The potential is evaluated on a grid of points on the van der Waals molecular surface, derived with the Connolly algorithm,[20] using a surface density of $\delta = 30$ pt/Å$^2$ and no solvent probe. The atomic radii are those in Ref. 23. (Regarding changes in MEP distribution for other choices of atomic radii, see discussion in Ref. 29).

**3.** The MLP has been computed at the points on the molecular surface with the lipophilicity parameter sets given in Refs. 6 and 10.

These steps are sufficient to our present discussion of qualitative behavior on the electrostatics–lipophilicity maps. If desired, more refined diagrams can be constructed by using a larger basis set of atomic orbitals and a higher density of points on the molecular surface.

First, we have analyzed the (*V, L*) maps of pyrazole derivatives with substituents of different lipophilicity nature. Figure 4 shows the results for 4-propylpyrazole and 4-hydroxycarbonylpyrazole, corresponding to lipophilic and hydrophilic substituents, respectively. Figure 4 gives the (*V, L*) maps computed with both sets of lipophilicity parameters. All results are presented in the same scale of *V* and *L*. For a simpler analysis, we have dropped the graphic distinction between the various atoms in the maps.

As found in the previous section, the main effect of a change in parameter set is a uniform shift along the *L* axis: when using the parameters given in Ref. 6, the *L* values are reduced by approximately 0.1. In addition, the map for 4-hydroxycarbonylpyrazole shows a smaller span of *L* values. The two compounds in Figure 4 can readily be distinguished by the interrelation of their molecular surface properties.

In 4-propylpyrazole, the more hydrophilic region (low *L*) and more electrophilic region (high *V*) of the map correspond to the H atom on the pyrazole ring that is bonded to N (position 1). Conversely, the more nucleophilic part of the map corresponds to the N atom on the pyrazole ring not bonded to hydrogen (position 2). In this latter case, the position exhibits only intermediate values of lipophilicity. In the case of 4-hydroxycarbonylpyrazole, we observe that not only is the –COOH substituent group more lipophobic than the $-(CH_2)_2CH_3$ group, but that the entire molecule is more lipophobic as well.

The electrostatic and lipophilicity properties of these molecules can be further clarified by analyzing only the

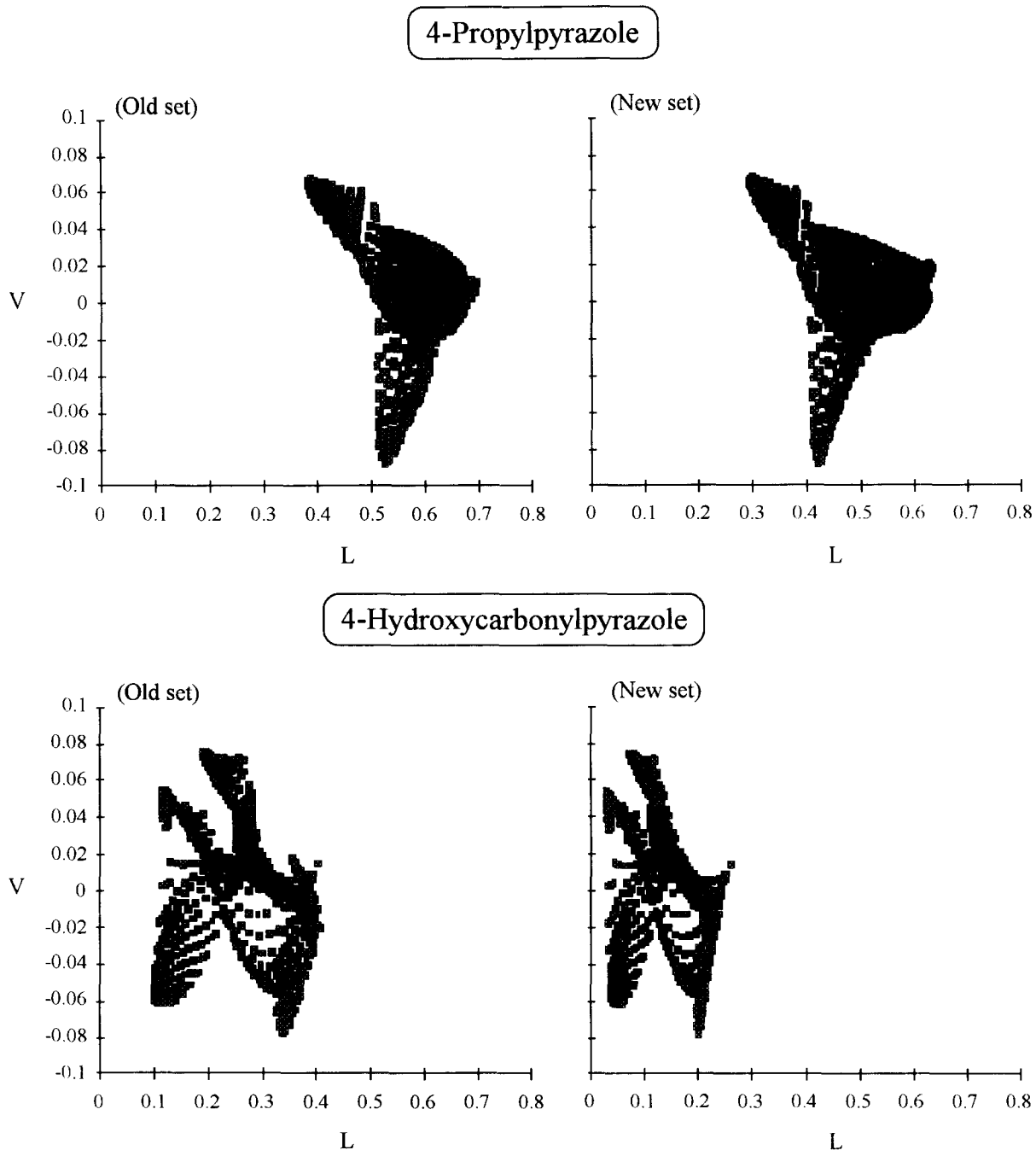


*Figure 4. Two-dimensional (V, L) maps of two 4-substituted pyrazoles. The top diagrams display 4-propylpyrazole, and the bottom ones display 4-hydroxycarbonylpyrazole. For a rough estimation, the electrostatic potential has been evaluated at the ab initio level with a minimum basis set, on the AM1-optimized global minimum. Both potentials are computed on a van der Waals surface with a density* $\delta = 30$ *pt/Å*$^2$. *The atomic lipophilicity parameters are taken from Ref. 10 (left-hand plots) and Ref. 6 (right-hand plots). MEP values (V) are in atomic units.*

partial (*V*, *L*) maps for the side chains and the pyrazole ring. Figures 5 and 6 display these partial maps for 4-propylpyrazole and 4-hydroxycarbonylpyrazole, respectively. Only the lipophilicity values computed with the "new data set" (Ref. 6) are given. The left-hand diagrams in Figures 5 and 6 correspond to the (*V*, *L*) maps for all the atoms in the substituent groups. The right-hand diagrams correspond to only the main atoms on the pyrazole ring ($C_3N_2$). From Figures 5 and 6, we can make the following observations:

- The propyl group appears as a compact region on the map, with high values of MLP and intermediate values of MEP. No internal structure can be discerned within the set of points. In contrast, the –COOH substituent group appears as an open region, with low values of MLP and spanning all possible values of MEP. In this case, the internal structure of the substituent group is evident: the two extended regions in the low MEP values correspond to two different types of oxygen atoms.

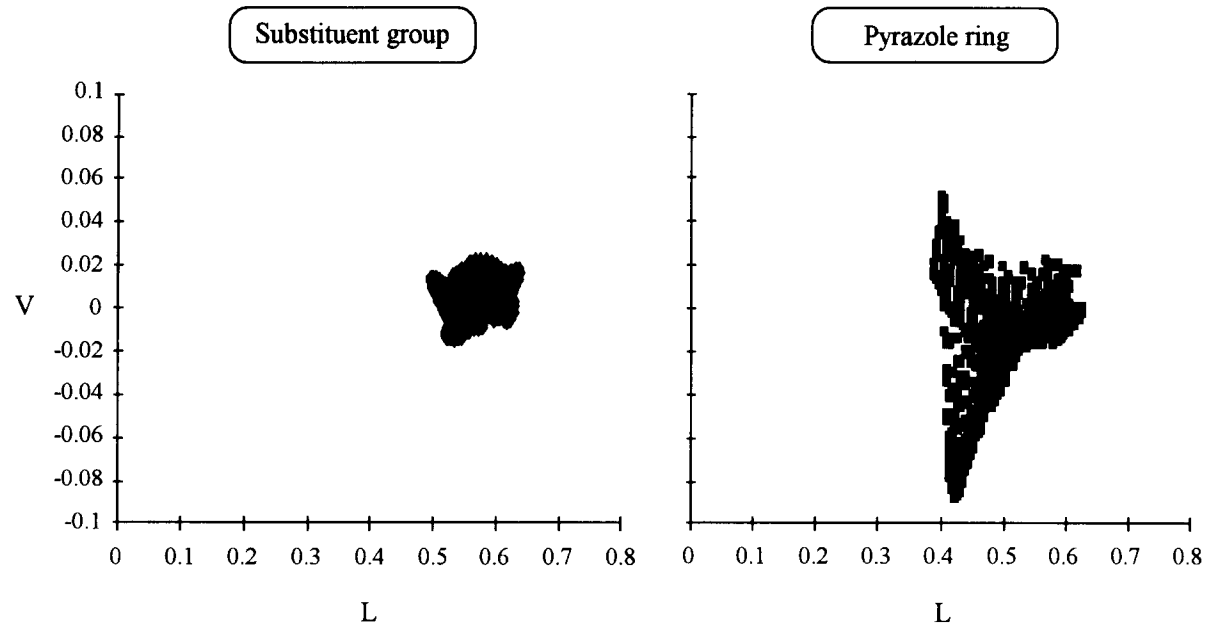


*Figure 5. Partial electrostatics–lipophilicity maps, restricted to substituent groups (left) and ring atoms (right), in 4-propylpyrazole. The MEP value is computed at the ab initio level with a minimal basis set at the AM1-optimized global minimum. The MLP value is computed with the parameters in Ref. 6. The MEP values (V) are in atomic units. The diagram on the right shows only the main atoms on the ring (i.e., only C and N). The diagram on the left corresponds to all the atoms in the $-(CH_2)_2CH_3$ side chain.*

- Although the pyrazole ring is affected by the nature of the substituent group, some common features can be noticed between the two molecules (cf. right-hand diagrams in Figures 5 and 6). In both cases, the highest and lowest values of MEP correspond to nitrogen atoms. The highest lipophilicity is found at the carbon atoms, which for both molecules take intermediate values of MEP.

In conclusion, the (*V*, *L*) maps quantitate the change in properties of pyrazoles caused by two different substituents

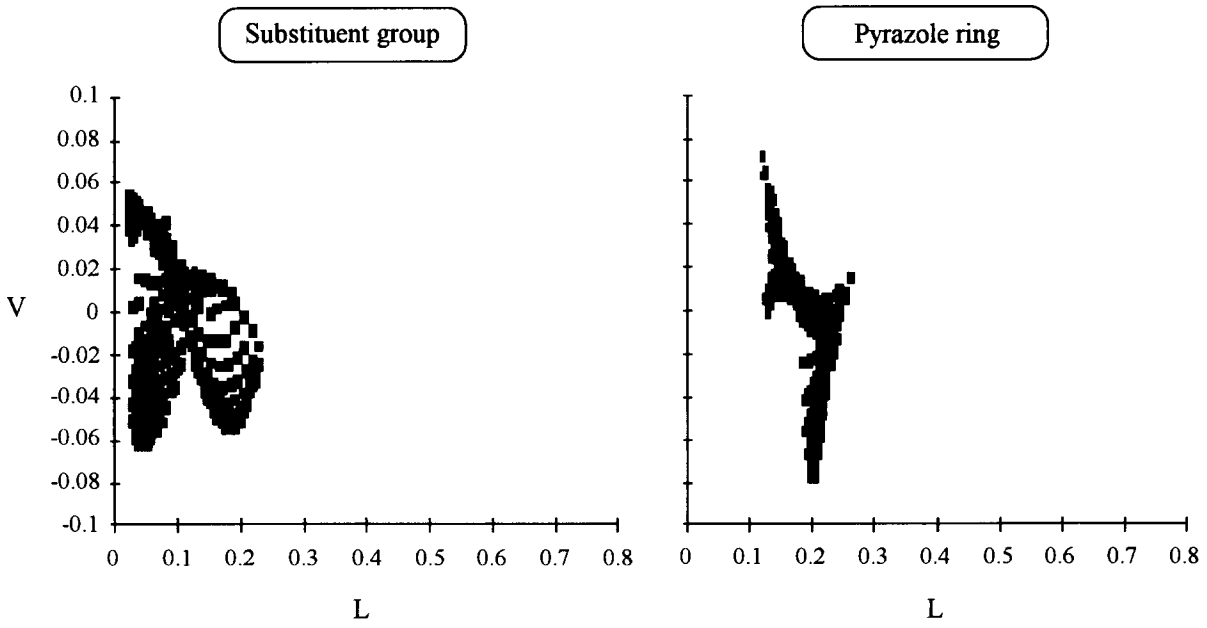


*Figure 6. Partial electrostatics–lipophilicity maps, restricted to substituent groups (left) and ring atoms (right) in 4-hydroxycarbonylpyrazole. The MEP value is computed at the ab initio level with a minimal basis set at the AM1-optimized global minimum. The MLP value is computed with the parameters in Ref. 6. MEP values (V) are in atomic units. The diagram on the right shows only the main atoms on the ring (i.e., only C and N). The diagram on the left corresponds to all the atoms in the –COOH side chain.*

in position 4. The distribution of points on a (*V*, *L*) map allows one to recognize similarities and differences between two related molecules.

Figure 7 shows the result of a change in the substituent position. The diagrams present 3-methylpyrazole and 4-methylpyrazole computed with the lipophilicity parameters given in Refs. 6 and 10. The MEP has been computed as in Figure 4. The (*V*, *L*) maps for the two methylpyrazoles exhibit many common features. Yet the following differences are noted.

- On average, the surface of 3-methylpyrazole is displaced toward lower lipophilicity values with respect to 4-methylpyrazole.
- The nucleophilic N atom in the pyrazole ring (low MEP value) is more lipophilic in the 3-methyl derivative.

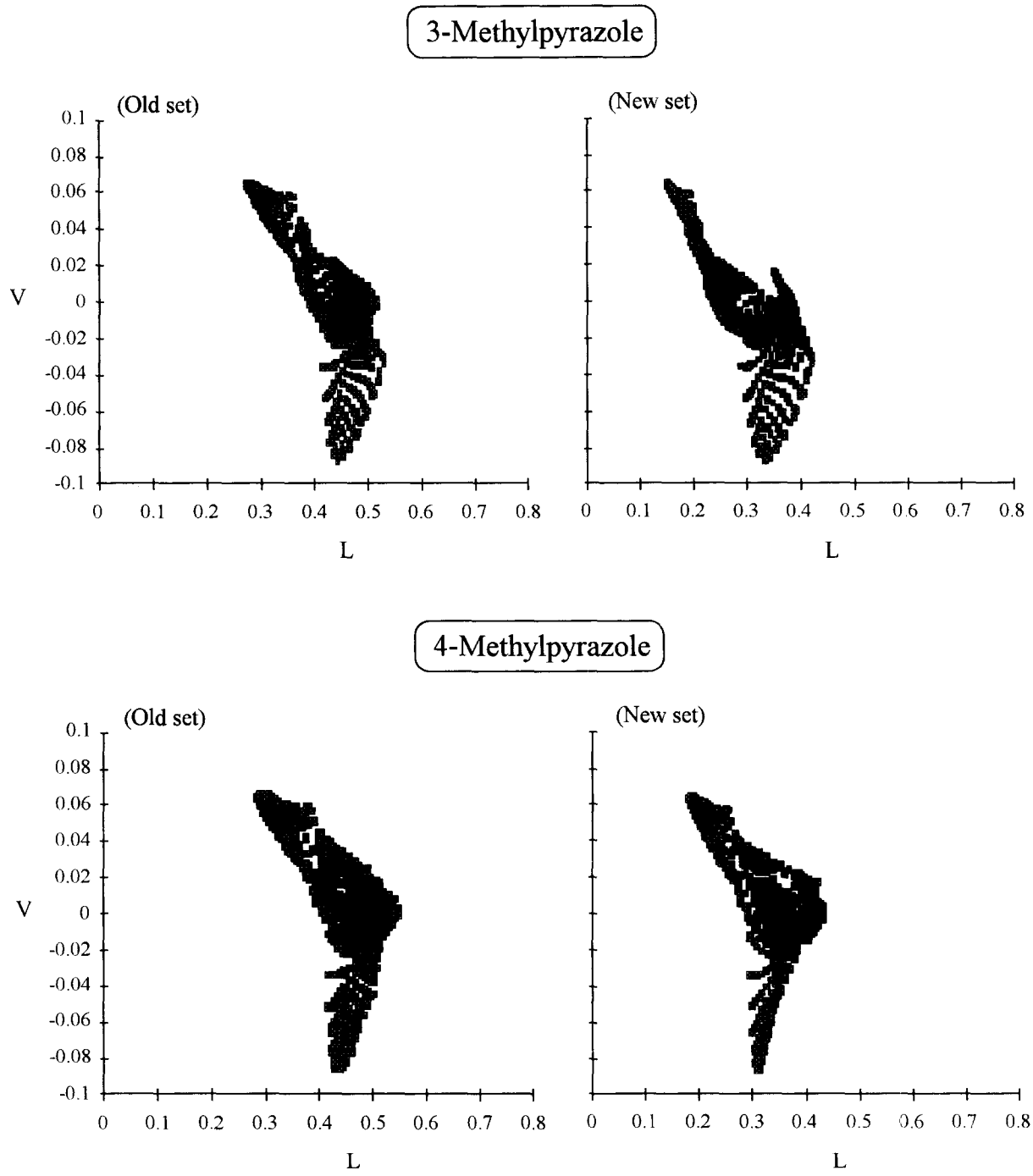


*Figure 7. Changes in (V, L) maps due to the position of a substituent group on the pyrazole ring. The MEP and MLP have been calculated as explained in Fig. 4 for other substituted pyrazoles. Note that even a small difference in compounds, such as a 3-methyl and 4-methylpyrazole, can be detected in the (V, L) maps, especially when using the set of lipophilicity parameters in Ref. 6. MEP values (V) are in atomic units.*

- The electrophilic N–H group in the pyrazole ring (high MEP values) is more lipophobic in the 3-methyl derivative.

These small differences distinguish the two molecules, although they may not be sufficient to rationalize their different inhibitory power. Nevertheless, by comparing the distributions of MEP and MLP on the molecular surfaces (cf. Figures 4 and 7), 4-methylpyrazole would appear to be more akin to 4-propylpyrazole than to 3-methylpyrazole. This similarity is in agreement with the experimental inhibitory tendencies,[27] and it illustrates the potential applications of an analysis of MEP–MLP maps.

Further insight into the properties of pyrazole derivatives can be derived from Figure 8. Figure 8 presents the partial (*V*, *L*) maps for the substituent groups in all four compounds, labelled (a) to (d). (Only the results with the set of lipophilicity parameters from Ref. 6 are shown). The regions enclosed in dashed lines define the ranges of MEP and MLP for the entire molecules. These diagrams give an absolute comparison between the side chain and the rest of the molecule, as well as a relative comparison between compounds. From the results in Figure 8, the following conclusions on molecular similarity are drawn.

- Alkyl substituents span a small region in a (*V*, *L*) map. This region is mostly lipophilic and involves intermediate values of MEP. The carboxylic chain, in contrast, spans almost the same range of *V* and *L* values as the entire molecule.
- The 4-methyl and 4-propyl substituents occupy a similar position with respect to the rest of the molecules. [That

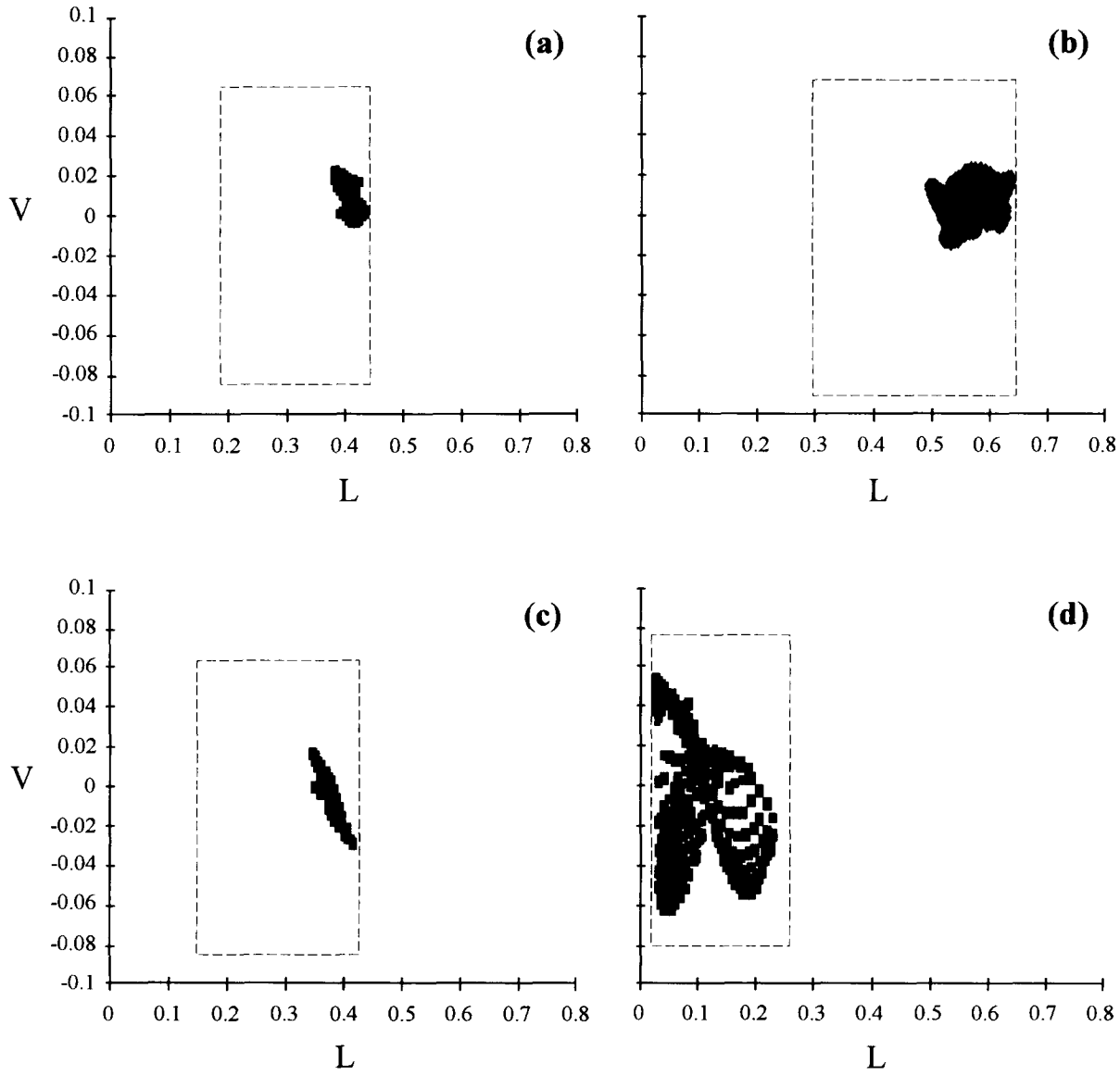


*Figure 8. Comparison of the partial (V, L) maps for the substituent groups of four substituted pyrazoles. The dashed-line boxes indicate the full span of V and L values for each molecule. (a) 4-Methylpyrazole; (b) 4-propylpyrazole; (c) 3-methylpyrazole; (d) 4-hydroxycarbonylpyrazole. Alkyl chains are lipophilic groups, yet they can be differentiated by their (V, L) maps. The MEP value has been computed at the ab initio level with a minimal basis set, using AM1-optimized geometries. The MEP values (V) are in atomic units. Lipophilicity potential is computed with the parameter set in Ref. 6.*

is, by shifting the diagrams, result (a) can be superimposed completely within result (b).] The 3-methyl chain, in contrast, appears less lipophilic and more nucleophilic than the latter two.

In summary, we observe that a proper comparison of these compounds requires more than a mere study of electrostatics, because the four molecules span almost the same range of values of MEP. The inclusion of MLP in the analysis permits a more detailed assessment of dissimilarities among the various compounds.

The fact that different regions on the molecular surface may appear superimposed on an MEP–MLP map entails, of course, a loss of information. However, the resulting picture conveys those features that are relevant for understanding the recognition of a small molecule by an active site. The MEP–MLP maps can be used as a first step in computer-assisted receptor modelling. For instance, the analysis of optimal steric similarity could be restricted only to those molecules most similar to a reference ligand in terms of overall electrostatics and hydrophobicity. The present approach can also be used for improved drug–receptor matching, by analyzing the complementarity of MEP–MLP maps of an active site and its agonists and antagonists.

## FURTHER COMMENTS AND CONCLUSIONS

In this work, we have computed electrostatics–lipophilicity maps and studied some of their properties. The approach is based on mapping two properties [$V(\mathbf{r})$ and $L(\mathbf{r})$] on a molecular surface and analyzing their interrelations. From another perspective, the resulting ($V$, $L$) maps can be viewed as a new representation of the 3D molecular surface as a 2D projection onto a property plane.

These maps characterize the molecular surface in a manner that is different from observing the molecular surface in three dimensions. Whereas the space occupied by each atom in the molecular surface is determined by its atomic radius, the space it occupies on a ($V$, $L$) map reflects its reactivity and solvent affinity. In this manner, small regions and crevices that are relevant to molecular recognition and binding may appear as large, easily recognizable regions on a ($V$, $L$) map.

We have presented results for derivatives of pyrazole and observed the role of various substituents on the maps. The differences found provide structural information that can be correlated with experimental activity measures. For these applications, it is important to compare maps quantitatively. Several alternatives for absolute and relative comparison are possible.

The shift between maps conveys the difference in absolute lipophilicity between two or more compounds. A simple measure of molecular similarity can be a "distance" between two such maps. For instance, consider two molecules A and B. The sets of points on their respective ($V$, $L$) maps will be indicated by $\{(V_i^{(A)}, L_i^{(A)})\}$ and $\{(V_i^{(B)}, L_i^{(B)})\}$. If the total number of points on the van der Waals surfaces is $N_A$ and $N_B$, for A and B, respectively, the geometrical center of a map is given by a pair of values $(V^{(k)}, L^{(k)})$:

$$V^{(k)} = \frac{1}{N_k}\sum_i V_i^{(k)},\ L^{(k)} = \frac{1}{N_k}\sum_i L_i^{(k)} \qquad (3)$$

where $k$ = A or B

Finally, a single number $d_{AB}$ measures the distance between two maps:

$$d_{AB} = [(V^{(A)} - V^{(B)})^2 + (L^{(A)} - L^{(B)})^2]^{1/2} \qquad (4)$$

For other comparisons, it may be more useful to analyze the relative difference between two maps instead of their absolute distance. To this purpose, we can build the difference diagram of two ($V$, $L$) maps. One possible approach would be to superimpose the geometric centers of the maps and retain only the parts where they do not overlap with each other. Another possibility would be to find out the translation along $V$ and $L$ axes that maximize the overlap between two diagrams and then define a similarity index as the one commonly used for molecular density functions.[30] These alternatives can easily be implemented and should facilitate the use of the present approach for the study of large series of molecules.

## ACKNOWLEDGMENTS

This work has been supported by an individual operating grant (to G.A.A.) from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

## REFERENCES

1 Tanford, C. *The Hydrophobic Effect*. Wiley, New York, 1973; Israelachvili, J. *Intermolecular and Surface Forces*. Academic Press, London, 1991
2 Mezey, P.G. In: *Reviews of Computational Chemistry*, Vol. 1 (Lipkowitz, K.B. and Boyd, D.B., Eds.). VCH, New York, 1990
3 Richards, F.M. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151
4 Langridge, R., Ferrin, T.E., Kuntz, I.D., and Connolly, M.L. *Science* 1981, **211,** 661
5 Richards, W.G. *Quantum Pharmacology*. Butterworths, London, 1983
6 Viswanadhan, V.N., Ghose, A.K., Revankar, G.R., and Robins, R.K. *J. Chem. Inf. Comput. Sci.* 1989, **29,** 163
7 Hansch, C. In: *Correlation Analysis in Chemistry* (Chapman, N.B. and Shorter, J., Eds.). Plenum, New York, 1978
8 Hansch C. and Leo, A.J. *Substituent Constants for Correlation Analysis in Chemistry and Biochemistry*. Wiley, New York, 1977
9 Rekker, R.F. and de Kort, H.M. *Eur. J. Med. Chem.* 1979, **14,** 479
10 Ghose, A.K. and Crippen, G.M. *J. Comput. Chem.* 1986, **7,** 565
11. Furet, P., Sele, A., and Cohen, N.C. *J. Mol. Graphics* 1988, **6,** 182
12 Fauchère, J.-L., Quarendon, P., and Kaetterer, L. *J. Mol. Graphics* 1988, **6,** 203
13 Audry, E., Dubost, J.-P., Colleter, J.-C., and Dallet, P. *Eur. J. Med. Chem.* 1989, **24,** 71
14 Audry, E., Dubost, J.-P., Dallet, P., Langlois, M.-H., and Colleter, J.-C. *Eur. J. Med. Chem.* 1989, **24,** 155
15 Croizet, F., Langlois, M.-H., Dubost, J.-P., Braquet,

P., Audry, E., Dallet, P., and Colleter, J.-C. *J. Mol. Graphics* 1990, **8,** 153
16 Iwase, K., Komatsu, K., Hirono, S., Nakagawa, S., and Moriguchi, I. *Chem. Pharm. Bull.* 1985, **33,** 2114
17 Akahane, K., Nagano, Y., and Umeyama, H., *Chem. Pharm. Bull.* 1989, **37,** 86
18 Kantola, A., Villar, H.O., and Loew, G.H. *J. Comput. Chem.* 1991, **12,** 681
19 Alkorta, I. and Villar, H.O. *Int. J. Quantum Chem.* 1992, **44,** 203
20 Connolly, M.L. *J. Am. Chem. Soc.* 1985, **107,** 1118
21 Binkley, J.S., Whiteside, R.A., Krishnan, R., Seeger, R., DeFrees, D.J., Schlegel, H.B., Topiol, S., Kahn, L.R., and Pople, J.A. *QCPE Bull.* 1981, **13**
22 Singh, U.C. and Kollman, P.A. *QCPE Bull.* 1982, **117,** prog. no. 446
23 Gavezzotti, A. *J. Am. Chem. Soc.* 1983, **105,** 5220
24 Rekker, R.F. *The Hydrophobic Fragmental Constant.* Elsevier, Amsterdam, 1977
25 Theorell, H., Yonetani, T., and Sjöberg, B. *Acta Chem. Scand.* 1969, **23,** 255
26 Fries, R.W., Bohlen, D.P., and Plapp, B.V. *J. Med. Chem.* 1979, **22,** 356
27 Rozas, I., Arteca, G.A., and Mezey, P.G. *Int. J. Quantum Chem., QBS* 1991, **18,** 269
28 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., and Stewart, J.J.P. *J. Am. Chem. Soc.* 1985, **107,** 3902
29 Alkorta, I., Villar, H.O., and Arteca, G.A. *J. Comput. Chem.* 1993, **14,** 530
30 Carbó, R., Leyda, L., and Arnau, M. *Int. J. Quantum Chem.* 1980, **17,** 1185