Theor Chim Acta (1991) 81: 79–93

Theoretica Chimica Acta

A measure of roughness of cross sections of molecular surfaces

Gustavo A. Arteca[1] and Paul G. Mezey[2]

[1,2] Department of Chemistry, [1]Department of Physics, and [2]Department of Mathematics, University of Saskatchewan, Saskatoon, Canada S7N 0W0

Received November 11, 1990; received in revised form January 27, 1991/Accepted January 30, 1991

Summary. It is possible to describe some aspects of the 3D shape features of molecules by characterizing a selection of 2D cross sections (planar curves) of a formal molecular surface. In this work we propose a new method to analyze the complexity of the shape of planar curves, in particular, to recognize and measure their roughness. Our approach provides a discrete "*roughness spectrum*" of the curve, which, in turn, can be characterized by other, simpler shape descriptors. The procedure is applied to cross sections of molecular electronic isodensity contours, which are classified and compared in terms of their roughness measures.

Key words: Molecular shape – Cross sections of molecular surfaces – Roughness measures – Shape descriptors – Roughness spectrum

1. Introduction

Some aspects of the three-dimensional shape of molecules can be described by appropriately chosen, formal molecular surfaces, such as electronic isodensity contours and molecular electrostatic isopotential contours. The characterization and comparison of such molecular surfaces is of importance in several areas of theoretical and applied chemistry, biochemistry, and pharmacology [1–4].

Molecular shapes are often represented in terms of *cross sections* of such molecular surfaces. In this case, the problem of studying of 3D shape is transformed into a 2D problem: the analysis of a number of planar curves, the cross sections. The problem is further simplified if these *continuous* curves are characterized by applying methods of *discrete* mathematics, which are often more suitable for computer applications. In this work we formalize a discrete characterization of molecular surface cross sections. The method uses some graph-theoretical notions already discussed in the literature [5], but goes further into the characterization of the roughness of a curve, as measured in terms of its oscillations or undulations. The procedure is easily programmable as an algorithm, permitting an automatic evaluation of the shape descriptors for a large number of cross sections and molecules. This feature is of importance, for example, in computer-assisted drug design where studies of molecular similarity are needed.

Graph theory has been used to characterize molecular surfaces [6–8] and their cross sections [5]. Previous works were based on visibility properties [5, 8], i.e., on the condition of special points of the curve (vertices) seeing each other [5], or seeing sections of the closed curve from points interior or exterior to it [8]. In this work we use distance properties between vertices of the planar curve representing the cross section. These distances play a key role in constructing a roughness measure for planar curves. Our approach allows one to quantify absolute roughness and provides an algorithm to compare cross sections of model molecular surfaces. This quantitative characterization is of potential use in the study of correlations between chemical structure and properties such as sweetness or odour, which can be related to the roughness of a molecular surface [1–3].

This work has been organized as follows. In Sect. 2 we define a shape descriptor of planar curves. The information is presented in terms of a formal spectrum: the *roughness spectrum of a planar curve*. Properties of this spectrum and some illustrative computations are given in Sect. 3. Section 4 presents a one-dimensional roughness measure of a curve, derived from its roughness spectrum. This measure involves a further reduction of the original information, and it provides an easily computable numerical value to assess degrees of similarity between related compounds. This similarity measure may serve as an alternative to earlier choices for measures and degrees of shape similarity [9–12]. Section 5 provides an illustration of the concepts and the technique for several cross sections of the electron density function of planar molecules. The molecular surfaces of water, ethene, formaldehyde, and formic acid are considered as examples. In Sect. 6 we discuss the extension of the method to a number of anomalous curves, e.g., sectionally-generated cross sections (i.e., cross sections composed from pieces) and nondifferentiable curves. As well, we discuss the generalization of this method to 3D surfaces. Conclusions are found in Sect. 7.

2. Roughness spectrum of a planar curve

Let C be a non-self-intersecting planar curve, defined by a parametric function $\phi(t): I \to {}^2\mathbb{R}$, where I is the unit interval, $I = [0, 1]$. As for now, we will require C to be continuous, twice differentiable, bounded, and of finite arc length, with no self-crossings. Let $\{V_i\}$ be the set of "*vertices*" of the curve. This will be formed by the initial and final points $V_0 = \phi(0)$ and $V_{N+1} = \phi(1)$ of the curve (for open curves), and all its N local inflexion points, as discussed in [5]:

$$V_i \in C, \quad i = 1, 2, \ldots, N; \qquad V_i = \phi(t_i), \tag{1}$$

where $\{t_i\}$ is the corresponding set of parameter values, with $t_i \in (0, 1)$. The local system of coordinates is defined by the straight line tangent to the planar curve at any given point, and by a direction perpendicular to it, lying on the same plane. By regarding the coordinate along the tangent as the variable and the curve as a locally defined function along the other coordinate, all points of C are critical points in their corresponding local systems. However, only some of them will be local inflexion points in the same coordinate systems. Notice that a curve everywhere locally convex will have $N = 0$; if it is open it will possess two vertices and one if it is closed, since the formally designated initial and final points coincide.

The index of ordering for the vertices is defined by their occurrence along the oriented curve. If the curve is open, the initial and final points of the curve, $\phi(0)$

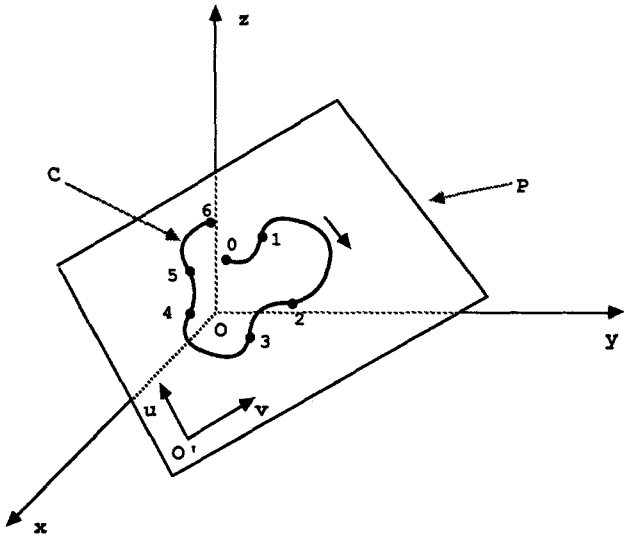


Fig. 1. Laboratory frame and planar coordinate systems for the characterization of planar curves. (The numbers 1, 2, 3, 4, and 5 on the oriented curve C identify inflexion points. Number 0 corresponds to the initial point of the curve and 6 to the final point)

and $\phi(1)$, respectively, are considered as formal vertices. We shall discuss below the criterion to choose the initial point $\phi(0)$ of a closed curve. Let $P(A_1, A_2, A_3, A_4)$ be the plane in 3-space where the curve C is embedded:

$$P(A_1, A_2, A_3, A_4) = \{\boldsymbol{r} = (x, y, z) \in {}^3R : A_1x + A_2y + A_3z = A_4\}. \quad (2)$$

Let v and u be the two Cartesian variables of a system of coordinates spanning the plane P, with an arbitrary origin within the plane. The relation between the various coordinate systems and the curve C is illustrated in Fig. 1. The equation for C can now be written as $u = h(v)$, with h a multi-valued function related to $\phi(t)$. See also the discussion in [5].

Our aim is to measure quantitatively the "roughness" of the curve C. In our interpretation, the roughness will be regarded as a manifestation of the undulations of the curve. In all cases, the character of a differentiable curve as being rough or rugged will be given in terms of its curvature properties. The design of a roughness measure appropriate to our needs is based on meeting the following intuitive notions: A curve will be rougher if it has more undulations, and for an equal number of undulations, if they are densely packed and if their amplitudes are greater. Recall that all the curves we deal with are smooth, in the sense that they are at least twice differentiable. As a consequence, a curve with fewer or more damped oscillations will be called "less rough", and the term "not smooth" will be used for curves which are not twice differentiable.

The vertices of the curve define the limits of an undulation, which resembles a half-wave [5]. In the case of sine waves, two consecutive vertices enclose exactly a half wave.

In order to characterize these undulations, two parameters appear naturally: the distance between two consecutive inflexion points $\boldsymbol{V}_i$ and $\boldsymbol{V}_{i+1}$, and the arc length of the section of the curve for which they provide end points. We shall take the distance between two consecutive inflexion points (or vertices) as:

$$L_{i,i+1} = \|\boldsymbol{V}_i - \boldsymbol{V}_{i+1}\|, \quad i = 0, 1, 2, \ldots, N, \quad (3)$$

and the arc length s of the section of the curve between the same two inflexion points as:

$$s_{i,i+1} = \int_{v(t_i)}^{v(t_{i+1})} [1 + (dh/dv)^2]^{1/2}\, dv. \tag{4}$$

Here $v(t_i)$ and $v(t_{i+1})$ are the values of the variable v on plane P for the consecutive inflexion points. If $(v(0), h(v(0)))$ is the initial point of the curve in the system of coordinates for the plane P, then $s_{0,1}$ denotes the arc length of the curve from the initial point to the first inflexion point. Analogously, $s_{N,N+1}$ will be the arc length measured from the last inflexion point to the end of the curve $(v(1), h(v(1)))$. The total length S of the curve is then $S = s_{0,N+1}$. Notice that the curve is oriented from $t = 0$ to $t = 1$; if the curve is closed, the first and last vertices coincide, $V_0 = V_{N+1}$. Conventionally, the orientation of a closed curve will be chosen as clockwise, as viewed from a point of the positive w axis in a right-handed (v, u, w) local coordinate system.

There is a degree of arbitrariness in the choice of the starting point $(v(0), h(v(0)))$ for closed curves C. A convenient choice is one that makes the location of $(v(0), h(v(0)))$ independent from the placement or orientation of the curve in 3-space. A criterion that satisfies this condition would be to take as starting point an inflexion point with extremal properties, for example, having the largest value for the local third derivative. For the sake of simplicity, in this work we have resorted to a simpler choice. The starting point will be the inflexion point of smallest v value; if there are more than one such inflexion points, then we shall choose the one with the smallest u value. This criterion provides a starting point which is not invariant if the curve is rotated. However, we shall see that many important features of the shape characterization will not depend on this choice.

Our goal is to characterize the curve by a discrete spectrum, that is, by a finite sequence of lines, whose intensity (length) and position will characterize the shape features of the curve. Since the main property we seek to characterize is the roughness of the curve, the final description will be called *a roughness spectrum.*

The spectrum is completely specified by expressing the relation between the original curve C and the position and intensity of the lines.

a. Line position. A line will indicate the occurrence of a vertex. The position of the i-th spectral line will be given by the arc length of the curve from the starting point of the i-th vertex, measured as a fraction of the total arc length of S. The position p_i of the i-th line, $i \geqslant 0$, is then given as:

$$p_i = \frac{1}{S} \sum_{k=0}^{i} s_{k,k+1}. \tag{5}$$

Notice that the spectrum has a peak at the end of the interval, that is, for $p = 1$, since the final point V_{N+1} is also considered a vertex.

b. Line intensity. Our evaluation of the line intensity takes into account two competing factors: the accumulation of inflexion points and the amplitude of the oscillations. A curve may exhibit a rough section if a large number of vertices accumulates in the section, but it would not be rough if the oscillations were small. In order to describe these features, one must evaluate the intensities,

dependent on parameters determined by at least two inflexion points. We take into account these two factors by expressing the intensity of the i-th line as:

$$I_i = \lambda_i \omega_i, \tag{6}$$

where λ_i is a factor describing the proximity of two vertices, and ω_i is a weight factor for the amplitude. We shall require the following properties to be satisfied:

(i) $|\lambda_i| \leqslant 1$; $\lambda_i \to 0$ if the $(i-1)$-th vertex is infinitely far from the i-th vertex, and $|\lambda_i| \to 1$ if the two vertices are infinitesimally close.

(ii) $I_i \to 0$, if $\lambda_i \to 0$.

(iii) $\omega_i \geqslant 1$; the weight ω_i will be a minimum ($\omega_i = 1$) when the section of the curve joining the V_i and V_{i-1} vertices becomes a straight line segment, that is, when $s_{i-1,i} = L_{i-1,i}$. For more curved segments, ω_i is a greater number.

There is an infinite number of functions that satisfy the above conditions. We shall use a very simple representation. To this end, it is only necessary to introduce one more parameter, measuring the overall size of the curve in space.

Note that a curve can have a very large total length S, but actually occupy a very small area on the plane. Let us introduce a parameter L_M, defined as:

$$L_M = \max\{|v(t) - v(t')|, |h(v(t)) - h(v(t'))|\}, \quad t, t' \in I. \tag{7}$$

Note that L_M is the maximum span of v or u coordinates of the curve in the $w = 0$ plane (i.e., L_M is *not* a distance between inflexion points). As the curve C is then enclosed in a square of sidelength L_M, this parameter provides the bound $2^{1/2}L_M$ for the diameter of the curve. This parameter has a similar orientation dependence as the choice of a starting vertex of loops and it is the yardstick against which we shall compare the distances between vertices. A simple choice of functions λ_i and ω_i is:

$$\lambda_i = (L_M - L_{i-1,i})/L_M, \tag{8.1}$$

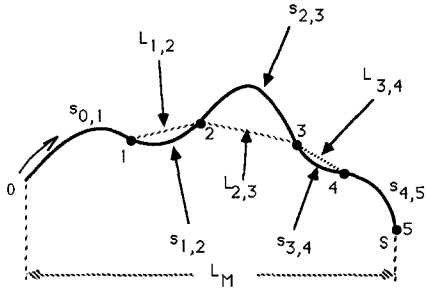
$$\omega_i = s_{i-1,i}/L_{i-1,i}. \tag{8.2}$$

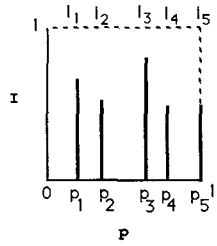
The set of values $\{(p_i, I_i)\}$ obtained with the above choices specifies a roughness spectrum of the curve. Notice that positions and intensities of lines are dimensionless. This feature allows one to address separately the shape (roughness) and size aspects of the description of a molecular surface.

Note that, since from Eqs. (8) one has $i > 0$, we have chosen not to define the intensity I for the position p_0. Accordingly, a closed curve with no inflexion points will generate no spectral lines. On the other hand, an open curve with no inflexion points will generate one spectral line, resulting from the initial and final points of the curve. Accordingly, closed and open convex curves can be distinguished. Note that, if one reverses the orientation of the curve, the line intensities are not affected, and the line positions simply change from p_i to $1 - p_i$. In this sense, the essential features of the spectrum are not affected, but one can nevertheless distinguish two curves differing only in their orientation.

Figure 2 gives an example of an arbitrary open curve, where all the parameters discussed above are indicated. The drawing at the bottom represents qualitatively the corresponding spectrum.

The intensity I_i, $i > 0$, is not bounded. Therefore, large oscillations in sections where the curve has many inflexion points may lead to very large intensities. Curves exhibiting these features are classified as very rough. Figure 3





$$I_i = (L_M - L_{i-1,i}) \frac{s_{i-1,i}}{L_M L_{i-1,i}}$$

$$P_i = \frac{\sum_{k=0}^{i} s_{k,k+1}}{S}$$

Fig. 2. Curve parameters necessary to define the lines and intensities of the roughness spectrum. (*Upper* diagram shows an open curve *C*; *lower* diagram corresponds to a schematic representation of the spectrum)

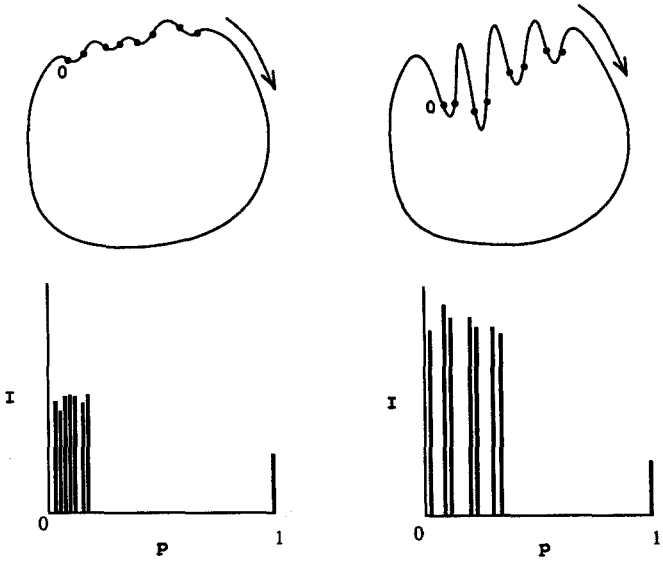


Fig. 3. Comparison of roughness spectra for closed curves with the same number of inflexion points, but different undulation patterns

provides an example of how the roughness spectrum characterizes two curves with different types of undulations. Both curves have the same number of spectral lines. However, the curve on the left-hand side is less undulating, and hence has peaks with smaller intensity. Although they have comparable spacings, the different intensities of lines indicate that the curve to the left is less rough. In another section, we discuss a quantitative measure of the difference between two spectra.

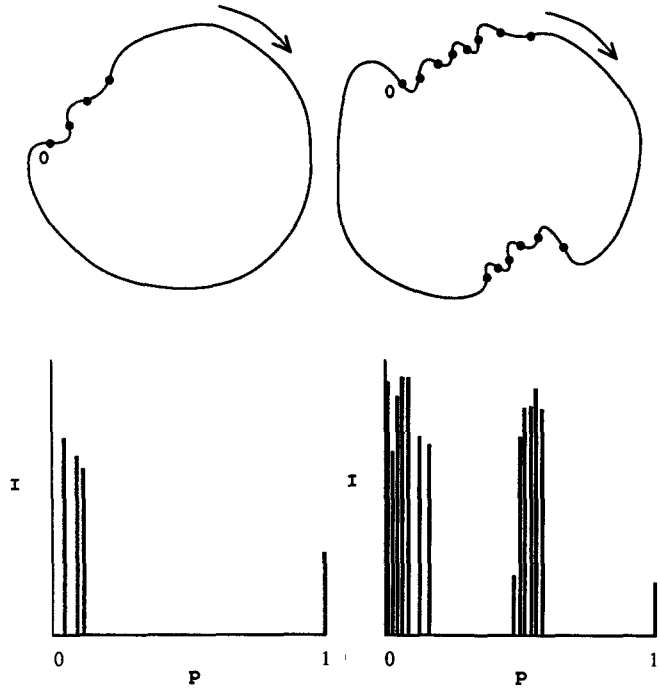


Fig. 4. Comparison of roughness spectra for closed curves with different number and location of inflexion points

Figure 4 compares two other closed curves. The curve on the right-hand side is rougher, but the roughness pattern varies along sections of the curve. The spectrum makes this fact evident since the peaks cluster forming two "bands", indicating the presence of two separated rough sections of the curve C.

We are now in the position to classify and compare planar curves in terms of their roughness spectra, completely specified by a set of ordered pairs $\{(p_i, I_i)\}$, with $i = 1, 2, \ldots, N+1$. In what follows we discuss some properties of the roughness spectra and provide illustrative examples.

3. Properties of the roughness spectra and some simple examples

Regarding the comparison of curves, we shall say that two curves C_1 and C_2 have identical roughness spectra if they possess the same set $\{(p_i, I_i)\}$. There are a number of properties that follow from the definition of the roughness spectrum. One important property is the following: the roughness spectra remain invariant when deforming the curve by multiplying simultaneously the u and v coordinates by the same scaling factor (double-scaling operation or "blow-up" of the curve). This feature guarantees that our descriptor measures *shape*, and is independent of *size*.

The effect of a one-variable scaling on the roughness of the curve varies, depending on whether the scaling produces an amplitude or frequency change in the undulations, or a contribution of both. For example, the following property is found: the lines in the spectra shift with damping since the relative locations of inflexion points change. Figure 5 illustrates this behavior. The example shows three curves (C_1, C_2, and C_3) with four inflexion points; the distances between

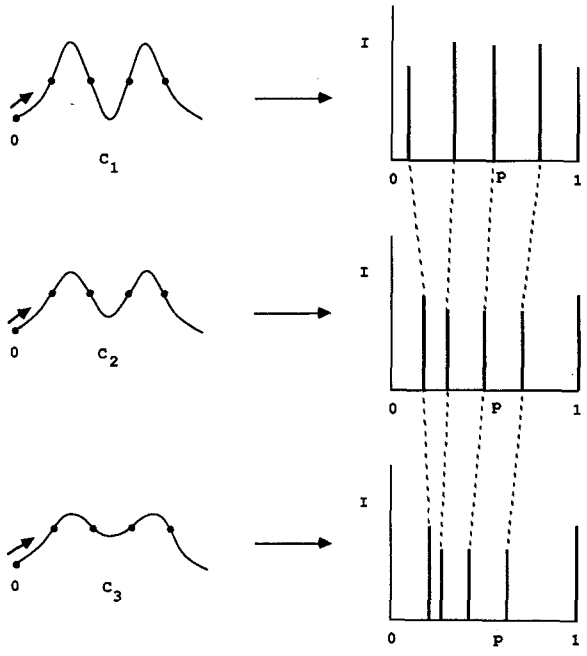


Fig. 5. Shift in spectral lines due to damping in the oscillations of planar curves with the same relative location of inflexion points

inflexion points are the same in the three cases. For simplicity, we have assumed that all spectral lines have the same intensity. However, the curve's oscillation between inflexion points becomes damped as one passes from C_1 to C_3. The effect on the spectrum is to shift the second peak (the first depending on an actual inflexion point) to a larger value of the position variable, and all other ones to smaller values of the position variable. The result is that the lines of the spectrum cluster in a more compact region. Moreover, the intensity of the lines decreases. A more compact spectrum with lower intensity lines describes a less rough curve.

In order to illustrate these properties we have considered some analytical examples.

a. Sine waves. The curve is given by $u = \sin kv$, with k a constant, and v defined over the interval $[0, 2\pi]$. In this case we have $L_M = 2\pi$. If k is an integer number, then the curve will have $2k + 1$ vertices (inflexion points). Notice that in this particular case the initial and final points are also local inflexion points. The following elementary results hold ($i \geqslant 1$):

$$L_{i,i+1} = \pi/k,$$
$$S = 4k\sigma,$$
$$p_i = i/2k,$$
$$I_i = (2k - 1)\sigma/\pi,$$

where σ is the following constant:

$$\sigma = \frac{1}{k} \int_0^{\pi/2} (1 + k^2 \cos^2 x)^{1/2}\, dx.$$

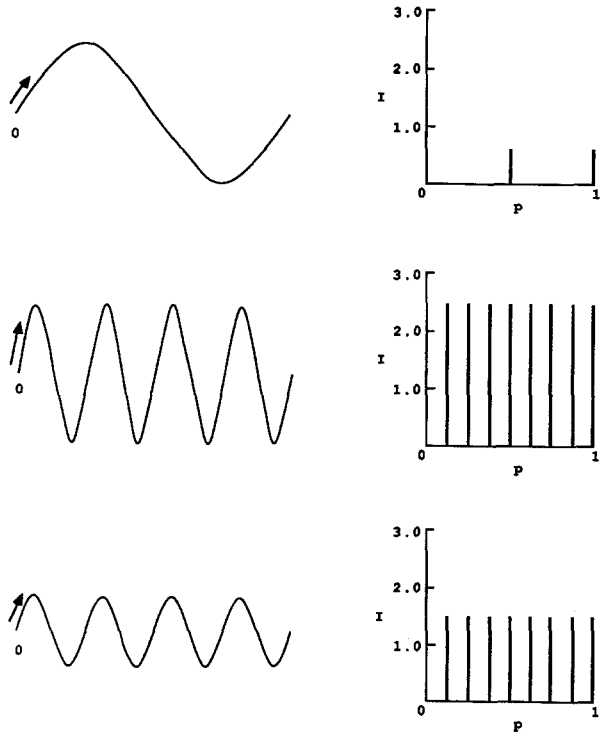


Fig. 6. Effect of frequency and amplitude changes on the roughness spectra of sine waves. (*Upper*, *middle*, and *bottom* diagrams correspond to the functions $u = \sin v$, $u = \sin 4v$, and $u = (\sin 4v)/2$, respectively)

These results show that the spectrum has more lines (and with larger relative intensity) when the frequency k increases. Figure 6 (top and middle diagrams) illustrates this. A smaller frequency makes the curve less rough. On the other hand, a change in amplitude can have a similar effect. The bottom diagram in Fig. 6 shows the result of decreasing the amplitude by a half, for the function $u = \sin 4v$ (middle diagram). The line positions are not affected, but the intensity of the peaks decreases by about 40%. The curve becomes less rough if its amplitude decreases, while leaving the positions of the inflexion points unchanged.

b. Round-edge cross. This example is a closed curve having alternating concave and convex sections. Each section can be represented by an arc of a reference circle. The curve is depicted in Fig. 7. Notice the location of the starting point of the curve chosen according to our convention. If R is the radius of the reference circle, then the following relations hold:

$$L_M = 6R,$$

$$L_{i,i+1} = 2R, \qquad L_{i+1,i+2} = 2^{1/2}R,$$

$$p_{2j+1} = (3j+2)/12, \qquad p_{2j+2} = (j+1)/4,$$

$$I_{2j+1} = \pi/3, \qquad I_{2j+2} = (2^{1/2} - 1/3)\pi/4,$$

where $i \geqslant 1$ and $j \geqslant 0$. The bottom diagram in Fig. 7 shows the resulting spectrum. The distribution of lines clearly reveals the existence of periodicity in

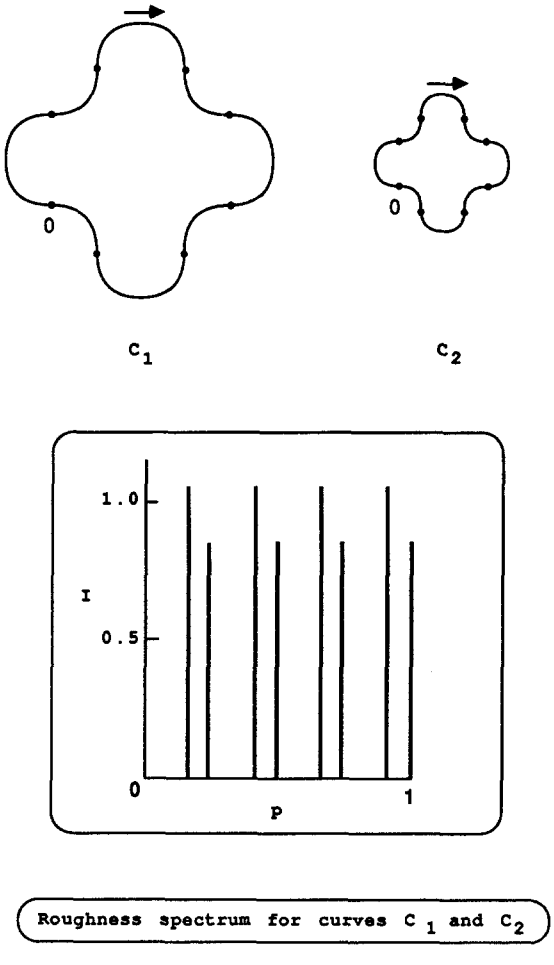


Fig. 7. Invariance of the roughness spectrum by double scaling. [Notice the periodicity of the spectral lines, as a result of the symmetry in the round-edge cross]

the original curve. The curve C_2 on the right-hand side of the diagram has been obtained from C_1 by uniform scaling (contraction). Therefore, the two curves have exactly the same spectrum (property mentioned before).

We discuss below more chemically relevant examples, namely, the characterization of curves representing cross sections of the electron density function of molecules.

4. A simple roughness measure derived from the roughness spectrum

Although the roughness spectrum conveys the degree of undulation (roughness) of the planar curve in a concise pictorial way, occasionally it is necessary to have a simpler, one-dimensional function that describes some of the features of the spectrum. We propose one such function here.

Let us choose a real function δ_R, which will be completely determined from the roughness spectrum. This function will be called a "roughness measure". In a

discrete spectrum this function δ_R will play the role similar to a spectrum integral of a continuous function $I(p)$.

We propose the following definition:

$$\delta_R = \sum_{i=0}^{N} w_i \langle I_i \rangle \Delta p_{i,i+1}, \tag{9}$$

where $\langle I_i \rangle = (I_i + I_{i+1})/2$ is the average intensity value for two consecutive peaks, $\Delta p_{i,i+1}$ is the difference in position of these two peaks, and w_i is a weight function. This weight function is introduced in order to reduce the contribution of two peaks with large separation. The only requirement for w_i is: $w_i \to 1$ if $\Delta p_{i,i+1} \to 0$, with $w_i \leqslant 1$. Notice that, with the above definitions, if the spectrum becomes continuous (that is, if there are infinitely many lines with infinitesimal separation $\Delta p_{i,i+1} \to 0$) then δ_R reduces to the integral of the function $I(p)$, for $0 \leqslant p \leqslant 1$.

There are infinitely many functions w_i that satisfy the above condition. We have sought a function that falls rapidly within the interval [0, 1], so that only close neighboring lines will give a large contribution to δ_R. We have tried a number of possibilities; the following appears to be an appropriate choice:

$$w_i = \exp(-\Delta p_{i,i+1}/(1 - \Delta p_{i,i+1})). \tag{10}$$

In the next section we discuss some constant electron density cross sections (closed planar curves), and we compare them in terms of their roughness spectra and roughness measures δ_R.

5. Characterization of constant electron density contour lines

We have considered a number of planar molecules and computed the cross section of the total electron density ϱ in the molecular plane. The molecules are studied in their ground states and at their equilibrium geometries. The electron density has been computed at an *ab initio* level using a 3-21*G* basis set [13]. For the sake of simplicity, we present the results for isodensity contours with a single value of density.

We have considered four molecules: water, ethene, formaldehyde, and formic acid. The electron density value chosen is 0.0035 a.u. This value defines a density contour surface which is a reasonable, "realistic" molecular surface [14]. At this density value the cross sections are single, closed curves. For larger values of density the molecular cross section may be formed by a number of disjoint pieces. In this case we have to modify the approach to build the spectrum, since one has to proceed from curve to curve when moving along the interval for p. This extension is discussed later.

The four contour lines are shown in Fig. 8. The characterization in terms of their roughness spectra appears in Fig. 9. Observe that the spectrum of ethene is the one with more peaks, but the spectrum of formic acid exhibits the largest intensities. It is here where a roughness measure can provide a simple classification. For these molecules we have obtained the following results: $\delta_R = 0.53$, 0.45, 0.40, and 3.9×10^{-6}, for formic acid, ethene, formaldehyde, and water, respectively. The very small value of δ_R for water describes clearly the almost convex character of the planar cross section at $\varrho = 0.0035$ a.u. Notice that formic acid is the system classified as roughest due to a marked change of curvature of the electron density contour.

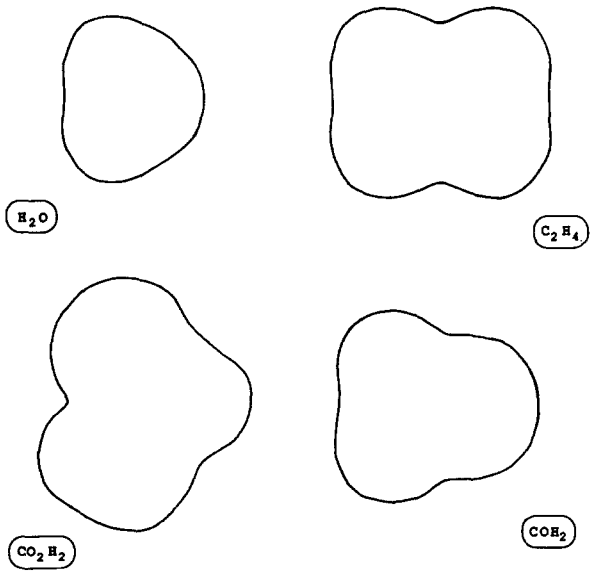


Fig. 8. Cross sections of the constant electron density contours at the molecular plane of several molecules. [The density value is 0.0035 a.u.; the electron density has been computed at RHF/3-21G *ab initio* level, for the equilibrium geometries. The molecules are water, ethene, formic acid, and formaldehyde]

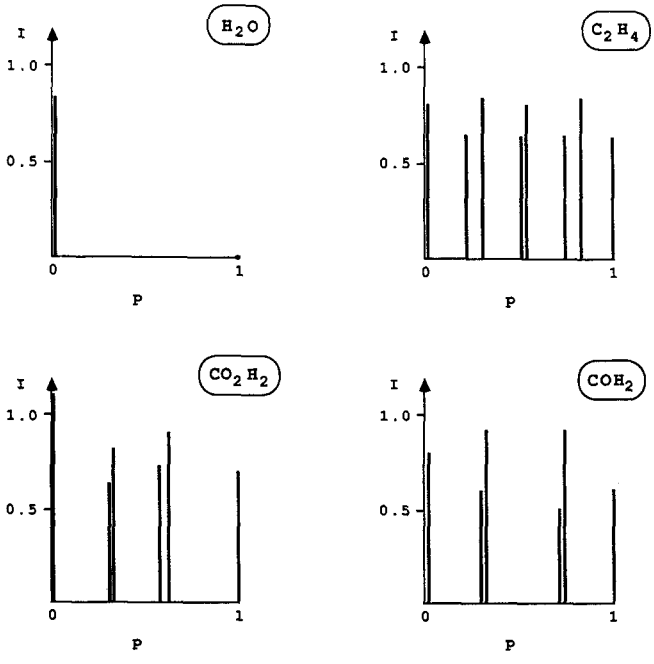


Fig. 9. Roughness spectra of the cross sections of the molecular electron density contours shown in Fig. 8

Our procedure provides a simple description of some of the curvature features that affect the roughness of curves. The construction of the spectrum and its characterization can be performed automatically, and the comparison of curves becomes straightforward. This approach may be potentially useful as a

tool, integrated with other techniques in studies requiring the assessment of molecular shape similarity.

6. Extensions of the method

We comment briefly in this section on the generalization of the procedure to a number of situations we have so far excluded.

Let us consider the case of a cross section C of a surface, formed by a number of disjoint pieces (maximum connected components of C). Suppose C is made of M disjoint sets C_i, $1 \leqslant i \leqslant M$, each of them, in turn, a closed planar curve. How does one pass from one curve to another? A simple solution consists of ordering them according to some criterion and defining a starting point for each C_i. We propose the following procedure, in the spirit of the choice of starting points made before:

(i) Starting point for the i-th curve: If the curve has local inflexion points, the starting point of C_i will be the inflexion point with the smallest v value (and, if there are other inflexion points with the same v value, then the point with the smallest u value will be taken). If the curve has no local inflexion points (a convex curve), then the starting point will be (min v, min h(min v)).

(ii) Ordering of curves C_i: The index labelling of the curves C_i is chosen in increasing order, $i = 1, 2, \ldots, M$, according to the values of min v. If two curves have the same min v, then the lowest index i is assigned to the connected component showing the smallest h(min v) value.

With these definitions, the spectrum is constructed by joining the spectra for each piece C_i. The overall length S is taken as: $S = S(C_1) + S(C_2) + \cdots + S(C_M)$, for the calculation of the line positions; however, the calculation of intensities is performed independently for each curve at a time.

Figure 10 illustrates how this is accomplished. The figure shows a cross section formed by four disjoint closed curves. To construct the roughness spectrum one passes successively from C_1 to C_2, C_2 to C_3, and so forth. The bottom diagram in the figure represents the roughness spectrum, indicating the contributions from each connected component C_i. Notice that C_3, being a convex curve, does not contribute with any line to the spectrum.

The approaches just discussed are all based on the fact that the curves are everywhere twice differentiable. With this condition it is possible to define vertices on the curve as inflexion points. However, these ideas can be extended to other types of curves. For example, if the curve is everywhere differentiable, except at a finite number of points, these points could be taken as the vertices of the curve. This approach can be useful to describe cross sections of fused-sphere surfaces, e.g., van der Waals surfaces. These molecular surfaces have cross sections which are locally convex, except for a finite number of points where the curves are not differentiable (points belonging to two or more atomic spheres).

Another definition of vertices of a planar curve can be of interest for some other applications. For example, the definition of $\{V_i\}$ as the set of points where the local second derivative is an extremum (instead of zero local second derivative) can be useful. This approach would differentiate among convex curves, which otherwise will always be 'unrough', due to their lack of local inflexion points.

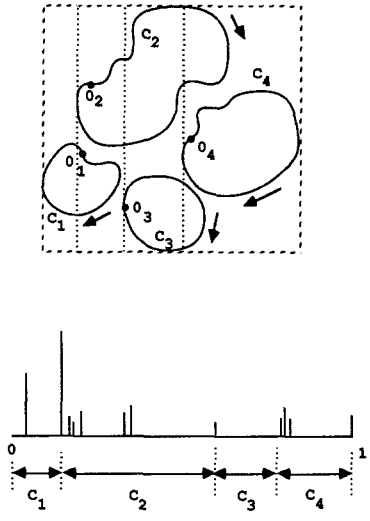


Fig. 10. Roughness spectrum of a cross section formed by a number of disjoint closed curves. (The starting point for each curve is the inflexion point with smallest v and smallest u values. If no inflexion point is found, as for the curve C_3, the origin is given as (min v, min h(min v)). The *dashed-line rectangle* identifies the domain in the u, v-plane where the cross section C is enclosed. The *dotted vertical lines* indicate the smallest v values for every curve, which gives the ordering to label the curves as C_1, C_2, C_3, and C_4. The curves contribute to the spectrum according to this ordering. Notice that C_3 does not contribute with any line)

A further aspect can be explored. Our roughness measure characterizes a cross section, which is a 1D object. The description of roughness in a 2D surface is a more complicated task, but it nevertheless can be accomplished by an extension of our approach. To this end, we can consider the infinitely-many cross sections of a surface obtained by slicing it with parallel planes. The choice of the normal vector of these planes introduces a direction dependence. Let dh be the distance between two of such planes, and $h = 0$ and $h = H$ two values of a height variable specifying the location of two parallel planes between which the actual surface is found. For each cross section one gets a roughness spectrum $\{(p_i, I_i)\}$ and a direction-dependent roughness measure $\delta_R(h)$, deduced from it. In principle, a characteristic function describing some roughness features of the surface can be calculated as:

$$\delta = \int_0^H \delta_R(h)\, dh. \tag{11}$$

Equation (11) can be evaluated numerically. Examples of these and similar generalizations will be discussed elsewhere.

7. Conclusions

We have proposed a mathematical approach to define a degree of 'roughness' in planar, differentiable surfaces. The procedure uses the changes in curvature along the cross section to derive a discrete spectrum which characterizes in a concise manner some of the shape features related to roughness. Moreover, some essential information contained in the roughness spectrum can be represented by a simple one-dimensional function (the roughness measure δ_R). This function provides a criterion to classify cross sections according to roughness in terms of a single number. Furthermore, the integration of δ_R over all parallel cross sections of a molecular surface may serve as a direction-dependent criterion to assess roughness for 3D surfaces.

It must be mentioned that other approaches to describe the roughness of curves can be followed. For example, a measure can be defined as the one minimized by a spline curve. This measure could be seen, approximately, as the average over the whole length of C of the root square curvature at each point of C. This measure could also easily be made scale invariant.

These quantitative measures can be useful for testing some theories proposing a correlation between "molecular roughness" and macroscopic properties, such as sweetness or odour.

Acknowledgments. This work has been supported by both strategic and operating research grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada. We would like to acknowledge a stimulating discussion with Prof. A. Laforgue (Reims) on this subject. We also would like to thank a referee for calling our attention to alternative roughness measures based on spline curves.

References

1. Richards WG (1983) Quantum pharmacology. Butterworths, London
2. Dearden JC (1983) Quantitative approaches to drug design. Pharmacochemical Library, vol 6. Elsevier, Amsterdam
3. Franke R (1984) Theoretical drug design methods. Pharmacochemical Library, vol 7. Elsevier, Amsterdam
4. Mezey PG (1990) Molecular surfaces, in: Lipkowitz KB, Boyd DB (eds) Reviews of computational chemistry, chap 7. VCH Publ, New York
5. Arteca GA, Mezey PG (1989) Theor Chim Acta 75:333
6. Mezey PG (1988) J Math Chem 2:299
7. Arteca GA, Mezey PG (1988) Int J Quantum Chem 34:517
8. Harary F, Mezey PG (1988) J Math Chem 2:377
9. Mezey PG (1991) J Math Chem 7, in press
10. Harary F, Mezey PG (1990) Chiral and achiral square-cell configurations and the degree of chirality, in: Mezey PG (ed) New developments in molecular chirality. Kluwer, Dordrecht
11. Mezey PG (1990) A global approach to molecular chirality, in: Mezey PG (ed) New developments in molecular chirality. Kluwer, Dordrecht
12. Arteca GA, Mezey PG (1989) Proc Conf IEEE Eng Med Biol Soc 11:1907
13. Frisch MJ, Head-Gordon M, Schlegel HB, Raghavachari K, Binkley JS, González C, Defrees DJ, Fox DJ, Whiteside RA, Seeger R, Melius CF, Baker J, Martin RL, Kahn LR, Stewart JJP, Fluder EM, Topiol S, Pople JA (1986) Gaussian 88. Gaussian Inc, Pittsburgh
14. Arteca GA, Grant ND, Mezey PG J Comp Chem (to be published)