

## SHAPE GROUP THEORY OF VAN DER WAALS SURFACES

Gustavo A. ARTECA\* and Paul G. MEZEY\*,<sup>‡</sup>

\* *Department of Chemistry, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0*

<sup>‡</sup> *Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0*

Received 22 October 1987  
(in final form 29 February 1988)

### Abstract

In this article we present a method for the study of shapes of general, asymmetric van der Waals surfaces. The procedure is simple to apply and it consists of two steps. First, the surface is decomposed into spherical domains, according to the interpenetration of the van der Waals atomic spheres. Each domain defines a topological object that is either a 2-manifold or some truncated 2-manifold. Second, we compute the homology groups for all the objects into which the surface is divided. These groups are topological and homotopical invariants of the domains, hence they remain invariant to conformational changes that preserve the essential features of these domains of decomposition. In particular, these homology groups do not depend explicitly on the molecular symmetry. Major rearrangements of the nuclear configurations, however, do alter the decomposition into spherical domains, and the corresponding variation of the homology groups can be followed easily under conformational rearrangements. We discuss a partitioning of the metric internal configuration space  $M$  into shape regions of van der Waals surfaces, which allows one to identify those rearrangements which introduce an essential change in shape and to distinguish them from those which do not alter the fundamental shape of the molecular surface. The dependence of the shape group partitioning of  $M$  on the symmetry under permutation of nuclear changes is discussed briefly, considering a simple illustrative example.

### 1. Introduction

The van der Waals surface (VDWS) is a simple and useful tool for the description of the volume and size effects in molecular interactions [1–3]. VDWSs are also important tools for the characterization of molecular shapes. The simplicity of their computation has made the VDWSs a test of first choice to study molecular similarity and correlations with other properties (see, for instance, refs. [4–8] on methods for

its computation). In certain fields, as in theoretical biochemistry and pharmacology, the VDWSs may constitute the main model used to represent the molecule (as, for example, in refs. [1,9,10]). In computer-aided drug design, an initial screening of molecules that may act as potential drugs can be made by comparing the shapes of their molecular surfaces. Due to the fact that these surfaces are most often asymmetrical objects, the use of point group theory provides no advantage. Therefore, a different approach is needed to reveal the similarity characteristics between different surfaces.

In certain cases, it is important to know whether a given conformational rearrangement leads to significant changes in shape or not. Even though some changes in configurational space may modify some features of the VDWS, it is clear that not every rearrangement of the nuclei represents an essential alteration of shape features. There is a need for a precise definition of the shape characteristics of the three-dimensional "body" of the molecule, as distinguished from the nuclear configuration and from a formal bond skeleton. This would allow one to compare the shapes of VDWSs for different molecules, as well as following the changes in the VDWS of a single molecule undergoing rearrangements in its internal geometry.

Ideally, the characterization of "shape" should be simple enough to allow a quick, non-visual computation, leading to descriptive information easy to store and analyze by a computer in an algebraic way. Moreover, the analysis of shape features should not be dependent solely on symmetry, in order to obtain non-trivial conclusions.

In this article we propose a method for such an analysis. The procedure is based on the computation of the symmetry-independent homology groups of algebraic topology of topological objects related to the molecular VDW surface, that is either a manifold, a manifold with boundary (or a collection of them), or topological objects derived from them by various truncations (for algebraic topological texts see, for instance, refs. [11,12]). Several proposals have been presented recently to analyze the shape of molecular surfaces, for fixed or varying molecular conformations [13–20]. These techniques are based on constructing new entities from the molecular surface, whose topological properties are then studied. The definition of these objects follows from the definition of domains on a molecular surface. Some alternatives are:

- (1) Definition of domains of specified curvature properties on an isodensity or electrostatic isopotential contour surface [13,14,17,20], as well as contours defined under the constraint of constant volume or constant fractional charge enclosed by the associated surface [15].
- (2) Definition of domains on a surface according to the pattern of interpenetration by another, physically different surface [16]. For instance, we have studied the interrelations of VDWSs and electrostatic potential contour surfaces [16,17].

The detailed analysis of a *single* VDWS requires a different strategy. The fact that every point on the VDWS lies at least on one spherical surface (a 2-sphere) makes the earlier partitioning based on curvature a rather trivial and not too informative one (at every point where the surface is differentiable, it is also locally convex and has curvature fixed by the radius of the VDW sphere). Accordingly, the curvature analysis is an inappropriate tool for the shape characterization of the VDWS [18,19]. Furthermore, although a general VDWS can be homeomorphic to a 2-torus, or to manifolds with higher genus, in most cases it will be topologically equivalent to a 2-sphere. Hence, VDWSs are usually topologically rather simple objects, and the characterization of their shapes in terms of their homeomorphism type reveals next to nothing about the chemically important details.

In this article we propose a simple, alternative method of analysis, well-suited to the problems of VDWSs. Our discussion will focus on the *decomposition* of the VDWS into subsets which are in turn characterized topologically; the information on how these subsets are glued together to reconstruct the complete VDWS can be treated by the incidence graph or shape graph techniques, proposed earlier [35] for general surfaces. It is the decomposition of a VDWS that allows for the utilization of special features, inherent in the definition of a VDWS, and these are the very features that simplify the analytic description of shape variations with conformational changes. The method proposed in this study does not give a complete description of the VDWS; instead, it extracts topologically significant shape information that can be stored by a computer and compared in a simple, non-visual, numerical form. The method may be used by itself, or it may be combined with an incidence graph method [35] if the emphasis is placed on the global, rather than on the local shape properties.

On a VDWS, there are two types of points: those which lie on the surface of a single sphere, and those which belong simultaneously to the surfaces of two or more spheres. In this latter case, a tangent plane to the surface cannot be defined. (Here, we are not concerned with *smoothed* VDWSs, such as those derived when computing solvent-accessibility surfaces [21–24]). If one eliminates the points of the second type from the surface, then the VDWS will be decomposed into open, spherical domains. These domains represent a collection of topological objects (spheres and punctured spheres) whose homology groups can be computed, that in turn provides a characterization of the shape of the VDWS. *Due to the fact that the VDWS has a simple structure, built from spherical domains, it is easy to compute the dependence of the homology groups on conformational rearrangements.* This is the particular advantage of the method proposed here, since one is able to describe boundaries of shape regions of the nuclear configuration space by simple, analytic formulas (*vide infra*). It is expected that the insight gained will be helpful in more general analyses of other molecular surfaces. The characterization we discuss here is very detailed, and it allows one to describe appropriately the shape features for all nuclear configurations. Moreover, the characterization of the VDWS in terms of a number of disjoint entities may be advantageous, since they are often *chemically*

*different* domains on the VDWS, which are better identified and described by considering them as separate objects. Even though for simple molecules the description might seem unnecessarily complicated for the VDWS associated with the equilibrium geometry, it reveals clearly a number of far from trivial results for nonequilibrium configurations.

The paper is organized as follows. In section 2 we discuss briefly the essentials of the method, and the computation of the homology group is illustrated with a simple example of the VDWS. In section 3 we analyze the partitioning of the configurational space into regions ("shape regions") where, within each shape region, the essential shape features of the VDWS are preserved, although they change when a configuration change leads from one shape region to another. The molecule of water is used as an example. Section 4 contains a brief discussion about the dependence of the shape region partitioning of the configurational space on the symmetry under nuclear permutation. Conclusions are found in section 5.

## 2. Shape characterization of a VDWS

Consider a molecule with  $N$  nuclei, where each of the nuclei (even those of identical nuclear charges) has a unique distinguishing label, as well as a serial index of an arbitrary but fixed ordering. A molecular geometry for this system is characterized by a vector in the  $3N$ -dimensional Euclidean space  ${}^{3N}\mathcal{R}$ , spanned by the  $3N$  (mass-weighted, laboratory frame) Cartesian coordinates of the ordered set of the  $N$  nuclei. Note that a fixed index ordering of the nuclei is necessary for a one-to-one assignment of each laboratory frame nuclear configuration to a unique point of the configuration space  ${}^{3N}\mathcal{R}$ . Furthermore, the molecule is characterized by the set of nuclear charges  $Z_1, Z_2, \dots, Z_N$ , represented by a formal vector

$$\mathbf{z} = (Z_1, Z_2, \dots, Z_N) \quad (1)$$

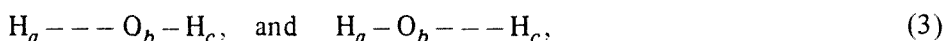
in an  $N$ -dimensional space, the nuclear charge space  ${}^N\mathcal{Z}$  [25–27] where, again, the component index follows the fixed ordering of the nuclei. We assume that the ordering is the same for the two spaces.

Any physical property that is suitable to describe the shape of the molecule (as, for instance, a VDWS assigned to it) has to be rotationally and translationally invariant within the laboratory frame, as long as there are no external fields. Accordingly, the reduced nuclear configuration space  $M$ , a *metric* space of dimension  $3N - 6$  for  $N > 2$ , where different points represent different *internal* configurations [28–31], is an appropriate space to study the transformations of the VDWS under conformational changes. This metric space  $M$  is a quotient space of  ${}^{3N}\mathcal{R}$  with respect to the *ty*-equivalence relation [28–31] in  ${}^{3N}\mathcal{R}$  defined by superposition under translation and rigid rotation of the *labeled and ordered* set of nuclei within the laboratory frame.

(Note that the actual, operational definition of those transformations in space  ${}^{3N}\mathcal{R}$  which represent rigid displacements of configurations within the laboratory frame require a fixed ordering of the nuclei [28–31]; in the absence of a fixed ordering of the nuclei, the rigid rotation and rigid translation operators  $Y$  and  $T$  of space  ${}^{3N}\mathcal{R}$  used in the definition of the above  $\boldsymbol{ty}$ -equivalence relation would not be sufficient to generate all superimposable configurations, that is, all those nuclear arrangements that are equivalent as internal configurations.) The elements of the quotient space  $M$  are the  $\boldsymbol{ty}$ -equivalence classes  $K$ , which represent the distinct internal configurations of the labeled and ordered nuclei,

$$M = \{K_\alpha\}, \quad \cup_\alpha K_\alpha = {}^{3N}\mathcal{R}. \quad (2)$$

Due to the individual labeling and ordering of the nuclei, two configurations for which rigid displacements can result in a superposition where the only requirement is that each nucleus is coincident with another one of the same nuclear charge, do not necessarily belong to the same equivalence class  $K$  if the superimposed nuclei differ in the nuclear labeling and serial indices. For example, take two asymmetrically stretched configurations of the  $\text{H}_2\text{O}$  molecule,



(denoted by  $\alpha$  and  $\beta$ , respectively), which are mirror images of each other, and where the serial indices of the sequence of nuclei  $\text{H}_a$ ,  $\text{O}_b$ , and  $\text{H}_c$  of labels  $a$ ,  $b$ , and  $c$  are 1, 2, and 3, in the order as listed from left to right, respectively. The two configurations are superimposable, with nuclear charges matching, if one disregards labels and ordering indices. However, there exist no rigid rotations and rigid translations within the laboratory frame which could bring these two configurations into superposition, with nuclear labels and serial indices matching. Consequently, these two configurations belong to two different  $K$  equivalence classes of space  ${}^{3N}\mathcal{R}$ .

Let

$$\boldsymbol{x} \in K \subset {}^{3N}\mathcal{R} \quad (4)$$

be any element of the equivalence class  $K$ . The vectors  $\boldsymbol{x}$  and  $\boldsymbol{z}$  contain an essential information about the molecular system that can be represented by a pair  $(\boldsymbol{x}, \boldsymbol{z})$  and may be regarded as an element of a product space,

$$(\boldsymbol{x}, \boldsymbol{z}) \in {}^{3N}\mathcal{R} \times {}^N\mathcal{Z}. \quad (5)$$

Our purpose is to characterize various VDWSs as a function of the internal geometry of the nuclear configuration, and to identify domains of nuclear configuration spaces  ${}^{3N}\mathcal{R}$  and  $M$  with similar VDWSs. For the purely shape features of the

VDWS, the actual serial indices and labeling of identical nuclei are irrelevant; however, these indices and labeling are essential in choosing an actual, operational representation of molecules in spaces  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$ . The conflict between the above two requirements can be circumvented by regarding certain  $K$  sets equivalent if they are related to one another by permutations of identical nuclei, and by introducing into spaces  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$  further equivalence classes, using various permutation operators.

In our later discussion, the symmetry properties of the product space  ${}^{3N}\mathcal{R} \times {}^N\mathcal{Z}$  under permutation of nuclei are of interest; some additional background information will be reviewed here as preparation for further applications. The three-dimensional symmetry properties of a nuclear configuration can be characterized by the effects of permutations of the triplets of Cartesian coordinates of the nuclei. In fact, one may consider three different types of permutations:

- (i) those acting in space  ${}^N\mathcal{Z}$ , permuting labeled nuclear charges with respect to the serial indices,
- (ii) those acting simultaneously on both spaces  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$ , permuting the serial indices of nuclei, thereby changing the serial index convention used in the definitions of these spaces, as well as the physical meaning of the vectors  $\mathbf{x}$  and  $\mathbf{z}$  in the two spaces  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$ , but leaving the *numerical values* of vector components assigned to each labeled nucleus unchanged, and
- (iii) those acting in space  ${}^{3N}\mathcal{R}$ , permuting triplets of coordinate values, with respect to the serial indices of nuclei.

Let  $\mathbf{p}'_{ij}$  be the permutation operator that exchanges the components  $i$  and  $j$  of the nuclear charge vector  $\mathbf{z}$ :

$$\begin{aligned}
 \mathbf{p}'_{ij}(\mathbf{x}, \mathbf{z}) &= \mathbf{p}'_{ij}(\mathbf{x}, (z_1, z_2, \dots, z_i, \dots, z_j, \dots, z_N)) \\
 &= \mathbf{p}'_{ij}(\mathbf{x}, (Z_a, Z_b, \dots, Z_r, \dots, Z_t, \dots, Z_w)) \\
 &= (\mathbf{x}, (Z_a, Z_b, \dots, Z_t, \dots, Z_r, \dots, Z_w)) \\
 &= (\mathbf{x}, (z'_1, z'_2, \dots, z'_i, \dots, z'_j, \dots, z'_N)) \\
 &= (\mathbf{x}, \mathbf{z}'),
 \end{aligned} \tag{6}$$

where in vector  $\mathbf{z}$  the values of components  $z_1$ ,  $z_2$ ,  $z_i$ , and  $z_j$  are the charges of nuclei of labels  $a$ ,  $b$ ,  $r$ , and  $t$ , respectively, whereas after the permutation in the resulting vector  $\mathbf{z}'$  the values of components  $z'_1$ ,  $z'_2$ ,  $z'_i$ , and  $z'_j$  are the charges of nuclei of labels  $a$ ,  $b$ ,  $t$ , and  $r$ , respectively. Operation  $\mathbf{p}'_{ij}$  does not affect the components of vector  $\mathbf{x}$ .

For the same index pair  $i, j$ , there are two other permutation operators of interest. The operator  $\mathbf{p}_{ij}''$  permutes the indices of ordering of the nuclei in the definition of both spaces  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$ :

$$\begin{aligned}
 \mathbf{p}_{ij}''(\mathbf{x}, \mathbf{z}) &= \mathbf{p}_{ij}''((\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots, \mathbf{x}_N), (z_1, z_2, \dots, z_i, \dots, z_j, \dots, z_N)) \\
 &= \mathbf{p}_{ij}''((R_a, R_b, \dots, R_r, \dots, R_t, \dots, R_w), (Z_a, Z_b, \dots, Z_r, \dots, Z_t, \dots, Z_w)) \\
 &= ((R_a, R_b, \dots, R_t, \dots, R_r, \dots, R_w), (Z_a, Z_b, \dots, Z_t, \dots, Z_r, \dots, Z_w)) \\
 &= ((\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_i, \dots, \mathbf{x}'_j, \dots, \mathbf{x}'_N), (z'_1, z'_2, \dots, z'_i, \dots, z'_j, \dots, z'_N)) \\
 &= (\mathbf{x}', \mathbf{z}'), \tag{7}
 \end{aligned}$$

where, for example, the symbol  $\mathbf{x}_i$  represents the triplet of the first three coordinate values in  $\mathbf{x}$ , that actually corresponds to the three laboratory frame Cartesian coordinates  $R_a$  of nucleus of label  $a$ . This permutation interchanges the original serial indices  $i$  and  $j$  of nuclei of labels  $r$  and  $t$  (of coordinates  $R_r$  and  $R_t$ , and of nuclear charges  $Z_r, \dots, Z_t$ , respectively). Note that operation by  $\mathbf{p}_{ij}''$  does not represent any physical change, it merely leads to an equivalent but different assignment of physical reality to elements of abstract spaces  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$ . These spaces are regarded as a common basis for describing and exploiting different representations, in particular those which differ merely in the serial indices of nuclei. Using more than one representation, one shall be able to describe those symmetry properties of configurations which are relevant to the shape characterization of VDW surfaces.

The third operator  $\mathbf{p}_{ij}'''$  permutes the two triplets of Cartesian coordinates of nucleus  $i$  and nucleus  $j$  in the  $3N$ -dimensional vector  $\mathbf{x}$ :

$$\begin{aligned}
 \mathbf{p}_{ij}'''(\mathbf{x}, \mathbf{z}) &= \mathbf{p}_{ij}'''((\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots, \mathbf{x}_N), \mathbf{z}) \\
 &= \mathbf{p}_{ij}'''((R_a, R_b, \dots, R_r, \dots, R_t, \dots, R_w), \mathbf{z}) \\
 &= ((R_a, R_b, \dots, R_t, \dots, R_r, \dots, R_w), \mathbf{z}) \\
 &= ((\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_i, \dots, \mathbf{x}'_j, \dots, \mathbf{x}'_N), \mathbf{z}) \\
 &= (\mathbf{x}', \mathbf{z}). \tag{8}
 \end{aligned}$$

Operation  $\mathbf{p}_{ij}'''$  does not affect the components of vector  $\mathbf{z}$ .

Operations analogous to  $\mathbf{p}_{ij}'$  and  $\mathbf{p}_{ij}'''$  may be regarded to act within spaces  ${}^N\mathcal{Z}$  and  ${}^{3N}\mathcal{R}$ , respectively. If, however, one considers all possible index orderings of

the nuclei, that is, all the possible  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$  representations, then it is more natural to regard all three of the above permutation operators as acting on vectors of the form  $(\mathbf{x}, \mathbf{z})$ . These three operators are then interdependent for any fixed pair of indices  $i$  and  $j$ :

$$\mathbf{p}'_{ij} \mathbf{p}''_{ij} = \mathbf{p}''_{ij} \mathbf{p}'_{ij} = \mathbf{p}'''_{ij}, \quad (9)$$

$$\mathbf{p}''_{ij} \mathbf{p}'''_{ij} = \mathbf{p}'''_{ij} \mathbf{p}''_{ij} = \mathbf{p}'_{ij}, \quad (10)$$

and

$$\mathbf{p}'''_{ij} \mathbf{p}'_{ij} = \mathbf{p}'_{ij} \mathbf{p}'''_{ij} = \mathbf{p}''_{ij}. \quad (11)$$

Consequently, if one denotes the identity operation on vectors  $(\mathbf{x}, \mathbf{z})$  by  $\mathbf{p}^\circ$  then for each index pair  $i, j$ , the set

$$\mathbf{P}_{ij} = \{\mathbf{p}^\circ, \mathbf{p}'_{ij}, \mathbf{p}''_{ij}, \mathbf{p}'''_{ij}\} \quad (12)$$

is a group, isomorphic to point group  $D_2$ , that provides an internal structure for the group  $\mathbf{P}$  containing as elements all possible products of all pairwise permutations

$$\mathbf{p}'_{ij}, \mathbf{p}''_{ij}, \mathbf{p}'''_{ij}, \quad 1 \leq i, j \leq N. \quad (13)$$

Consider again the example of the two distorted configurations of the water molecule (3), denoted by  $\alpha$  and  $\beta$ , and the pair permutations  $\mathbf{p}'_{13}, \mathbf{p}''_{13}, \mathbf{p}'''_{13}$ , involving the two H atoms of serial indices  $i = 1, j = 3$ . For the sake of simplicity, we assume that all  $Y$  and  $Z$  coordinates are zero, and the  $X$  coordinates are  $-\tau, 0, \rho$  in configuration  $\alpha$  and  $-\rho, 0, \tau$  in configuration  $\beta$ , for nuclei of indices 1, 2, and 3, respectively. We shall consider all real positive values for  $\rho$  and  $\tau$ . Consider the permutation  $\mathbf{p}'''_{13}$  of the Cartesian coordinate triplets of identical nuclei  $H_a$  and  $H_b$ ,

$$\begin{aligned} & \mathbf{p}'''_{13}(\mathbf{x}^\alpha, \mathbf{z}) \\ &= \mathbf{p}'''_{13}((\mathbf{x}_1^\alpha, \mathbf{x}_2^\alpha, \mathbf{x}_3^\alpha), (z_1, z_2, z_3)) \\ &= \mathbf{p}'''_{13}((R_a, R_b, R_c), (Z_a, Z_b, Z_c)) \\ &= \mathbf{p}'''_{13}((( -\tau, 0, 0), (0, 0, 0), (\rho, 0, 0)), (1, 8, 1)) \\ &= ((R_c, R_b, R_a), (Z_a, Z_b, Z_c)) \\ &= (\mathbf{x}', \mathbf{z}) \\ &= (((\rho, 0, 0), (0, 0, 0), (-\tau, 0, 0)), (1, 8, 1)). \end{aligned} \quad (14)$$



This permutation does not lead to a physically different molecule, as long as nucleus  $H_a$  is physically indistinguishable from  $H_b$ . However, in the definition of the actual spaces  ${}^{3N}\mathcal{R}$  and  ${}^N\mathcal{Z}$  the nuclear labels  $a$ ,  $b$ , and  $c$  are assigned to serial indices 1, 2, and 3, respectively, and the permutation  $\mathbf{p}'_{13}$  does alter this assignment. Hence, permutation  $\mathbf{p}'_{13}$  also alters the interpretations of elements of the space  ${}^{3N}\mathcal{R}$ . This is manifested in the fact that the two  ${}^{3N}\mathcal{R}$  space vectors,

$$\mathbf{x}^\alpha = ((-\tau, 0, 0), (0, 0, 0), (\rho, 0, 0)), \quad (15)$$

and

$$\mathbf{x}' = ((\rho, 0, 0), (0, 0, 0), (-\tau, 0, 0)) \quad (16)$$

are not necessarily *ty*-equivalent, in spite of the fact that they describe indistinguishable physical conformations (of two different label conventions). The two vectors may (but do not have to) belong to two different  $K$  equivalence classes of space  ${}^{3N}\mathcal{R}$ :  $\mathbf{x}^\alpha \in K_\alpha$  and  $\mathbf{x}' \in K'$ , where  $K_\alpha \neq K'$ . The special case of  $K_\alpha = K'$  arises if and only if  $\tau = \rho$ ; this latter condition, however, implies a *symmetry* of the configuration  $\alpha$ .

Note, on the other hand, that in general a rigid rotation *can* bring the labeled and indexed molecular structures of  $(\mathbf{x}', \mathbf{z})$  and  $(\mathbf{x}^\beta, \mathbf{z})$  into coincidence, where the vector

$$\mathbf{x}^\beta = ((-\rho, 0, 0), (0, 0, 0), (\tau, 0, 0)) \quad (17)$$

represents the collection of coordinates of configuration  $\beta$ . Consequently, vectors  $\mathbf{x}'$  and  $\mathbf{x}^\beta$  are necessarily *ty*-equivalent,

$$\mathbf{x}', \mathbf{x}^\beta \in K' = K_\beta. \quad (18)$$

The above observations on the symmetry properties of the example are easily generalized: the three-dimensional point group symmetry properties of nuclear configurations can be checked and detected by products of  $\mathbf{p}'_{i_k j_k}$  permutations. If and only if a  $\prod_k \mathbf{p}'_{i_k j_k}$  permutation of a configuration  $\mathbf{x}$  does not leave the corresponding equivalence class  $K$  that contains  $\mathbf{x}$ , that is, if and only if

$$\mathbf{x}, \left( \prod_k \mathbf{p}'_{i_k j_k} \right) \mathbf{x} \in K \quad (19)$$

for some  $K$ , then there exists a symmetry operation  $S$  of the molecular configuration  $\mathbf{x}$ , such that

$$\left( \prod_k \mathbf{p}'_{i_k j_k} \right) S\mathbf{x} = I\mathbf{x}, \quad (20)$$

where  $I$  is the identity operation.

Our purpose is to carry out a simplification of the shape analysis of all possible nuclear configurations, based on the identity of some nuclear charges. In addition, we shall investigate the effects of gradually increasing the number of identical nuclear charges by replacing some of the nuclei. Although it is natural to use spaces  ${}^{3N}\mathcal{R}$  and  $M$  for a study involving all possible nuclear configurations, nuclear charge variations are the easiest to monitor within the nuclear charge space  ${}^NZ$ ; hence, it is advantageous to reformulate coordinate permutations and symmetry in terms of nuclear charge permutations and formal index permutations. For this reason, we shall consider the products defined as

$$p_{i_k j_k} = p'_{i_k j_k} p''_{i_k j_k} \quad (21)$$

where the permutations  $p'_{i_k j_k}$  and  $p''_{i_k j_k}$  operate in space  ${}^NZ$  and on the index ordering of the nuclei, respectively. The group properties pointed out above imply that the effect of  $p_{i_k j_k}$  is equivalent to that of  $p'''_{i_k j_k}$ ; hence, one may replace the action of the  $p'''_{i_k j_k}$  permutations in space  ${}^{3N}\mathcal{R}$  by those of  $p'_{i_k j_k}$  and  $p''_{i_k j_k}$  operating in space  ${}^NZ$  and on the nuclear indices, respectively. This allows one to regard spaces  ${}^{3N}\mathcal{R}$  and  $M$  as abstract, mathematical entities without further concern to the specific index convention used for the nuclei. The symmetry condition can be restated as follows:

$$x, \left( \prod_k p_{i_k j_k} \right) x \in K \quad (22)$$

for some  $K$  if and only if there exists a symmetry operation  $S$  of the molecular configuration  $x$ , such that

$$\left( \prod_k p_{i_k j_k} \right) Sx = Ix, \quad (23)$$

Consider a product  $(\prod_k p_{i_k j_k})$  of permutations  $p_{i_k j_k}$  where all indices  $i_k$  and  $j_k$  refer to nuclei having the *same* nuclear charge. Such permutations leave the topological shape features of molecules invariant. Consequently, it is useful to introduce an equivalence relation of internal configurations  $K$  of metric space  $M$ , as follows:  $K_1$  and  $K_2 \in M$  are *zpe*-equivalent,

$$K_1 \text{ zpe } K_2, \quad (24)$$

if and only if there exist points  $x_1 \in K_1$  and  $x_2 \in K_2$ ,  $K_1, K_2 \subset {}^{3N}\mathcal{R}$ , and a product  $(\prod_k p_{i_k j_k})$  of permutations  $p_{i_k j_k}$  where all indices  $i_k$  and  $j_k$  refer to nuclei having the *same* nuclear charge, such that

$$\left( \prod_k p_{i_k j_k} \right) x_1 = x_2. \quad (25)$$

The **zpe** equivalence class to which  $K_1$  belongs is denoted by  $\mathbf{K}_1$ : in the above example of the definition,  $K_1, K_2 \in \mathbf{K}_1$ . The **zpe** notation in the equivalence relation stands for **z** permutation equivalence (accompanied by a re-indexing of the nuclei, according to the associated  $\mathbf{p}''_{i_k j_k}$  permutations).

Accordingly, the permutational symmetry in space  ${}^N Z$  is reflected in the internal structure of the space  $M$  in terms of equivalence classes  $\mathbf{K}$ . If the index pair  $i, j$  is involved in the construction of equivalence class  $\mathbf{K}$ , then we say that space  $M$  possesses  $i$ - $j$  symmetry under nuclear charge permutation.

For example, a triatomic molecule ABC has no permutational symmetry in the space  $M$  ( $Z_A \neq Z_B \neq Z_C \neq Z_A$ ), and a molecule  $AB_2$  has only  $i$ - $j$  symmetry in  $M$  under the B-B permutations. All these symmetry properties of the metric space  $M$  are inherited by any VDWS, attached to the molecule, which is chosen to describe the molecular shape. The relationship between the configurational space, the molecular shape, and the  $i$ - $j$  permutational symmetry will be used to introduce shape classes into configurational space  $M$ .

Let  $f(\mathbf{r})$  be an auxiliary function, which is a function of the variable  $\mathbf{r}$  in the physical three-dimensional space. In order to define the VDWS, one considers first a fixed configuration for the nuclei in the space  $M$ . Suppose that  $\mathbf{R}_i$  is the three-dimensional position vector of the  $i$ th nucleus ( $i = 1, 2, \dots, N$ ) and  $\rho_i$  is the corresponding atomic van der Waals radius. (According to the definition of space  ${}^{3N}R$ , the origin of the coordinate system for the vectors  $\mathbf{R}_i$  is arbitrary.) A sphere with radius  $\rho_i$  is introduced about each nucleus. The envelope surface of the object formed by the superposition of interpenetrating spheres is the VDWS [21–24]. We choose the function  $f(\mathbf{r})$  such that it takes a value equal to unity at the VDWS and zero elsewhere [16]. This allows one to use the standard description of the surface by formal level sets  $G$ . According to our choice, the VDWS is given as follows:

$$G(K, \mathbf{z}) = \{ \mathbf{r} \in {}^{3N}R : f(\mathbf{r}) = 1 \} , \quad (26)$$

where the notation  $G(K, \mathbf{z})$  stands for a given molecular configuration  $K$  and the set of nuclear charges represented by vector  $\mathbf{z}$ . The VDWS, given formally by eq. (26), is built up from pieces of spherical surfaces. In order to introduce the desired partitioning of set (26), we will distinguish those points on  $G(K, \mathbf{z})$  that lie on one spherical surface from those which belong to more than one. Let us introduce the following set  $D(\mathbf{r})$ , that contains as elements the distances from the point  $\mathbf{r}$  on the VDWS to the centres  $\mathbf{R}_i$  of various spheres:

$$D(\mathbf{r}) = \{ d_i : d_i = \|\mathbf{r} - \mathbf{R}_i\|, i = 1, 2, \dots, N, \mathbf{r} \in G(K, \mathbf{z}) \} . \quad (27)$$

Set  $D(\mathbf{r})$  is an ordered set,  $D(\mathbf{r}) = \{d_1, d_2, \dots, d_N\}$ , with indices identical to those of the ordering of the nuclei. The subset  $G'(K, \mathbf{z})$  of  $G(K, \mathbf{z})$ , defined as follows,

$$G'(K, \mathbf{z}) = \{r \in G(K, \mathbf{z}) : \exists d_i, d_j \in D(r), d_i = \rho_i, d_j = \rho_j, i \neq j\}, \quad (28)$$

contains all the points on the surface for which no unique tangent plane can be defined, for they belong to two or more interpenetrating spheres simultaneously. We are now in the position to define the set of topological objects, derived from the VDWS, that will be the subject of analysis. In general, this set is a collection of  $L$  disjoint sets, obtained as:

$$G(K, \mathbf{z}) = G(K, \mathbf{z}) \setminus G'(K, \mathbf{z}) = \bigcup_{i=1}^L G_i^{\circ}(K, \mathbf{z}), \quad L \geq 1, \quad (29)$$

where the  $G_i^{\circ}(K, \mathbf{z})$  sets are the maximum connected components of  $G(K, \mathbf{z})$ .

$$G_i^{\circ} \cap G_j^{\circ} = \emptyset, \quad i \neq j. \quad (30)$$

Here, the lines belonging to more than one VDW sphere are used to subdivide the VDWS. This subdivision itself may be considered as a graph (possibly a multigraph) that characterizes the shape of the VDWS. The actual analysis will be carried out using the closures  $G_i^{\circ}(K, \mathbf{z})$  of these sets  $G_i^{\circ}(K, \mathbf{z})$ . Note that sets  $G_i^{\circ}(K, \mathbf{z})$  do not contain their boundary lines, while sets  $G_i^{\circ}(K, \mathbf{z})$  do. For the sake of simplicity, we shall not always indicate  $K$  and  $\mathbf{z}$  when referring to the sets  $G_i^{\circ}$ . Intuitively, the family  $G_i^{\circ}$  corresponds to the collection of the closures of separate spherical pieces of a VDWS, obtained after "cutting" the VDWS along the lines generated by the overlapping atomic spheres.

An appropriate triangulation of a set  $G_i^{\circ}$  is equivalent to the construction of a simplicial complex. This triangulation can correspond only to that of a 2-sphere (a 2-pseudomanifold) or to that of the objects formed by removing a number of disks from a 2-sphere. The corresponding simplicial complexes are characterized by their  $p$ -dimensional homology groups  $H^p(G_i^{\circ})$  and their corresponding ranks, the Betti numbers  $b_p(H^p(G_i^{\circ}))$ ,  $p = 0, 1, 2$  [11,12]. These numbers, which are topological and homotopical invariants, provide a classification of the surface. The Euler–Poincaré characteristic  $\chi(G_i^{\circ})$  can also be used for classification [11,12], but the set of Betti numbers provides the more detailed, standard topological information used for description.

Our proposal is to describe the shape of the VDWS by the set of Betti numbers associated to each of the pieces into which is divided. By their definition, all the  $G_i^{\circ}$  sets are connected, and hence their zero-dimensional homology groups are isomorphic to the additive group of integers  $\mathbb{Z}$ , i.e.  $H^0(G_i^{\circ}) \cong \mathbb{Z}$ , for all indices  $i = 1, 2, \dots, L$ . This gives us the following trivial result for the corresponding Betti number:

$$b_0(H^0(G_i^c)) = 1, \quad i = 1, 2, \dots, L. \quad (31)$$

Hence, we may omit the zero-dimensional groups from our discussion. Note that from the number of holes obtained by the removal of disks one may deduce the remaining Betti numbers, and one could formulate a shape characterization based on the number of holes. However, the description based on Betti numbers is more general, and it allows direct comparisons with other topological characterizations where more complicated cases may occur, which cannot be characterized by the number of holes. Whereas the one-dimensional Betti number  $b_1$  provides the most important shape information, it is not sufficient by itself: a complete VDW sphere and a sphere from where one disk has been removed have the same value for Betti number  $b_1$ , ( $b_1 = 0$ ), and it is the two-dimensional Betti number  $b_2$  that distinguishes these two cases ( $b_2 = 1$  and  $b_2 = 0$ , respectively).

The analysis of the one- and two-dimensional groups can be summarized in a compact fashion by listing the Betti numbers in a matrix  $\mathcal{B}$ :

$$\mathcal{B} \in {}^{2 \times L} \mathbb{Z}; \quad (\mathcal{B})_{pi} = b_p(H^p(G_i^c)) \geq 0, \quad p = 1, 2; \quad i = 1, 2, \dots, L. \quad (32)$$

The row index of  $\mathcal{B}$  identifies the dimension of the homology group, whereas the column index corresponds to the serial index  $i$  of the simplicial cell complex obtained from  $G_i^c$ . It is noteworthy that the choice of index  $i$  given to a domain  $G_i^c$  is arbitrary. Therefore, any *column permutation* of the matrix  $\mathcal{B}$  merely leads to a different index convention and identifies the same physical object.

As indicated by eq. (29), the shape of the VDWS is dependent in general on the configuration  $K$  and the defining set of nuclear charges. Accordingly, the shape of the VDWS will be characterized by a matrix  $\mathcal{B}(K, \mathbf{z})$ . However, not every change in configuration  $K$  leads to a significant difference in shape. In fact, the purpose of our approach is to classify the configurations in space  $M$  according to the property of having the same intrinsic, topological shape for the VDWS. To this end, we will define here "VDWS shape regions" in space  $M$  as subsets  $M'_\alpha(\mathbf{z})$ , where a shape region contains all the configurations  $K$  characterized by a specified matrix  $\mathcal{B}_\alpha(\mathbf{z})$  or matrices equivalent to it by column permutations. That is to say:

$$M'_\alpha(\mathbf{z}) \subset M, \quad M'_\alpha(\mathbf{z}) = \{K_i \in M : \mathcal{B}(K_i, \mathbf{z}) \approx \mathcal{B}_\alpha(\mathbf{z})\}, \quad (33)$$

where the symbol " $\approx$ " stands to indicate that the matrices  $\mathcal{B}(K_i, \mathbf{z})$  and  $\mathcal{B}_\alpha(\mathbf{z})$  must have the same elements after a finite number of column permutations.

The matrix  $\mathcal{B}_\alpha(\mathbf{z})$  describes, in general, the shape of the VDWS for a family of related, but infinitely many configurations. Moreover, the VDWS shape regions  $M'_\alpha$  provide a rigorous partitioning of the overall configurational space:

$$M = \bigcup_{\alpha} M'_\alpha(\mathbf{z}). \quad (34)$$

It is worth mentioning that a similar partitioning was discussed previously, in a qualitative approximation, for isodensity contour surfaces [15], and in the context of the "dynamic shape space" [35].

According to our notation, two sets  $M'_\alpha$  and  $M'_\beta$ , with  $\alpha \neq \beta$ , have different matrices  $\mathcal{B}$ , i.e. different VDWS shape groups. We say in this case that the two sets  $M'_\alpha$  and  $M'_\beta$  belong to two different *VDWS shape classes*.

Notice that the set  $M'_\alpha$  is not necessarily pathwise connected, it may be formed by several disjoint subsets. To discuss this point, we shall regard a path in  $M$  as a mapping between two topological spaces:

$$P: (I, T) \rightarrow (M, T'_M), \quad I = [0, 1], \quad (35)$$

where  $I$  is the unit interval of parameter values,  $T$  and  $T'_M$  are the respective metric topologies. Although comparisons with other, chemically motivated topologies of space  $M$ , for example those that are defined in terms of the energy hypersurface (see, for instance, refs. [28–31]), may lead to interesting insight, in this study we shall consider only metric topologies. Since our primary interest is the study of shape changes due to *conformational rearrangements* (as opposed to actual chemical reactions), we consider here only those transformations of the configurational framework that preserve the connectedness of the VDWS. Note, however, that an analogous treatment is also applicable to general chemical reactions. A conformational rearrangement can be viewed as a process that continuously interconverts a series of molecular surfaces, and in the present context, energy changes may be disregarded. We shall discuss paths  $P$  of configuration changes in  $M$  with emphasis on the geometrical properties and we shall not exclude possible shapes based on their energetic instability.

Let  $K_i \in M'_\alpha$  and  $K_j \in M'_\alpha$ ,  $i \neq j$ , be the endpoints of a path  $P$ , so that:  $P(0) = K_i$  and  $P(1) = K_j$ . We say that  $M'_\alpha$  is pathwise connected if there exists a path  $P(u)$  for each pair  $K_i, K_j \in M'_\alpha$ , so that:  $P(u) \in M'_\alpha$  for all  $u \in (0, 1)$ . Otherwise, the set is pathwise disconnected. In the latter case, it is possible to connect  $K_i$  and  $K_j$  by a path only by crossing through at least one different shape region in  $M$ . If  $M'_\alpha$  is pathwise disconnected, then it can be written as the union of its maximum pathwise connected subsets  $M'_{\alpha i}$ .

In configurational space, the sets  $M'_{\alpha i}$ , with different  $i$  indices but with the same  $\alpha$  index, can be seen as separate, disjoint *shape regions*. However, they all belong to the same *shape class*  $M'_\alpha$ . In the next section we study some examples showing this property.

The shape class specified by  $\mathcal{B}_\alpha(\mathbf{z})$  is associated to a given set of nuclear charges  $\mathbf{z}$ . A variation of the values of  $\{Z_i\}$  may or may not, in general, lead to essential differences in the shape of a molecular surface. The topologically significant changes of a VDWS, induced by a change  $\mathbf{z} \rightarrow \mathbf{z}'$  of nuclear charges, leading to a change of the *shape regions*,

$$M'_\alpha(\mathbf{z}) \neq M'_\alpha(\mathbf{z}'), \quad \mathbf{z} \neq \mathbf{z}', \quad (36)$$

are discussed further in the following sections.

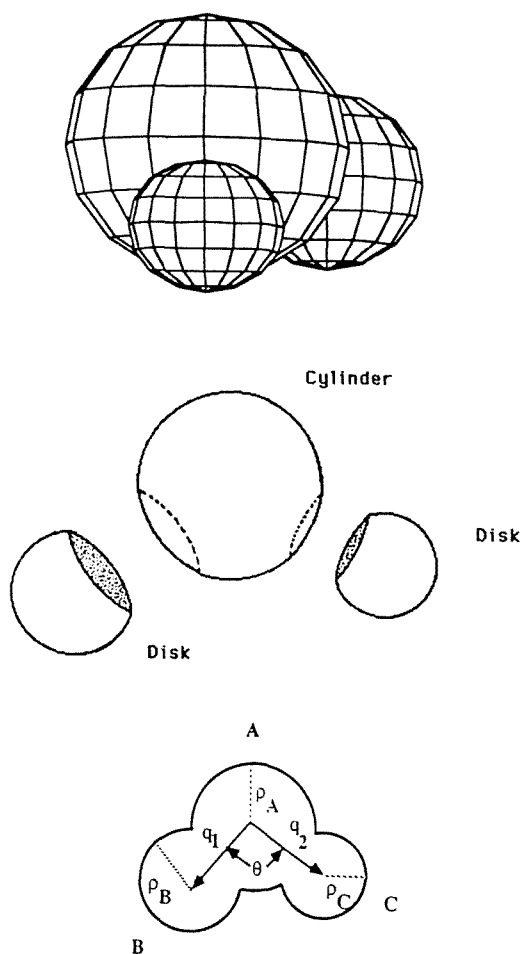


Fig. 1. Partitioning of the van der Waals surface for an ABC triatomic molecule (the notation used for the internal coordinates is also given).

We consider an example to illustrate the actual computation of the matrix  $\mathcal{B}$ . Figure 1 shows a typical VDWS for a triatomic molecule ABC, with all the van der Waals atomic radii different. The study of properties of this system in the reduced nuclear configuration space  $M$  is equivalent to considering a space of only three internal coordinates, the internuclear distances  $q_1(A-B)$ ,  $q_2(A-C)$ , and the bond angle  $\theta$  (see fig. 1). The spatial array chosen for the nuclei defines a configuration, belonging, say, to the equivalence class  $K_i$ . In fig. 1 we also display the resulting domain partitioning of the VDWS according to the discussion above. In this case, we find the set  $G^{\circ}(K_i, z)$  (eq. (29)) composed by three disjoint pieces. One of them is a 2-sphere with two holes, which is homeomorphic to a cylinder. The other two pieces are spheres with a single hole, which are homeomorphic to the two-dimensional disk. All this information is condensed within the following matrix:

$$\mathcal{B}(K_i, \mathbf{z}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{z} = (Z_A, Z_B, Z_C), \quad (37)$$

where the first column stands for the cylinder, and the last two for the disks. The same procedure can be applied to any other conformation. The triatomic system is the simplest, non-trivial one that we can consider to study the shape region partitioning of  $M$ . Its simplicity allows one to carry out the computations analytically, as shown in the next section.

### 3. Shape region partitioning of space $M$ for the system $AB_2$

In the case of analyzing a triatomic VDWS, the shape study of the surface is reduced to the analysis of the geometrical relationships of at most three circles obtained when intersecting the van der Waals atomic spheres with the plane defined by the three nuclei. Our problem now is to determine the relations among the coordinates  $q_1, q_2$  ( $q_i \geq 0, i = 1, 2$ ) and  $\theta$  ( $0 \leq \theta \leq 180^\circ$ ) corresponding to the *boundaries* for the shape regions.

Suppose that  $\rho_A$  and  $\rho_B$  are the van der Waals atomic radii for the atoms A and B, respectively. We assume that  $\rho_B < \rho_A$ . For the shapes of the objects obtained after the partitioning into domains, only the following situations are expected:

(i) *A single 2-sphere*

(both spheres of B are contained inside A, and the surfaces of A and either sphere of type B have no point in common):

$$\mathcal{B}(K, \mathbf{z}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (38)$$

(ii) *Two 2-spheres*

(either *one* sphere B is contained inside A, and the other is separate, or the two B spheres are separated from A, but having their centres coincident):

$$\mathcal{B}(K, \mathbf{z}) = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}. \quad (39)$$

If the centres coincide, i.e. if the two 2-spheres have all their points in common, then we regard it *a single 2-sphere*. This case corresponds to a degenerate situation in which the number of nuclei of the system can be formally considered as reduced by one (e.g. by setting the number of nuclei equal to  $N - 1$ ). The term "separate spheres" means that the two spheres have no common points.



*(iii) A cylinder and two disks*

(one sphere intersects the other two, but the latter ones do not intersect each other *outside* the former). (This situation is described by eq. (37).)

*(iv) Two disks*

(two spheres overlap partly, and either the two spheres B coincide, or one of the B spheres is totally contained within A):

$$\mathcal{L}(K, \mathbf{z}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \quad (40)$$

*(v) Three disks*

(each sphere intersects partly with the other two, and the overall VDWS is homeomorphic to a 2-sphere):

$$\mathcal{L}(K, \mathbf{z}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (41)$$

*(vi) Three cylinders*

(each sphere intersects partly with the other two, and the overall VDWS is homeomorphic to a 2-torus):

$$\mathcal{L}(K, \mathbf{z}) = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \quad (42)$$

*(vii) One 2-sphere and two disks*

(one sphere is separate and the other two intersect each other, without having coincident centres):

$$\mathcal{L}(K, \mathbf{z}) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \quad (43)$$

*(viii) Three 2-spheres (all are separate)*

$$\mathcal{L}(K, \mathbf{z}) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}. \quad (44)$$

To analyze the occurrence of the above shape features, we have performed a study by considering two subsets of space  $M$ . One corresponds to the plane  $\theta, q$ , with  $q = q_1 = q_2$  (that is, to bending deformations and symmetry stretchings), and the other to the plane  $q_1, q_2$ , for a variety of fixed values of the angle  $\theta$  (that is, to stretchings at various fixed bond angles).

In the first case, the subsets of the shape regions are defined by the following boundaries:

$$(1) \quad q < \rho_A - \rho_B, \quad 0 \leq \theta \leq 180^\circ: \quad (45)$$

a single 2-sphere.

$$(2) \quad \rho_A - \rho_B \leq q \leq \rho_A + \rho_B, \quad 0 < \theta(q) < g_1(q),$$

$$g_1(q) = 2 \arccos \{(q^2 + \rho_A^2 - \rho_B^2)/2q\rho_A\}: \quad (46)$$

three disks.

$$(3) \quad \rho_A - \rho_B \leq q < (\rho_A^2 + \rho_B^2)^{1/2}, \quad g_1(q) \leq \theta(q) \leq 180^\circ, \quad \text{and}$$

$$(\rho_A^2 + \rho_B^2)^{1/2} \leq q \leq \rho_A + \rho_B, \quad g_2(q) = 2 \arcsin \{\rho_B/q\} \leq \theta(q) \leq 180^\circ: \quad (47)$$

one cylinder and two disks.

Observe that at the critical distance  $q^* = (\rho_A^2 + \rho_B^2)^{1/2}$ , one finds:  $g_1(q^*) = g_2(q^*)$ , and  $(\partial g_1/\partial q)_{q=q^*} = (\partial g_2/\partial q)_{q=q^*} = -(\rho_B/\rho_A)(\rho_A^2 + \rho_B^2)^{-1/2}$ . Consequently, the functions  $g_1$  and  $g_2$  osculate at  $q^*$ , and they determine a single boundary function for the shape region, defined sectionally as follows:  $g(q) = g_1(q)$ , if  $q \in [\rho_A - \rho_B, q^*]$ , and  $g(q) = g_2(q)$ , if  $q \in [q^*, \rho_A + \rho_B]$ . It is noteworthy that  $q^*$  is an inflexion point of  $g(q)$ .

$$(4) \quad (\rho_A^2 + \rho_B^2)^{1/2} \leq q \leq \rho_A + \rho_B, \quad 0 < \theta(q) < g_2(q): \quad (48)$$

three cylinders.

$$(5) \quad q > \rho_A + \rho_B, \quad 0 < \theta(q) < g_2(q): \quad (49)$$

one 2-sphere and two disks.

$$(6) \quad q > \rho_A + \rho_B, \quad g_2(q) \leq \theta(q) \leq 180^\circ: \quad (50)$$

three 2-spheres.

$$(7) \quad \rho_A - \rho_B \leq q \leq \rho_A + \rho_B, \quad \theta = 0: \quad (51)$$

two disks.

$$(8) \quad q > \rho_A + \rho_B, \quad \theta = 0: \quad (52)$$

two 2-spheres.

The above shape regions are shown in fig. 2, using the example of the molecule of water. In this case, A corresponds to the oxygen atom ( $\rho_A = 1.4 \text{ \AA}$ ) and B to the

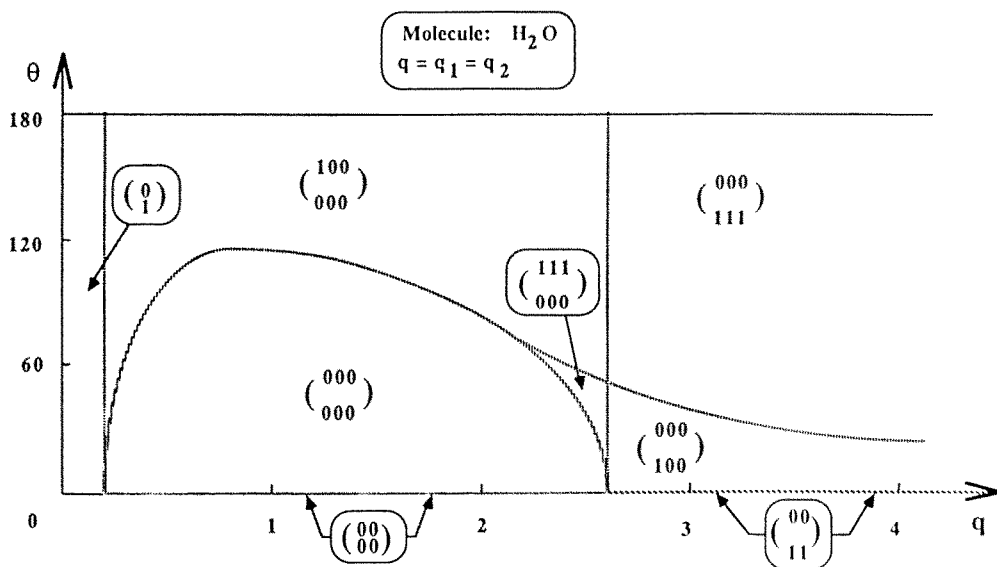


Fig. 2. Shape regions in the plane  $\theta$  (degrees),  $q$  (Å) for the molecule of water (the two coordinates describe bending and symmetric stretching, respectively).

hydrogen atoms ( $\rho_B = 1.2$  Å). The values for the van der Waals radii were taken from ref. [32]. The values agree approximately with an average of the values given by Bondi [33] and the "effective" van der Waals radii of ref. [34]. As we notice from fig. 2, there exists a critical bond angle  $\theta^*$  above which the shape region corresponding to the three disks cannot be reached by symmetrical stretching. This angle  $\theta^*$  is easily shown to be:

$$\theta^* = 2 \arccos\{(\rho_A^2 - \rho_B^2)^{1/2}/\rho_A\}, \quad (53)$$

that is,  $\theta^* \approx 118^\circ$  for  $H_2O$ .

It is worth reiterating that the diagram represents a cross section of the shape regions along the plane  $q_1 = q_2$ . The description of the overall shape regions is completed with the analysis of the planes  $q_1, q_2$  for different values of the bond angle  $\theta$ . Notice that fig. 2 shows the occurrence of two-dimensional, as well as one-dimensional shape regions.

The system considered ( $AB_2$  molecule) possesses a reduced nuclear configuration space  $M$  with B–B permutational symmetry in  $^N Z$  (see section 2). This assures us that only a subset  $S$  in the plane  $q_1, q_2$  is needed for our present study. In our case, we choose  $S = \{(q_1, q_2): 0 \leq q_1 \leq q_2\}$ , and the pattern of shape regions in other parts of the plane follows from symmetry.

Figure 3 displays the shape region partitioning of the plane  $q_1, q_2$  ( $\theta = 180^\circ$ ) (that is, the linear system B–A–B). In this case, the boundaries are straight lines

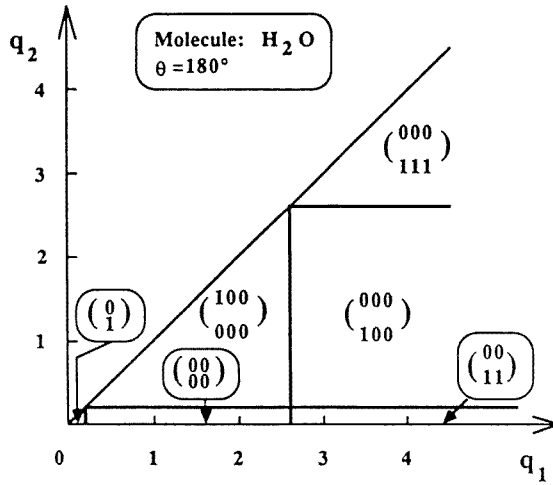


Fig. 3. Shape regions in the plane  $q_1, q_2$  (in Å) for the molecule of water, at  $\theta = 180^\circ$ .

parallel to the axes  $q_2$  and  $q_1$ . The pattern is very simple and no further comment is necessary. Notice, however, the occurrence of boundary points in space  $M$  where two or more shape regions meet. Figure 3 shows one point where *four* regions meet (that point is  $(q_1, q_2) = (\rho_A - \rho_B, \rho_A + \rho_B)$ ). According to our discussion above, following the definition of the shape regions, the *boundary point itself* belongs to the region characterized by eq. (37).

Figure 4 shows similar results for the linear system A–B–B (plane  $q_1, q_2$  ( $\theta = 0^\circ$ )). We notice here two distinct features, when comparing it to fig. 3:

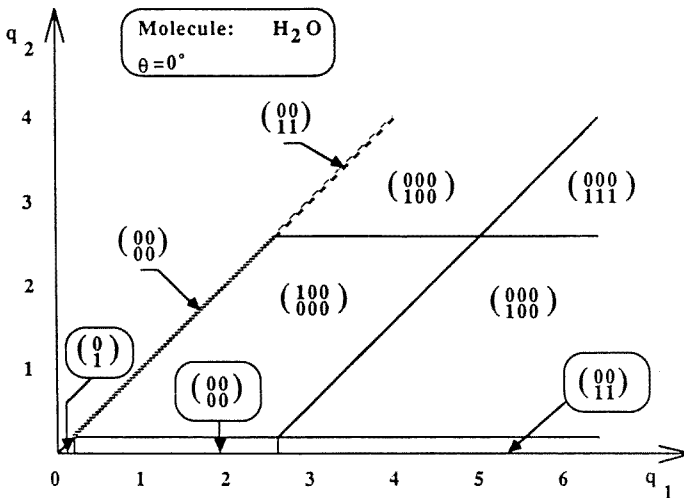


Fig. 4. Shape regions in the plane  $q_1, q_2$  (in Å) for the molecule of water, at  $\theta = 0^\circ$ .



(i) There exist boundaries which are straight lines parallel to the axes  $q_2$  and  $q_1$ , and others parallel to the diagonal  $q_2 = q_1$ . (ii) Some new one-dimensional shape regions also occur.

Figures 5 and 6 display more interesting results, corresponding to the situation with intermediate values of angle  $\theta$ . Figure 5 corresponds to the angle  $\theta = 100^\circ$  for water (that is, to the plane  $q_1, q_2$  ( $\theta = 100^\circ$ )). The features shown in this figure are characteristic for any other angle satisfying the following relation:

$$2 \arccos \{ [1 + (\rho_B/\rho_A)^2]^{1/2} \} < \theta \leq 2 \arcsin \{ [1 - (\rho_B/\rho_A)^2]^{1/2} \}. \quad (54)$$

The most characteristic feature in the diagram is, of course, the occurrence of a region with a curved boundary. A simple geometric reasoning shows that the boundary can be represented by a multivalued function  $q_2 = h_1(q_1; \theta)$  obtained from the roots of the following quadratic equation:

$$q_2^2 - q_2 \{ 2\rho_A \cos[\theta - \arccos\{(q_1^2 + \rho_A^2 - \rho_B^2)/2\rho_A q_1\}] \} + (\rho_A^2 - \rho_B^2) = 0. \quad (55)$$

The function  $h_1(q_1; \theta)$  defines a region corresponding to an "island" of configurations where the shape of the VDWS is characterized by eq. (41) (three disks). When considering the entire parameter plane, this region is totally surrounded by another region characterized by eq. (37). Two interesting consequences arise from the properties of  $h_1(q_1; \theta)$ , solution of eq. (55):

- (i) The boundary becomes perpendicular to the diagonal  $q_1 = q_2$ , i.e.  $(\partial h_1 / \partial q_1)_{[q_1 = q_2]} = -1$ .
- (ii) The region enclosed by  $h_1(q_1; \theta)$  is a convex point set. Consequently, the region characterized by eq. (37) is not convex if  $\theta$  satisfies eq. (54).

Figure 6 shows the results for  $H_2O$  with  $\theta = 50^\circ$ . The features present are characteristic of all angles satisfying the condition:

$$0 < \theta < 2 \arcsin \{ \rho_B / (\rho_A + \rho_B) \}. \quad (56)$$

Comparing fig. 5 with fig. 6, one notices a pattern much richer in details in the latter case. The former "island" region whose shape descriptors are condensed in eq. (41) has now frontiers with additional shape regions.

A new curved boundary  $q_2 = h_2(q_1; \theta)$  also appears. It is defined by the roots of the equation:

$$q_2^2 - q_2 (2q_1 \cos \theta) + (q_1^2 - 4\rho_B^2) = 0, \quad (q_1, q_2) \in S, \quad (57)$$

as long as  $q_1 < q_1^*$ , where  $q_1^*$  is the maximum  $q_1$  value compatible with the condition  $h_2(q_1; \theta) = \rho_A + \rho_B$ . Notice that the point  $(q_1, q_2) = (q_1^*, \rho_A + \rho_B)$  is another

boundary point where four regions meet. Two of the regions are of the same shape class (see section 2), for both are characterized by eq. (43). However, the point  $(q_1^*, \rho_A + \rho_B)$  belongs strictly to the region characterized by eq. (37); therefore, the above two regions are *pathwise disconnected* subsets (in the sense of the discussion in section 2) of the same shape class. This disconnectedness can be interpreted physically; although both regions belong to the same shape class, they identify two different situations. In one of the cases the separate 2-sphere is of type B, in the second case it is of type A.

Analogously, we notice the presence of several regions of the shape class specified by eq. (37). In one of the cases, two of these regions meet in another boundary point. At this point, the boundaries  $h_1(q_1; \theta)$  and  $h_2(q_1; \theta)$  merge, resulting in a unique, sectionally-defined boundary. The point itself belongs to the shape region characterized by eq. (41); consequently, the two previously mentioned regions specified by eq. (37) are pathwise disconnected subsets of the same shape class. Once again, they represent two situations that are physically different. In one of the cases the topological cylinder is of type B, in the second it is of type A.

#### 4. Shape region partitioning and symmetry under nuclear charge permutation

One obvious question that arises after the analysis performed in section 3 can be expressed as follows. Let us fix the number of nuclei at  $N$ ; in this case, how does the shape region partitioning  $\{M'_\alpha\}$  depend on  $\mathbf{z}$ ? In this section we discuss briefly this problem, upon analyzing the changes in the partitioning induced by a transformation  $\mathbf{z}_1 \rightarrow \mathbf{z}_2$ , and specifically, under permutations of the nuclear charges, so that the symmetry of the respective  $M$  space is altered.

Let us consider a point  $\mathbf{z}_n \in {}^NZ$ ,  $\mathbf{z}_n = (Z_{1n}, Z_{2n}, \dots, Z_{Nn})$ , where all the components of  $\mathbf{z}_n$  are numerically different. Of course, in this case, the space  $M$  has *no*  $i$ - $j$  permutational symmetry for all  $i, j \in A$ ,  $A = \{1, 2, 3, \dots, N\}$ . (See section 2 with regards to the permutation operator  $\mathbf{p}'_{ij}$  in  ${}^NZ$ , and the nature of the permutational symmetry in  $M$ .) The set  $A$  contains the  $N$  indices of the components of a nuclear charge vector. In this case, it is only the identity permutation that leaves  $\mathbf{z}_n$  invariant. Let us define a subset  $\Omega_1 \subset {}^NZ$  that contains all the points of the above property:

$$\Omega_1 = \{ \mathbf{z}_n \in {}^NZ : \mathbf{p}'_{ij} \mathbf{z}_n \neq \mathbf{z}_n, \quad i, j \in A, \quad i \neq j \}. \quad (58)$$

Our purpose is to generate sequences of changes in the nuclear charges and to follow the variation induced in the shape domain decomposition of the configuration space  $M$ . A family of important sequences can be constructed by systematically replacing different nuclei with identical ones, and one may use the permutational symmetry

of the resulting vectors of nuclear charges for classification. A detailed classification is provided by the corresponding permutation group; however, for certain special sequences, a simpler classification is sufficient. In this study we shall consider special sequences of nuclear charge vectors derived from elements of set  $\Omega_1$ : in each such vector  $\mathbf{z}$ , only one nuclear charge value may occur more than once as an element, and all other values are distinct. For such vectors, and for an arbitrary nuclear geometry in  ${}^3N\mathbb{R}$ , we can define a hierarchy of subsets of  ${}^NZ$  with increasing  $i$ - $j$  permutational symmetry. To that purpose, we introduce sets  $\Omega_\Gamma$ , where in accord with the notation in eq. (58), the subindex  $\Gamma$  in  $\Omega_\Gamma$  indicates the number of different permutation operations (including the identity) that leave each element  $\mathbf{z}_n \in \Omega_\Gamma$  invariant. In the following, we use index  $\Gamma$  for classifying the above special subsets  $\Omega_\Gamma$  of  ${}^NZ$ .

For the more general case, the integer number  $\Gamma$  depends on the number of classes of equivalent nuclei, and on the number of nuclei in each of them. For a general permutation, the number  $\Gamma$  does not provide a complete characterization, since different permutation groups may have the same order; in such cases, the permutation group itself may be used for indexing. However, if one considers the above family of subsets  $\Omega_\Gamma$  of  ${}^NZ$ , then the number  $\Gamma$  is sufficient for classification. If the subset  $\Omega_{\Gamma(p)}$  of  ${}^NZ$  contains all the points where there are precisely  $p$  equal nuclei, and the remaining  $N - p$  nuclei are all different from one another, as well as from the first  $p$  nuclei, then one has  $\Gamma(p) = p!$  different permutation operations that leave a vector  $\mathbf{z}_n \in \Omega_{\Gamma(p)}$  unaltered. (In fact, number  $p$  is an appropriate classification index for this special case; however, we shall use index  $\Gamma(p)$  in order to maintain reference to permutation groups.) Formally, for any  $p$ ,  $1 < p \leq N$ , the set  $\Omega_{\Gamma(p)}$  may be given as follows.

$$\Omega_{\Gamma(p)} = \{\mathbf{z}_n \in {}^NZ : \exists A' = (i_1, i_2, \dots, i_p) \subset A : \\ (\mathbf{z}_n)_\ell \neq (\mathbf{z}_n)_i = (\mathbf{z}_n)_j \neq (\mathbf{z}_n)_k \neq (\mathbf{z}_n)_\ell, \forall i, j \in A'; \forall k, \ell \notin A'\}. \quad (59)$$

As an example, for a triatomic system ( $N = 3$ ), any vector  $\mathbf{z}$  has the form  $(Z_1, Z_2, Z_3)$ . The set  $\Omega_{\Gamma(1)}$  corresponds to the systems of type ABC,  $\Omega_{\Gamma(2)}$  to those of type AB<sub>2</sub>, and  $\Omega_{\Gamma(3)}$  to those of type B<sub>3</sub>.

The dependence of the shape region partitioning  $M'_\alpha$  of nuclear configuration space  $M$  on the systematic replacement of elements of  $\mathbf{z} \in {}^NZ$  with identical nuclear charge values can be studied by computing the shape groups for each subset of  ${}^NZ$  obtained in the sequence  $\Omega_{\Gamma(1)} \rightarrow \Omega_{\Gamma(2)} \rightarrow \dots$ . A formal treatment can be given by introducing a family  $t$  of  $t_p$  endomorphisms in  ${}^NZ$ ,  $p = 1, 2, \dots$ , such that:

$$t = \{t_p : {}^NZ \rightarrow {}^NZ, p \geq 1, \text{ if } \mathbf{z}_n \in \Omega_{\Gamma(p)} \Rightarrow t_p(\mathbf{z}_n) \in \Omega_{\Gamma(p+1)}\}, \quad (60)$$

generating the sequence:



$$\begin{aligned} \{M'_\alpha(z)\} &\rightarrow \{M'_\alpha(t_1(z))\} \rightarrow \{M'_\alpha(t_2 t_1(z))\} \rightarrow \dots \\ &\dots \rightarrow \left\{ M'_\alpha \left( \left[ \prod_{i=1}^p t_i \right] (z) \right) \right\} \rightarrow \dots \quad z \in \Omega_1, \end{aligned} \quad (61)$$

where the continuing product  $\Pi$  refers to the composition of endomorphisms. The above sequence describes the changes of shape regions  $M'_\alpha$  of the nuclear configuration space  $M$  as gradually more and more nuclei become identical, following the chosen sequence of functions  $t_p$  in set  $t$ .

The above sequence, and more general sequences involving more complicated permutational equivalences of nuclei, provide descriptions of the *changes in the topology of the VDWS shape region partitioning of  $M$* , as induced by the associated nuclear charge changes. A similar, although somewhat more involved, method can be applied to other shape descriptors, such as shape groups or shape graphs of electronic charge densities (or electrostatic potentials, or selected molecular orbitals), based on a decomposition of the dynamic shape space [35], defined as  $D = (a, b) \otimes M$ , where  $a$  and  $b$  are an isodenisty contour value and a curvature parameter, respectively. The shape region decomposition of  $D$  is also dependent on the nuclear charges, and one may analyze this dependence using similar sequences of nuclear charge replacements. This interesting problem will be the subject of a subsequent study. We mention here a few of the properties expected for sequences (61):

- (i) Along the sequence, the number of shape classes could not increase.
- (ii) Along the sequence, disjoint shape regions of the same shape class cannot become pathwise connected. However, regions may become disconnected when the index  $p$  in (61) increases.
- (iii) The dimensionality of a shape region in space  $M$  cannot increase.
- (iv) The diagrams of shape regions encountered along the sequence (61) are homotopically related.

We consider here a simple, illustrative example of the above properties: the sequence  $ABC \rightarrow AB_2 \rightarrow B_3$  for a triatomic VDWS.

Figure 7 shows the results for the shape regions in the plane  $q_1, q_2$  ( $\theta = 0$ ). The values taken for the atomic radii were:  $\rho_A = 2.0$ ,  $\rho_B = 1.5$ , and  $\rho_C = 1.0$ . The middle diagram in fig. 7 is equivalent to fig. 4. All the features found in it agree with the properties listed above. Notice, for example, that region "c" (eq. (39)) is contracted to segments, and the same holds for region "b" (eq. (40)) which, in addition, becomes disconnected. Observe that region "a" (eq. (38)) is contracted to a point (see comment in section 3 regarding coincident spheres).

Figure 8 displays the view of the shape group partitioning of the plane  $\theta, q$  ( $q = q_1 = q_2$ ) for the system ABC with the atomic radii given above. This figure is

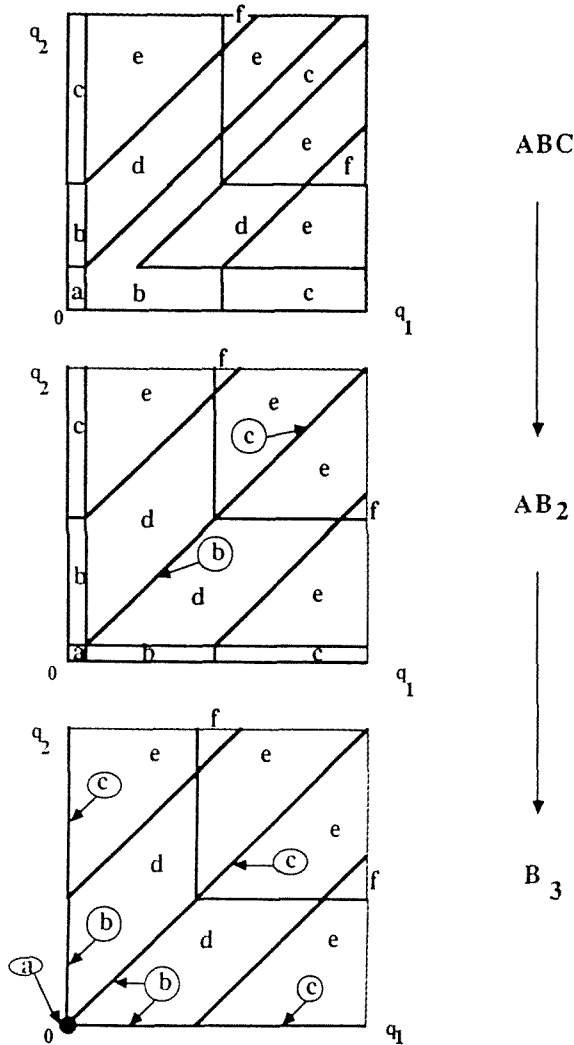


Fig. 7. Shape regions in the plane  $q_1, q_2$  ( $\theta = 0$ ) for the sequence of systems  $ABC \rightarrow AB_2 \rightarrow B_3$ . The notation a, b, c, d, e and f stands for the shape regions characterized by eqs. (38), (40), (39), (37), (43), and (44), respectively. (Radii used:  $\rho_A = 2.0$ ,  $\rho_B = 1.5$ ,  $\rho_C = 1.0$ .)

to be compared with fig. 2 for the  $AB_2$  system. The contraction of the region characterized by eq. (40) is a remarkable feature due to the change  $ABC \rightarrow AB_2$ .

The above results and examples illustrate the indirect role of symmetry and equivalence of nuclei in relation to the symmetry-independent homology groups of a VDW surface. The invariants of these groups serve as tools in the study of the shapes of simple VDWSs and also of more general molecular surfaces.



tional space that lead to essential changes in shape. We must add that the VDWS is a molecular surface for which the shape characterization of arbitrary molecular conformations can be done analytically, as shown here. The results found for a VDWS may also help to understand the general features expected when analyzing other molecular surfaces, such as isodensity contours or contours of electrostatic potentials, important in drug design.

## Acknowledgments

We would like to thank a referee for useful suggestions regarding this paper. This work was supported by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

## References

- [1] R. Langridge, T.E. Ferrin, I.D. Kuntz and M.L. Connolly, *Science* 211(1981)661.
- [2] P.K. Weiner, R. Langridge, J.M. Blaney, R. Schaefer and P.A. Kollman, *Proc. Natl. Acad. Sci. U.S.A.* 79(1982)3754.
- [3] R.J. Feldman, D.H. Bing, B.C. Furie and D. Furie, *Proc. Natl. Acad. Sci. U.S.A.* 75(1978) 5409.
- [4] T. Porter, *Comput. Graphics* 12(1978)282.
- [5] N.L. Max, *Comput. Graphics* 13(1979)165.
- [6] P.A. Basch, N. Pattabiraman, C. Huang, T.E. Ferrin and R. Langridge, *Science* 222(1983) 1325.
- [7] L.H. Pearl and A. Honegger, *J. Mol. Graphics* 1(1983)9.
- [8] M.L. Connolly, *J. Mol. Graphics* 3(1985)19.
- [9] C. Hansch, R.-L. Li, J.M. Blaney and R. Langridge, *J. Med. Chem.* 25(1982)777.
- [10] J.M. Blaney, E.C. Jorgensen, M.L. Connolly, T.E. Ferrin, R. Langridge, S.J. Oatley, J.M. Burridge and C.C.F. Blake, *J. Med. Chem.* 25(1982)785.
- [11] E.H. Spanier, *Algebraic Topology* (McGraw-Hill, New York, 1966).
- [12] J.R. Munkres, *Elements of Algebraic Topology* (Addison-Wesley, Menlo Park, 1984).
- [13] P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.* 12(1986)113.
- [14] P.G. Mezey, *J. Comput. Chem.* 8(1987)462.
- [15] G.A. Arteca and P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.* 14(1987)133.
- [16] P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.* 14(1987)127.
- [17] G.A. Arteca, V.B. Jammal, P.G. Mezey, J.S. Yadav, M.A. Hermsmeier and T.M. Gund, *J. Mol. Graphics* 6(1988)45.
- [18] G.A. Arteca and P.G. Mezey, *J. Comput. Chem.* 9(1988)554.
- [19] G.A. Arteca and P.G. Mezey, *J. Mol. Struct. Theochem.* 166(1988)11.
- [20] G.A. Arteca, V.B. Jammal and P.G. Mezey, *J. Comput. Chem.* 9(1988)608.
- [21] B. Lee and F.M. Richards, *J. Mol. Biol.* 55(1971)379.
- [22] F.M. Richards, *Annu. Rev. Biophys. Bioeng.* 6(1977)151.
- [23] J. Greer and B.L. Bush, *Proc. Natl. Acad. Sci. U.S.A.* 75(1978)303.
- [24] M.L. Connolly, *Science* 221(1983)709.
- [25] P.G. Mezey, *Int. J. Quant. Chem.* 22(1982)101.

- [26] P.G. Mezey, *Mol. Phys.* 47(1982)121.
- [27] P.G. Mezey, *J. Amer. Chem. Soc.* 107(1985)3100.
- [28] P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.* 17(1983)137.
- [29] P.G. Mezey, *Int. J. Quant. Chem.* 26(1984)983.
- [30] P.G. Mezey, *Theor. Chim. Acta* 63(1983)9; *ibid.* 67(1985)43.
- [31] P.G. Mezey, *Potential Energy Hypersurfaces* (Elsevier, Amsterdam, 1987).
- [32] R.C. Weast (ed.), *CRC Handbook of Chemistry and Physics*, 63rd edition (CRC Press, Boca Raton, 1983) p. D-195.
- [33] A. Bondi, *J. Phys. Chem.* 68(1964)441.
- [34] A. Hopfinger, in: *Conformational Properties of Macromolecules* (Academic Press, New York, 1973).
- [35] P.G. Mezey, *J. Math. Chem.* 2(1988)299; 325.