Classifying the conformations of a chemical system using matrices of integers

Peter L. Easthope

*Department of Pathology, University of British Columbia,
2211 Wesbrook Mall, Vancouver, BC, Canada V6T 1Z3*

A chemical system is a collection of atomic nuclei and accompanying electrons interacting by electromagnetic forces and organized into one or more molecules, ions, and transient structures. A general problem is to devise a mathematical representation of the conformation of a system which is useful in further mechanical analysis. Through dihedral rotations of bonds and relative motion of structures, many chemical systems can attain an infinite number of conformations. This paper is a theoretical presentation of a scheme to partition these conformations into a finite number of sets. The scheme involves a discrete aspect of chemical kinematics and uses matrices of integers called proximity matrices. The partitioning scheme is anticipated to be useful in studying reaction mechanisms and interactions between molecules, and in finding conformations of particular interest such as that with a potential energy near the global minimum for a molecule.

1.   Introduction

Computational methods for studying the mechanics of a chemical structure, an interaction between structures, and the mechanism of a reaction comprise an extensive field of research [1–4]. One of these analyses can involve large numbers of conformations [5] of a chemical system. For example, when a reaction occurs there can be relative motion of reactants, a catalyst, and ambient molecules such as water. One or more of these participants undergoes conformational change, and the distribution of electron density alters. Ultimately, the products and other participants assume stable conformations and disperse from the reaction site. Any mechanical analysis of such a system requires that a kinematic [6] model be established before force or energy can be considered.

For a given system, the set of all conformations will be called the conformational space. The number of conformations in a conformational space can be so large that there is currently no practical means of dealing with them all in a calculation. A natural way to resolve this difficulty at the kinematic level is to classify the conformations into subsets, and thereby to break the original problem into multiple

easier problems. This suggests that a scheme for classification of conformations could be a useful fundamental component of chemical kinematics.

Such a classification scheme should satisfy the elementary symmetry requirement that the classification be unaffected by a change in labelling of the nuclei in the system. For example, if the nuclei of a chain molecule are numbered the classification should be unchanged by reversing the numbers on the chain. Also, a reasonable scheme should classify together any two conformations that are similar in form, should cover the whole conformational space accessible to a system, and should provide the basis for construction of or traversal of the conformational space.

No previous works have focused on this problem, although some have marginal relevance to it. For many structures, prominent conformational patterns have been recognized and named. Thus, chain molecules have all cis and all trans conformations, and cyclohexane has chair and boat [7] conformations. Many of the conformational features of proteins determined by X-ray crystallography have descriptive names. For example, triose phosphate isomerase and certain other proteins have a conformation described as a singly wound parallel β barrel [8]. This visual recognition of conformational patterns is not adequate to solve the classification problem being considered because the boundaries of the classes would be indefinite and because there would be far too many conformational classes to identify and name. The D_k procedure [9] is a quantitative analysis yielding a spectrum of parameters representing the difference between two conformations. The spectrum could be used to measure similarity of conformations, but they would not be located in discrete subsets. Furthermore, the justification for using the D_k parameters applies best to a nonbranched chain. A general scheme for classifying conformations should apply to *any* chemical system. An algorithm for recognizing secondary structure in a protein is available [10], but it does not recognize global features; in protein conformation terminology, it does not recognize tertiary or quaternary structure. These existing schemes and analyses are applied to specified conformations, and there is no obvious way to use one of them as the basis for traversing or constructing a conformational space.

Chemical kinematics has been deficient in representing discrete relationships between internuclear distances. Consider, for example, a molecule containing nuclei A, B, and C. There are just three possible relationships between any two of the internuclear distances. The distance from A to B, for example, must be less than, equal to, or greater than the distance from B to C. This observation is the basis for the classification scheme presented.

2.   Theory

Before proceeding with the scheme for classifying conformations, the means of representing them is reviewed. Most chemical systems of more than three nuclei can attain an infinite number of conformations by rotation of dihedral bonds or by relative motion of structures. For a very simple molecule like ethane, the conformation

can be represented in symbolic terms, and problems can be solved directly using algebra and calculus. For a large system, symbolic calculations have been precluded by complexity. The usual way to proceed has been to represent a conformational space numerically by a finite number of conformations. This approach is incomplete if any significant portion of the space is not represented by at least one conformation. Representing other portions of the space to excess will only waste time in evaluation. These observations lead to a consideration of the relative merits of different representations of conformation.

Three mathematical representations of conformation have most frequent use. One is to give the positions of the nuclei in a coordinate system. To do this, the coordinate system must be defined, it must be located relative to the structure, and it must be oriented relative to the structure. Such a molecular coordinate system provides a precise means of specifying a conformation but does not provide the basis for a simple scheme of classification.

Conformation can also be specified by listing all dihedral angles for a molecule. This approach is simple and direct and is effective for small molecules. In large molecules, small adjustments to dihedral angles produce large changes in internuclear distances in regions distant from the adjusted dihedral bond. Consider, for example, a polymer in a long U-shaped conformation. A small adjustment of a dihedral angle near the bend will produce large changes in distances between nuclei at the ends of the molecule distant from the bend. This illustrates that if the conformational space of a molecule is sampled by incrementing dihedral angles, some important portions of the space could be omitted. The angular increment might be decreased to avoid this omission, but conformational changes would then be smaller than necessary when the molecule has a compact form. Thus, the dihedral angle representation of conformation is not perfectly suitable for large molecules. It also does not provide any obvious basis for a satisfactory scheme of classification.

A third specification of conformation is the matrix of distances between nuclei [11]. To construct a distance matrix, the nuclei of a molecular system must be numbered. A conformation of N nuclei is represented by an $N \times N$ matrix, where the element in row i and column j is the distance between nucleus i and nucleus j. The distance matrices for the all cis and all trans conformations of the carbon backbone of pentane are given subsequently.

A useful term in discussing distance matrices is the side diagonal [12]. The ith side diagonal contains each element where the column index equals the row index $+ i$. For example, in the matrix

$$\begin{pmatrix} 0 & g & v & d \\ g & 0 & g & v \\ v & g & 0 & g \\ d & v & g & 0 \end{pmatrix},$$

the first, second, and third side diagonals are comprised of the elements g, v, and d, respectively. If the matrix represents a chain with nuclei indexed consecutively from one end, the elements g, v, and d are geminal, vicinal, and dihedral distances.

Consider the distance specification applied to the carbon backbone of pentane. Two of an infinite number of conformations are illustrated in fig. 1. The geminal distances or bond lengths are all assumed to be equal in this example and are normalized to 1. The three vicinal distances are also assumed to be equal. Geminal

(a)   (b)

1   5   1   3   5

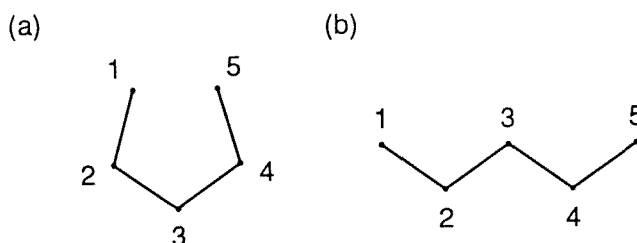2   4   2   4

3

Fig. 1. Two planar conformations for the carbon backbone
of pentane. $d_{12} = d_{23} = d_{34} = d_{45} = 1.0$; $d_{13} = d_{24} = d_{35} = 1.633$.
(a) All cis conformation: $d_{15} = 1.089$, $d_{14} = d_{25} = 1.667$.
(b) All trans conformation: $d_{15} = 3.266$, $d_{14} = d_{25} = 2.517$.

and vicinal distances are nearly constant for a given structure as conformation changes. On the other hand, each dihedral distance is directly related to a dihedral angle [13]. The two dihedral angles differ from fig. 1(a) to fig. 1(b), so the corresponding dihedral distances also differ. The $N - 1$ geminal distances, $N - 2$ vicinal angles, and $N - 3$ dihedral angles of an unbranched chain molecule completely determine its conformation. These bond lengths and vicinal and dihedral angles are equivalent to the $3N - 6$ distances of the first, second, and third side diagonals of the distance matrix. Consequently, the conformation of this structure is completely determined by these side diagonals. Nevertheless, the distances beyond the third side diagonal are useful because they show directly the relative position of nuclei that are not close neighbors in structure. The distance matrices for the all cis and all trans conformations of pentane are

$$\begin{pmatrix} 0.00 & 1.00 & 1.63 & 1.67 & 1.09 \\ 1.00 & 0.00 & 1.00 & 1.63 & 1.67 \\ 1.63 & 1.00 & 0.00 & 1.00 & 1.63 \\ 1.67 & 1.63 & 1.00 & 0.00 & 1.00 \\ 1.09 & 1.67 & 1.63 & 1.00 & 0.00 \end{pmatrix}$$

and

$$
\begin{pmatrix}
0.00 & 1.00 & 1.63 & 2.52 & 3.27 \\
1.00 & 0.00 & 1.00 & 1.63 & 2.52 \\
1.63 & 1.00 & 0.00 & 1.00 & 1.63 \\
2.52 & 1.63 & 1.00 & 0.00 & 1.00 \\
3.27 & 2.52 & 1.63 & 1.00 & 0.00
\end{pmatrix}.
$$

A classification of the conformations of a chemical system can be thought of as a partition [14] of the set of distance matrices representing the conformational space. Although the structure of most chemical systems permits an infinite number of conformations, the proposed scheme yields a finite number of subsets containing all the distance matrices (or all conformations) for any given system. The number of subsets in the partition being finite will be advantageous in some applications. This number increases with the number of nuclei in the system or elements in the distance matrix.

2.1. DEFINITION OF THE PARTITION OF PROXIMITY MATRICES

Partitioning is achieved by defining a many-to-one correspondence between the set of all distance matrices for a structure and a finite set of matrices, which will be called proximity matrices.[*] Because there will be a unique proximity matrix corresponding to every distance matrix, the correspondence is a function. Each subset of conformations, or member of the partition of distance matrices, is identified by a single proximity matrix according to this function.

A proximity matrix is essentially just an enumeration of the elements of a distance matrix in increasing order. By definition, a proximity matrix has the same dimension N as a corresponding distance matrix and has integer elements. Like the distance matrix, the proximity matrix is symmetric and has zeroes on the diagonal. If M is the number of nonzero distinct distances in a distance matrix and the distance from nucleus i to nucleus j is d_{ij}, then the numerical order of distances can be represented as

$$
0 < d_{(ij)_1} < d_{(ij)_2} \leq \ldots \leq d_{(ij)_s} \leq d_{(ij)_t} \leq \ldots \leq d_{(ij)_M}. \tag{1}
$$

The ij pairs are indexed with the whole numbers $1, 2, \ldots, s, t, \ldots, M$ such that $t = s$ if $d_{(ij)_s} = d_{(ij)_t}$, whereas $t = s + 1$ if $d_{(ij)_s} < d_{(ij)_t}$. Thus, if one or more equalities are present in (1), then $M < N(N-1)/2$. If there are no equalities, then $M = N(N-1)/2$. The proximity element p_{ij} corresponding to distance element d_{ij} is defined as s or t in the ordered sequence. This defines uniquely the function from distance matrices to proximity matrices. A computer language implementation is given in appendix A.

[*]Another application of distance geometry is the theory of multidimensional scaling (see Kruskal [15]). The term *proximity* has been used in multidimensional scaling with a meaning different to that in this paper (see Shepard [15]).
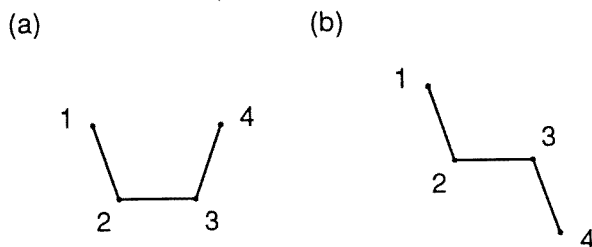
Fig. 2. (a) The cis conformation of the butane back-
bone. (b) The trans conformation of the butane backbone.

The planar cis conformation of the butane carbon backbone is illustrated in
fig. 2(a). For simplicity, all C–C bonds are assumed to have unit length. The
distance matrix for the cis conformation is

$$\begin{pmatrix} 0.00 & 1.00 & 1.63 & 1.67 \\ 1.00 & 0.00 & 1.00 & 1.63 \\ 1.63 & 1.00 & 0.00 & 1.00 \\ 1.67 & 1.63 & 1.00 & 0.00 \end{pmatrix}.$$

Note that the planar cis conformation minimizes d_{14} for the structure. The elements
of the matrix are ordered

$$0 < d_{(12)_1} = d_{(23)_1} = d_{(34)_1} < d_{(13)_2} = d_{(24)_2} < d_{(14)_3}.$$

Thus, the proximity matrix for the conformation is

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}.$$

In the planar trans conformation illustrated in fig. 2(b), the butane backbone has the
distance matrix

$$\begin{pmatrix} 0.00 & 1.00 & 1.63 & 2.52 \\ 1.00 & 0.00 & 1.00 & 1.63 \\ 1.63 & 1.00 & 0.00 & 1.00 \\ 2.52 & 1.63 & 1.00 & 0.00 \end{pmatrix}.$$

This conformation maximizes d_{14}, and aside from the increase in this element along with d_{41}, the matrix is identical to the distance matrix for the cis conformation. Expression (1) is identical for both conformations and for all other conformations of the molecule. There are an infinite number of distance matrices, but only one proximity matrix for this structure. The pentane backbone with just one more nucleus has several proximity matrices.

In the most general case where all parameters of a system including bond lengths and vicinal angles are variable, a distance matrix for a system of N nuclei has $L = N(N-1)/2$ variable elements. In a corresponding proximity matrix, each element must be one of the integers $\{1, 2, \ldots, L\}$. Consequently, there can be no more than L^L proximity matrices for a system of N nuclei. If the integer j is an element of a proximity matrix then, according to the definition, all of $\{1, 2, \ldots, j-1\}$ must also be present. For example, a matrix with every element off the main diagonal equal to 3 is not a valid proximity matrix. The number L^L includes such invalid matrices, so it is greater than the number of proximity matrices corresponding to attainable conformations. For $N = 10$, this upper bound on the number of proximity matrices is $45^{45} \doteq 2.4 \times 10^{74}$. Any computation involving such a large number of matrices is obviously infeasible. The refinements considered next restrict the set of proximity matrices for a given system.

2.2. CRITERIA FOR FEASIBILITY OF PROXIMITY MATRICES

The problem of whether a given distance matrix or set of lower and upper bounds on distances can be realized in a conformation in three-dimensional space, called the embeddability problem, has been illuminated by extensive work [16]. The proximity matrix corresponding to a distance matrix that is embeddable can be described as feasible. In some analyses, a candidate proximity matrix for a system will be available without a corresponding distance matrix, and a criterion of feasibility that does not resort to finding an embeddable distance matrix would be useful. Three such criteria of feasibility are described here. While these criteria are weak so that a significant proportion of proximity matrices they admit will probably not have embeddable distance matrices, they are also simpler to evaluate than the embeddability criteria. In some calculations, the application of feasibility criteria prior to embeddability criteria should prove beneficial.

2.2.1. Criteria of bond lengths and angles

Once the chemical structure of a molecule is established, some information about its proximity matrices can be obtained directly from bond lengths and vicinal angles, independent of conformation. With few exceptions, the proximity elements for pairs of atoms that are bonded in an organic molecule can be determined directly from bond radii. This is because the exposed van der Waals radius of most atoms is larger than the largest bond radius [17], and consequently all proximity elements

for non-bonded atom pairs are usually larger than any proximity element for a bonded pair. Consider pentane: the covalent bond radii for carbon, 0.077 nm, and hydrogen, 0.03 nm, are both smaller than the exposed radii, 0.17 and 0.11 nm. Therefore, all elements of the proximity matrix corresponding to C−H bonds must be 1 and all elements corresponding to C−C bonds must be 2.

This reasoning for first neighbors can be extended to second and more distant neighbors in chemical structure. In many structures, the distance between two nuclei separated by six or more bonds can be reduced until it is comparable to a bond length and van der Waals repulsion takes effect. For example, a chain of six carbon atoms attains this condition as the all cis conformation, where $d_{16} \doteq (0.17/0.077)d_{12}$ = 2.21 d_{12}, is approached. Atoms closer together in the chain are limited in their minimum separation by the geometry of the lengths and vicinal angles of bonds. The chain of five carbon nuclei illustrated in fig. 1 provides an example. Again, nuclei are numbered consecutively from one end of the chain to simplify the discussion, but a different numbering would yield equivalent results. The hydrogen atoms are ignored so that the principle can be illustrated with a single bond length and an unbranched chain. The geminal (d_{12}, d_{23}, d_{34}, d_{45}) and vicinal (d_{13}, d_{24}, d_{35}) distances are uniquely determined by bond lengths and atomic configurations; $d_{13} = d_{24} = d_{35}$ = $2(2/3)^{0.5}d_{12}$ = 1.63 d_{12}. The dihedral distances (d_{14}, d_{25}) are minimized in the all cis conformation illustrated in fig. 1(a), where $d_{14} = d_{25} = 5d_{12}/3 = 1.67\ d_{12}$, and maximized in the all trans conformation illustrated in fig. 1(b), where $d_{14} = d_{25}$ = $(19/3)^{0.5}d_{12}$ = 2.52 d_{12}. In the all cis conformation $d_{15} = (4/3)(2/3)^{0.5}$ = 1.09 d_{12}, and in the all trans conformation $d_{15} = 4(2/3)^{0.5}d_{12}$ = 3.27 d_{12}. Hence, for this example and other ordinarily occurring bond lengths and angles, the dihedral distance (never smaller than 1.67 d_{12}) and proximity are always greater than the vicinal distance (1.63 d_{12}) and proximity, respectively. In other words, d_{14} must be greater than d_{13} and p_{14} must be greater than p_{13}. On the other hand, d_{15} can be less than, equal to, or greater than a vicinal or dihedral distance, and consequently p_{15} can be less than, equal to, or greater than p_{13}, p_{24}, p_{35}, p_{14}, or p_{25}. To summarize, for a chain of five carbon nuclei, all geminal proximity elements are 1. Vicinal elements are 2 if $d_{12} \leq d_{15}$ and 3 if $d_{12} > d_{15}$.

In constructing a set of proximity matrices for a given molecule, the criteria of bond lengths and angles just described are significant for nuclei that are near neighbors in structure. With these criteria, the bound on the number of proximity matrices for a given structure can be improved. For a chain molecule, a minimum of $N-1$ proximity elements will usually be determined directly by structure. This leaves $(N-1)(N-2)/2$ variable elements in the distance matrix. Assuming the proximity matrix has the same number of variable elements, the upper bound on the number of possible proximity matrices for a molecule becomes

$$((N-1)(N-2)/2)^{((N-1)(N-2)/2)}.$$

The distance matrices for the all cis and all trans conformations of the pentane backbone have been given, and the corresponding proximity matrices are

$$
\begin{pmatrix}
0 & 1 & 3 & 4 & 2 \\
1 & 0 & 1 & 3 & 4 \\
3 & 1 & 0 & 1 & 3 \\
4 & 3 & 1 & 0 & 1 \\
2 & 4 & 3 & 1 & 0
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
0 & 1 & 2 & 3 & 4 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
4 & 3 & 2 & 1 & 0
\end{pmatrix}.
$$

For this structure, all vicinal distances, in addition to geminal distances, are uniquely determined. This leaves three variable distances (d_{14}, d_{15}, and d_{25}) and several proximity matrices are possible. Constraints of bond lengths and angles require that d_{14} and d_{25} be greater than d_{13} ($= d_{24} = d_{35}$), but d_{15} can be less than, equal to, or greater than d_{13}. For any proximity matrix for this structure, this implies that $p_{14} > p_{13} = p_{24} = p_{35}$, that $p_{25} > p_{13}$, and that p_{15} is less than, equal to, or greater than p_{13}. If $p_{15} < p_{13}$, two matrices are possible in addition to the all cis matrix:

$$
\begin{pmatrix}
0 & 1 & 3 & 4 & 2 \\
1 & 0 & 1 & 3 & 5 \\
3 & 1 & 0 & 1 & 3 \\
4 & 3 & 1 & 0 & 1 \\
2 & 5 & 3 & 1 & 0
\end{pmatrix},
\quad
\begin{pmatrix}
0 & 1 & 3 & 5 & 2 \\
1 & 0 & 1 & 3 & 4 \\
3 & 1 & 0 & 1 & 3 \\
5 & 3 & 1 & 0 & 1 \\
2 & 4 & 3 & 1 & 0
\end{pmatrix}.
$$

If $p_{15} = p_{13}$, three more matrices are possible:

$$
\begin{pmatrix}
0 & 1 & 2 & 3 & 2 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
2 & 3 & 2 & 1 & 0
\end{pmatrix},
\quad
\begin{pmatrix}
0 & 1 & 2 & 3 & 2 \\
1 & 0 & 1 & 2 & 4 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
2 & 4 & 2 & 1 & 0
\end{pmatrix},
\quad
\begin{pmatrix}
0 & 1 & 2 & 4 & 2 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
4 & 2 & 1 & 0 & 1 \\
2 & 3 & 2 & 1 & 0
\end{pmatrix}.
$$

If $p_{15} > p_{13}$ and at least two of p_{14}, p_{15}, and p_{25} are equal, there are six possibilities in addition to the matrix for the all trans conformation:

$$
\begin{pmatrix}
0 & 1 & 2 & 3 & 3 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
3 & 3 & 2 & 1 & 0
\end{pmatrix},
\quad
\begin{pmatrix}
0 & 1 & 2 & 3 & 3 \\
1 & 0 & 1 & 2 & 4 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
3 & 4 & 2 & 1 & 0
\end{pmatrix},
\quad
\begin{pmatrix}
0 & 1 & 2 & 4 & 3 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
4 & 2 & 1 & 0 & 1 \\
3 & 3 & 2 & 1 & 0
\end{pmatrix},
$$

$$
\begin{pmatrix}
0 & 1 & 2 & 4 & 3 \\
1 & 0 & 1 & 2 & 4 \\
2 & 1 & 0 & 1 & 2 \\
4 & 2 & 1 & 0 & 1 \\
3 & 4 & 2 & 1 & 0
\end{pmatrix},
\quad
\begin{pmatrix}
0 & 1 & 2 & 4 & 4 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
4 & 2 & 1 & 0 & 1 \\
4 & 3 & 2 & 1 & 0
\end{pmatrix},
\quad
\begin{pmatrix}
0 & 1 & 2 & 3 & 4 \\
1 & 0 & 1 & 2 & 4 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
4 & 4 & 2 & 1 & 0
\end{pmatrix}.
$$

Finally, if $p_{15} > p_{13}$ and all of p_{14}, p_{15}, and p_{25} are distinct, there are another six possibilities:

$$
\begin{pmatrix}
0 & 1 & 2 & 3 & 5 \\
1 & 0 & 1 & 2 & 4 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
5 & 4 & 2 & 1 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 1 & 2 & 4 & 5 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
4 & 2 & 1 & 0 & 1 \\
5 & 3 & 2 & 1 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 1 & 2 & 3 & 4 \\
1 & 0 & 1 & 2 & 5 \\
2 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 0 & 1 \\
4 & 5 & 2 & 1 & 0
\end{pmatrix},
$$

$$
\begin{pmatrix}
0 & 1 & 2 & 5 & 4 \\
1 & 0 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 2 \\
5 & 2 & 1 & 0 & 1 \\
4 & 3 & 2 & 1 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 1 & 2 & 4 & 3 \\
1 & 0 & 1 & 2 & 5 \\
2 & 1 & 0 & 1 & 2 \\
4 & 2 & 1 & 0 & 1 \\
3 & 5 & 2 & 1 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 1 & 2 & 5 & 3 \\
1 & 0 & 1 & 2 & 4 \\
2 & 1 & 0 & 1 & 2 \\
5 & 2 & 1 & 0 & 1 \\
3 & 4 & 2 & 1 & 0
\end{pmatrix}.
$$

These matrices have been constructed by selecting elements from $\{1, 2, 3, 4, 5\}$ and by permuting positions in the matrix. Any matrix is excluded if a member of $\{1, 2, 3, 4, 5\}$ between the smallest and largest elements of the matrix is absent or if a criterion of bond lengths and angles is violated. The pentane backbone has nineteen proximity matrices, and this number is substantially less than the bound of $((N-1)(N-2)/2)^{((N-1)(N-2)/2)} = 6^6 = 46656$.

2.2.2. Criteria of the face centered cubic lattice

A proof that the fcc lattice has the highest density of all packings of equal spheres has recently been announced [18]. This lattice probably also maximizes the number of equal distances between nodes.

(a) *Criterion of equal distances in an fcc lattice.* For a given number of nodes, this criterion establishes an upper limit on the number of equal elements in a proximity matrix or distance matrix. Every node in the interior of the fcc lattice is impinged by twelve edges of equal length. Since each edge is shared between two nodes, the interior of the lattice has six edges per node. Boundary nodes are impinged by fewer than twelve edges, so any finite lattice has no more than six edges per node in average.

The number of nodes may permit the boundary of the finite lattice to be a cubo-octahedron. For example, the smaller cubo-octahedron in fig. 3 contains thirteen nodes, but no cubo-octahedrally bounded fcc lattice can be constructed from eleven nodes. The cubo-octahedral boundary approximates a sphere and therefore minimizes the number of surface nodes. Consequently, the cubo-octahedron will maximize the number of equal edges over all possible boundaries. The size of a lattice bounded by a cubo-octahedron can be measured by the number of nodes counted along a series of edges coinciding with a straight segment from the center to the surface.
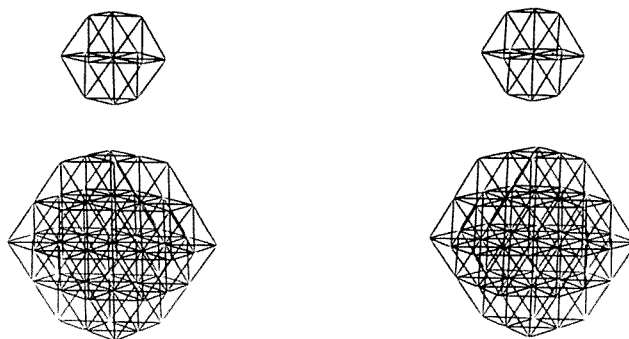
Fig. 3. The fcc lattices with cubo-octahedral boundaries; $r = 2$ and $r = 3$.

This number of nodes will be denoted r by analogy to the radius of a sphere, but the central node is $r = 1$ rather than $r = 0$. The $r = 2$ and $r = 3$ cases are illustrated in fig. 3. The number of nodes in terms of r, denoted $n(r)$, and the number of equal edges in terms of r, denoted $e(r)$, are calculated in appendix B. For $r \leq 3$, these numbers can be verified using a model (half the model in fig. 3 is adequate); for larger r, a model is not convenient. Table 1 lists $n(r)$ and $e(r)$ together with the number of elements above the diagonal in the distance matrix corresponding to $n(r)$ nodes, $n(r)(n(r) - 1)/2$, for r up to ten.

Table 1

Numbers of nodes, edges, and distances in an fcc lattice with a cubo-octahedral boundary.

r	$n(r)$	$e(r)$	$n(r)(n(r) - 1)/2$
1	1	0	0
2	13	36	78
3	55	216	1485
4	147	660	10731
5	309	1488	47586
6	561	2820	157080
7	923	4776	425503
8	1415	7476	1000405
9	2057	11040	2114596
10	2869	15588	4114146

For most systems, the total number of nuclei will not coincide with a number $n(r)$ of nodes in a cubo-octahedrally bounded lattice; whereas the criterion must apply to any chemical system. Rather than attempt to find the non-cubo-octahedral boundaries that maximize the number of equal edges for an arbitrary number of

nodes in a finite fcc lattice, an upper bound on the number of edges will be calculated from the values of $e(r)$ in table 1. When a node is added to an fcc lattice, the number of edges per node added never exceeds six. Therefore, an upper bound on the number of edges is obtained by extrapolation at the rate of six edges per node from the largest complete cubo-octahedron constructable with the given number of nodes. Obviously, the bound must be less than the number of edges in the next larger cubo-octahedron. For example, 17 nodes allows construction of a cubo-octahedron of $r = 2$ with four nodes extra ($n(2) = 13$). Thus, the upper bound on the maximum number of equal edges is

$$e(2) + 6(17 - n(2)) = 36 + 6 \times 4 = 60.$$

This upper bound does not exceed $e(3)$, which is 216. Now for 51 nodes,

$$e(2) + 6(51 - n(2)) = 36 + 6 \times 38 = 264 > e(3),$$

so the upper bound is 216. This calculation can be summarized as follows. The number of nodes is denoted by N and the upper bound on the number of equal edges is denoted by E. If $N = n(r)$ for some integer r, then a complete cubo-octahedron containing $e(r)$ edges can be assembled; else, extrapolation as described provides E.

(b) *Criterion of bond lengths and atomic volumes in an fcc lattice.* This criterion refers to elements in a tentative proximity matrix. Consider an arbitrary element p_{ij} and other elements in row i. If the criterion is violated, then the elements considered for the ith row are excluded from possibility.

Let p_{ik} be any element in the row containing p_{ij}. Let K be the set of column indices k where the inequality $p_{ik} \leq p_{ij}$ is satisfied. The definition of the proximity matrix yields the theorem: $p_{ik} \leq p_{ij}$ implies $d_{ik} \leq d_{ij}$. Now let S be the sphere centered at nucleus i and having radius d_{ij} plus the radius of the largest atom in K. According to the theorem just mentioned, all the atoms of K are contained in S. The bond lengths and vicinal angles connecting i and j imply a maximum value of d_{ij}, which is calculated in appendix C. If this maximum is denoted m_{ij} and the maximum radius of atoms in K is r_K, then the volume of S is $4\pi(m_{ij} + r_K)^3/3$. On the other hand, the bulk volume for the atoms in K can be approximated by summing the volumes of van der Waals spheres and dividing by the ratio of the volume of the spheres in an fcc lattice to the total volume of the lattice. The ratio [19] is 0.7405 (ref. [19] gives $\Lambda_3 = 0.74048$) and accounts for the interstitial volume. The structure is not exactly represented by uniform spheres contacting at points, but this approximation is made to simplify the calculation. The total volume, calculated in this way from the atoms in K, is denoted by V. The criterion of bond lengths and atomic volumes is the statement: V is within the volume of S; symbolically, $V \leq 4\pi(m_{ij} + r_K)^3/3$. If this relation is not satisfied, a tentative proximity matrix with the elements of K in the ith row cannot be realized. Butane can be used to illustrate the application of this criterion. We already know that

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}$$

is a valid proximity matrix for the butane backbone, so the criterion should not reject any elements in this matrix. We will apply the criterion to row 1 with $p_{ij} = p_{14} = 3$. $K = \{1, 2, 3, 4\}$ because p_{11}, p_{12}, and p_{13} are all less than p_{14}. Figure 4 illustrates a section of the sphere S through the center. Again, the length of the C—C bond is taken to be 1.0; r_K is then the ratio of the exposed radius for
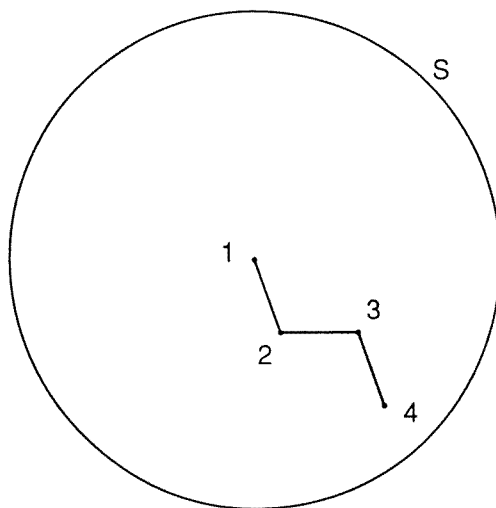

Fig. 4. Section of S through the center for $p_{ij} = p_{14}$.

a covalently bonded carbon atom to the C—C bond length [20]. The volume of S is $4\pi(m_{14} + r_K)^3/3 = 4\pi(2.52 + 1.10)^3/3 = 199$. For purposes of volume calculation, the surface of a carbon atom is a sphere of radius 1.10 with a spherical cap of height 0.6 removed for each bond. Thus, atoms 1 and 4 each have a volume of $5.58 - 1.02$, and atoms 2 and 3 have a volume of $5.58 - (2 \times 1.02)$. The total bulk volume is approximately $16.2/0.7405 = 21.9$. Now $21.9 \leq 199$, so the criterion is satisfied by a wide margin. This result is obvious in fig. 4 because the atoms occupy a small portion of S. When the criterion is applied in a system where S contains many tightly packed atoms, the result will not be trivial.

For proximity element p_{ij}, this criterion can be evaluated using any subset of the other elements in row i or in column j. The criterion can be applied to a matrix where some elements in the row or column used are not yet determined. This advantage weighs against the fact that the criterion is less likely to fail when fewer atoms are available for inclusion in the set K.

3. Discussion

3.1. TRAVERSING SETS OF CONFORMATIONS

A central problem in chemical kinematics is to generate or traverse a subset of the conformational space of a given system. The concept of a proximity matrix and the feasibility criteria provide a new approach to the solution of this problem. Completion of the solution requires an algorithm to construct proximity matrices and to construct distance matrices for each proximity matrix. Following the description of the criteria of bond lengths and angles, the number L^L was calculated as a weak upper bound on the number of proximity matrices. A refinement of this upper bound incorporating the criteria of the face centered cubic lattice is relevant to questions of computation time but cannot be easily performed. This number appears to be obtainable only by constructing the proximity matrices.

Generation of proximity matrices is essentially a process of permutation of integers and application of the feasibility criteria. Given the structure of a chain of N atoms joined by rotatable bonds, only $N - 3$ parameters are required to determine a conformation. If a set of conformations is generated as a sequence of distance matrices, only $N - 3$ elements in a distance matrix need be adjusted directly to alter the conformation, although the remaining distances must still be calculated to check embeddability. All permutations of more than about twenty distinct objects cannot be generated in a reasonable time [21], but the possibility of traversing a set of feasible proximity matrices by a combinatorial algorithm [22] remains. The task of developing such an algorithm is an intriguing challenge.

3.2. CONFORMATIONS OF GLOBAL MINIMUM ENERGY

Optimization algorithms that find a local minimum near a starting point are now known to be ineffective in finding the global minimum of a conformational energy function, in the general case, because of the existence of a large number of local minima [23,24]. The problem has a continuous nonlinear aspect [25], and the existence of proximity matrices shows that a combinatorial aspect is also present. If the number of local minima of energy over all conformations corresponding to one proximity matrix can be proven to be small or perhaps even one, then finding the smallest minimum for a given proximity matrix would be a relatively easy problem. Finding the global minimum of energy over the whole conformational space would then require a search through a finite number of proximity matrices.

3.3. CONCURRENCY IN COMPUTATION

At present, many problems for macromolecular systems cannot be solved because computation time is excessive [26]. Computation time can be reduced if the problem can be divided into smaller problems that can be solved concurrently. Proximity matrices provide a means for dividing some problems in chemical kinematics.

For example, to find an energy minimum a master processor would traverse a set of proximity matrices and delegate them one or more at a time to subsidiary processors. One subsidiary processor would take a proximity matrix and evaluate the optimal distance matrix for it using a relatively simple algorithm. This process would be carried out for different proximity matrices concurrently and independently by the subsidiary processors. When a subsidiary processor finishes evaluating the proximity matrices delegated to it, the minimum energy and corresponding proximity matrix would be returned to the master processor. The master processor would record the conformational energy that is minimal at the global level and the corresponding proximity matrix. After traversing all desired proximity matrices for the molecule, the minimum energy and corresponding optimal proximity matrix would be known. The corresponding conformation would be obtained by repeating the evaluation of the optimal distance matrix for the optimal proximity matrix.

4.   Conclusion

The enormous complexity of the kinematics of a large chemical system has motivated development of this scheme for classifying conformations. General criteria based on lengths and vicinal angles of bonds and on the limiting properties of face centered cubic lattices characterize the proximity matrices used in the classification. Criteria of feasibility stronger than those described may be possible. The classification scheme together with the feasibility criteria provide a conceptual basis for an algorithm that will generate conformations. If developed, this algorithm will allow a new approach to certain problems in computational chemistry.

Appendix A: The distance matrix to proximity matrix function

The following are extended and direct definitions of the function from distance matrices to proximity matrices in the computer languages APL and J [27].

```
      ∇ proximities←DIST2PROX distances;d;⎕IO
[1]    ⍝ Calculate the array of proximities for a given array
[2]    ⍝ of distances. The distances array can have any shape.
[3]    ⍝ ----------------------------------------------------------
[4]       ⎕IO←0
[5]    ⍝ Ravel distances. Append 0 to cover the case where
[6]    ⍝ the distance matrix is represented by a vector.
[7]       d←0,,distances
[8]    ⍝ Collect distinct elements:
[9]       d←((d⍳d) = ⍳⍴d)/d
[10]   ⍝ Sort them (⍋ denotes grade up):
[11]      d←d[⍋d]
```

[12] ⍴ Construct the proximity array:
[13]   proximities←(⍴distances)⍴dι,distances
     ∇

D2P: (⍴⍵)⍴d[⍙ d←((dιd) = ι⍴d)/d←(□IO←0),,⍵]ι,⍵
d2p=. (({~/:)@~.@(O&,)@,)i.]

Appendix B: Numbers of nodes and edges in a lattice of closed packed spheres with a cubo-octahedral boundary

This appendix considers a finite lattice in three-dimensional space comprised of nodes and connecting edges conforming to an fcc lattice with a cubo-octahedral boundary. A toothpick model of at least half an $r = 3$ cubo-octahedron like that in fig. 3 is helpful in following the derivation.

Inspection of the model reveals that the cubo-octahedron is composed of eight regular tetrahedrons and six pyramids. The center of the cubo-octahedron is a vertex of each tetrahedron and the vertex opposite the square base of each pyramid. The faces of the cubo-octahedron consist of a triangular face of each tetrahedron and the square base of each pyramid. The numbers of nodes and edges are calculated by summing over the tetrahedrons and pyramids and by counting only once shared nodes and edges where sectors are adjacent.

B.1.   NODES IN A CUBO-OCTAHEDRON

(a) *Tetrahedrons.* An equilateral triangular face of the cubo-octahedron is a planar lattice of equilateral triangles. If the face has j nodes along one side, it contains $j(j + 1)/2$ nodes in total. The number of nodes in the tetrahedron is

$$\sum_{j=1}^{r} j(j + 1)/2 = r(r + 1)(r + 2)/6.$$

Each tetrahedron shares r nodes with each of three tetrahedrons adjacent to it, and the eight tetrahedrons share the central node. The number of nodes in the eight tetrahedrons is

$$8 \times r(r + 1)(r + 2)/6 - 8 \times 3(r - 1)/2 - 7. \tag{B.1}$$

The second term is half of the number of shared nodes. The last term provides that the center is counted just once.

(b) *Pyramids.* A calculation like that for the tetrahedrons gives the number of nodes in a pyramid:

$$r(r + 1)(2r + 1)/6. \tag{B.2}$$

The nodes in the triangular faces of each pyramid must be excluded because they are counted in the tetrahedrons. The number of nodes in the four triangular faces is

$$4(r(r + 1)/2 - r) + 1 = 2r(r - 1) + 1. \tag{B.3}$$

By subtracting (B.3) from (B.2), the number of nodes in a pyramid, excluding triangular faces, is

$$r(r + 1)(r + 2)/6 - 2r(r - 1) - 1. \tag{B.4}$$

Adding (B.1) and (B.4) gives the number of nodes in the cubo-octahedron:

$$
\begin{aligned}
n(r) &= \text{(nodes in 8 tetrahedra)} + \text{(nodes in 6 pyramids)} \\
&= 8r(r + 1)(r + 2)/6 - 12(r - 1) - 7 + 6(r(r + 1)(2r + 1)/6 - 2r(r - 1) - 1) \\
&= (10r^3 - 15r^2 + 11r - 3)/3. \tag{B.5}
\end{aligned}
$$

B.2.   EDGES IN A CUBO-OCTAHEDRON

(a) *Tetrahedrons.* The total number of edges in a tetrahedron is

$$\sum_{i=1}^{r} (3(i - 1)i/2 + 3(i - 1)i/2) = (r - 1)r(r + 1).$$

The first of the two terms of the summand is the number of edges connecting layer $i - 1$ to layer i (i.e. $3 \times$ (number of nodes in layer i)), and the second term is the number of edges in layer i. The number of edges in eight tetrahedrons is

$$8(r - 1)r(r + 1) - 8 \times 3(r - 1)/2 = 8(r - 1)r(r + 1) - 12(r - 1). \tag{B.6}$$

The second term is for the edges shared by pairs of tetrahedons.

(b) *Pyramids.* The number of edges in a two-dimensional square lattice with i nodes on one side is $2i(i - 1)$. The total number of edges in a pyramid is

$$\sum_{\substack{\text{layers from vertex to base}}} \text{(edges in a layer)} = \sum_{i=1}^{r} 4(i - 1)^2 + 2i(i - 1)^2 = 2r(r - 1)^2.$$

The first of the two terms in the summand is the number of edges joining layer $i - 1$ to layer i, and the second term is the number of edges in layer i. The number of edges on one triangular face of a pyramid is $3r(r - 1)$. Two triangular faces of a pyramid that are adjacent share $r - 1$ edges that must be counted only once. Therefore, the total number of edges on all four triangular surfaces is $4(3 \times r(r - 1)/2 - (r - 1))$, which simplifies to $2(3r^2 - 5r + 2)$. Combining the two results gives the number of edges in a pyramid excluding triangular faces:

$$2r^2(r-1) - 2(3r^2 - 5r + 2) = 2(r^3 - 4r^2 + 5r - 2). \tag{B.7}$$

The total number of edges in a cubo-octahedron $e(r)$ is expression (B.6) plus six times expression (B.7):

$$e(r) = 8(r-1)r(r+1) - 12(r-1) + 6 \times 2(r^3 - 4r^2 + 5r - 2)$$
$$= 4(5r^3 - 12r^2 + 10r - 3). \tag{B.8}$$

Appendix C: Maximum distance between connected atoms

This appendix addresses the problem of calculating the maximum distance between two atoms connected by a chain of intervening atoms. Initially, there is assumed to be only one path connecting the two atoms of interest, and the atoms are numbered 1 through n for this discussion.

In the simplest case, all intervening bond lengths are equal, and all intervening vicinal angles are equal. The maximum distance from atom 1 to atom n occurs when all the atoms lie in a plane with all dihedral bonds in the trans conformation. This is illustrated in figs. 5(a) and 5(b). The maximum distance is easily calculated. For odd n (fig. 5(a)),

$$m_{1n} = (n \operatorname{div} 2) d_{i\ i+2},$$

and for even n (fig. 5(b)),

$$m_{1n} = \left[(n \operatorname{div} 2)^2 - (n \operatorname{div} 2) + \left(\frac{d_{i\ i+1}}{d_{i\ i+2}} \right)^2 \right]^{1/2} d_{i\ i+2}.$$

The notations div and mod refer to integer division. If $n = 2k$ or $n = 2k + 1$ (k a natural number), the operation div is defined so that $n \operatorname{div} 2 = k$; mod is defined so that $n \operatorname{mod} 2 = n - 2k$.

The more general case of various bond lengths and vicinal angles occurs in many molecules of interest. This is illustrated for a hypothetical molecule in fig. 5(c). m_{1n} is attained in the planar all trans conformation shown. In general, a tetrangle is a simplex of four vertices having zero volume. Therefore, if one diagonal and four sides of a tetrangle are known, the other diagonal can be obtained by solving an equation stating that the volume of the simplex [28] is zero. The simplex is specified by a distance matrix, so the square of the volume is calculated by a Cayley–Menger determinant. The determinant reduces to a quadratic expression in the square of the unknown distance, which is obtained as the largest root of the expression [29]. This calculation can be applied to successively larger tetrangles until the required distance is obtained. For example, consider that d_{16} in fig. 5(c) is required. d_{14} can be evaluated if d_{12}, d_{24}, d_{43}, d_{31}, and d_{23} are known. Similarly, d_{36} can be calculated. Then d_{16} can be calculated by considering the tetrangle 1463.
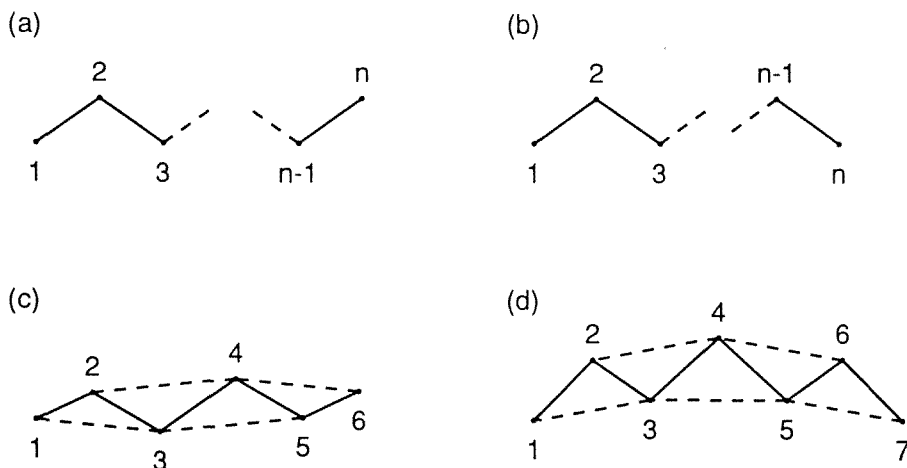
(a)

(b)

(c)

(d)

Fig. 5. (a) An all trans conformation with uniform bond lengths and vicinal angles for *n* odd. (b) An all trans conformation with uniform bond lengths and vicinal angles for *n* even. (c) An all trans conformation with non-uniform bond lengths and vicinal angles. (d) An all trans conformation with a nonconvex atom polygon.

Figure 5(d) illustrates another aspect of this maximum distance problem. Suppose the maximum value for the distance d_{17} is required. By suitable adjustment of the dihedral angles θ_{23}, θ_{34}, θ_{45}, θ_{56} with nuclei 2, 4, and 6 moving out of the plane of the figure, nuclei 1, 3, 5, and 7 can be made collinear. The maximum of d_{17} in this case is simply

$$m_{17} = d_{13} + d_{35} + d_{57}.$$

A nonplanar conformation has resulted in a simplification of the maximum distance expression. This result can be convincingly demonstrated if fig. 5(d) is traced onto cardboard and cut out along the segments joining 1, 3, 5, 7, 6, 4, 2, and 1. A crease along each dihedral bond will provide the required flexibility. Manipulating the cardboard strip while keeping tension between points 1 and 7 will reveal an infinite number of conformations where 1, 3, 5, and 7 are collinear.

The property distinguishing the structure of fig. 5(d) from those of figs. 5(a), 5(b), and 5(c) involves the concept of concavity. Any planar chain has an enclosing polygon of minimum area. In each of figs. 5(c) and 5(d), the polygon is denoted with broken line segments. Consider the chain in fig. 5(d). Atoms, 1, 3, 5, 7, 6, 4, and 2 are the vertices of a polygon; the vicinal segments (1, 3), (3, 5), (5, 7), (6, 4), and (4, 2) and geminal segments (7, 6) and (2, 1) are the sides. Such a polygon will be called the polygon of the chain. In fig. 5(c), no point of segment (1, 6) is outside the polygon, whereas in fig. 5(d), (1, 7) is outside the polygon. A polygon has

concavity when the segment of a line between any two vertices has at least one point outside the polygon. The chains in fig. 5(a), 5(b), and 5(c) do not have concavity, whereas the chain in fig. 5(d) does. Note that concavity of the polygon of the planar chain occurs only in the structure where a distance is not maximized. This observation regarding concavity in figs. 5(a)–5(d) can be generalized:

Let m_{ij} denote the distance between atoms i and j, connected by a chain of dihedral bonds, that is maximal over all conformations of the structure. t_{ij} is to denote the distance in the planar all trans conformation. Then $t_{ij} = m_{ij}$ if and only if no points of straight-line segment (i, j) are outside the polygon of the chain. $t_{ij} < m_{ij}$ if and only if at least one point of (i, j) is outside the polygon.

This statement does not guarantee that in the nonplanar case m_{ij} is the sum of collinear vicinal distances. In the most general case, both planar and nonplanar regions can occur in a single chain when d_{ij} is maximal.

A further complication in some molecules is the existence of more than one connection between pairs of atoms. In the simplest case d_{ij} can be calculated for each connection independent of the others. m_{ij} is then the minimum d_{ij} over the set of connections. Consider the structural formulas for cyclohexane and cyclododecane in fig. 6. m_{14} in cyclododecane can be calculated in this way. The maximum d_{14}
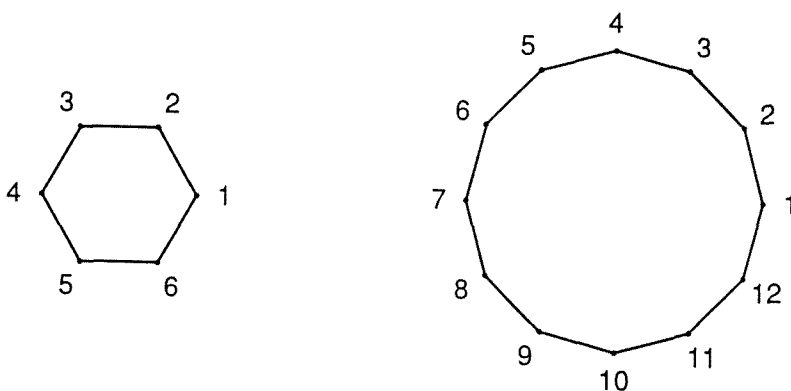

Fig. 6. Schematic illustrations of carbon backbones of cyclohexane and cyclododecane.

calculated through atoms 2 and 3 is $2.517\, d_{12}$, the same as m_{14} in butane. The maximum d_{14} calculated through atoms 12, 11, 10, 9, 8, 7, 6, and 5 is $7.365\, d_{12}$, the same as $m_{1\,10}$ in decane. Since the 5–12 connection does not interfere with the 1–4 connection,

$$m_{14\text{ cyclododecane}} = \min\{2.517\, d_{12},\ 7.365\, d_{12}\} = 2.517\, d_{12}.$$

In cyclohexane, on the other hand, the two connections between atoms 1 and 4 (through atoms 2 and 3, and through atoms 6 and 5) interfere so that $m_{14\,\text{cyclohexane}} < m_{14\,\text{butane}}$. The distances $d_{i\,i+3}$ in cyclohexane might be called *over constrained*. The exact calculation of the maximum value of an over constrained distance will not be pursued here.

Nonconvexity of atom polygons and multiple connections between atoms are significant features in many macromolecules. Many problems relating to these features remain unsolved.

Acknowledgement

References

[1] A. Szabo and N.S. Ostlund, *Modern Quantum Chemistry* (McGraw–Hill, New York, 1989).

[2] R.A. Fiesner, in: *Annual Review of Physical Chemistry*, Vol. 42, ed. H.L. Straus (Annual Reviews Inc., Palo Alto, 1991) pp. 341–367.

[3] J.M. Goodfellow, Molecular Simulation 5(1990)277.

[4] H. Hosoya, K. Morokuma, S. Yamabe and K. Ohno, J. Mol. Struct. (THEOCHEM) 84(1991)1.

[5] J.A. Simpson and E.S.C. Weiner, *The Oxford English Dictionary*, 2nd Ed. (Oxford University Press, Oxford, 1989), sense 3b.

[6] J.A. Simpson and E.S.C. Weiner, *The Oxford English Dictionary*, 2nd Ed. (Oxford University Press, Oxford, 1989), sense A1a.

[7] T.W.G. Solomons, *Organic Chemistry* (Wiley, New York, 1980) p. 104.

[8] J.S. Richardson and D.C. Richardson, in: *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. G.D. Fasman (Plenum Press, New York, 1989) p. 35.

[9] M.J. Sippl, J. Mol. Biol. 156(1982)359.

[10] W. Kabsch and Ch. Sander, Biopolymers 22(1983)2577.

[11] G.M. Crippen, J. Comput. Phys. 24(1977)96.

[12] M.J. Sippl, J. Mol. Biol. 156(1982)362

[13] M.H. Klapper and D. DeBrota, J. Comput. Phys. 37(1980)63.

[14] J.A. Simpson and E.S.C. Weiner, *The Oxford English Dictionary*, 2nd Ed. (Oxford University Press, Oxford, 1989), sense 6b.

[15] J.B. Kruskal, Psychometrika 29(1964)1;
R.N. Shepard, Psychometrika 27(1962)126.

[16] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation* (Wiley, New York, 1988) pp. 287–325.

[17] J.N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1985) p. 88, fig. 14.

[18] W.-Y. Hsiang, Reports PAM-530, PAM-535, PAM-536, Center for Pure and Applied Mathematics, University of California, Berkeley (1991).

[19] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices, and Groups* (Springer, New York, 1988) p. 15, table 1.2.

[20] J.N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1985) p. 88, fig. 14.

[21] R. Sedgewick, Computing Surveys 9(1977)137.

[22] M. Aigner, *Combinatorial Search* (Wiley and Teubner, Stuttgart, 1988).

[23] M. Karplus, A.T. Brunger, R. Elber and J. Kuriyan, in: *Structure and Expression*, Vol. 1, *From Proteins to Ribosomes*, ed. M.H. Sarma and R.H. Sarma (Adeline, New York, 1988) p. 55.

[24] K.D. Gibson and H.A. Scheraga, in: *Structure and Expression*, Vol. 1, *From Proteins and Ribosomes*, ed. M.H. Sarma and R.H. Sarma (Adeline, New York, 1988) p. 67.

[25] C.H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Prentice–Hall, Englewood Cliffs, 1982) pp. 1–16.

[26] D.S. Moss and T.P. Flores, in: *Supercomputational Science*, ed. R.G. Evans and S. Wilson (Plenum Press, New York, 1990) p. 252.

[27] *Programming Languages – APL* (Int. Organization for Standardization, Geneve, 8485:1989); K.E. Iverson, *Dictionary of J* (Iverson Software, Toronto, 1992).

[28] L.M. Blumenthal, *Theory and Applications of Distance Geometry* (Oxford University Press, Oxford, 1953) pp. 97, 98.

[29] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation* (Wiley, New York, 1988) p. 256.