A rapid method for comparing and matching the spherical parameter surfaces of molecules and other irregular objects

Peter Bladon

Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, UK

This paper describes a method for the rapid searching of rotational three-dimensional space, relying on the symmetries of the regular icosahedron and dodecahedron. This can be used to compare the spherical parameter surfaces of two molecules.

Keywords: icosahedral matching, molecular recognition, electrostatic potential surface

INTRODUCTION

In recent years, methods for comparing two (or more) molecules have received much attention.[1] One aspect of this work deals with the problems of defining and comparing molecular surfaces; another addresses the computational processes needed for measuring the degree of similarity.[2-5] The reason for all this activity stems from the need for methods of correlating the biological activity of molecules with chemical structure.

One way of approaching the comparison process is to align molecules pairwise, so that their surfaces show maximum similarity, and then to quantify the degree of similarity. Central to this process of alignment is the method whereby a complete search of three-dimensional (3D) space is carried out. Traditionally, this search has involved a preliminary low-resolution search, followed by refinement stages at higher resolution until the point of maximum similarity is found. The danger in these processes—of finding only a local point of maximum similarity and missing the global maximum—is well known. The only foolproof way of ensuring that the global maximum is found involves a complete search of 3-space with fine enough resolution; this "brute-force" approach is usually ruled out as being too costly in computer time. Under certain conditions, however, the efficiency of this brute-force method can be improved to the extent that it can become a method of choice for the rapid searching of 3D space.

A separate issue is the choice of parameters to be matched. In the particular case of assessing two molecules for similarity in biological activity, two factors are dominant: overall molecular shape and charge distribution. These factors can be considered separately or combined. In the latter case, the problem can be reduced to that of computing the energy needed to bring a positively (or negatively) charged particle from infinity to defined points in the region of the molecules.

The shape of molecules is difficult to define satisfactorily, although attempts are being made in this direction.[6-9] However, for many purposes, it is sufficient to consider points on spheres surrounding the molecules, and then it is simple to define parameters characteristic of either the shape or the charge distribution in the molecules at these points.[1] Obviously, this idea is more suitable for molecules that are globular in shape rather than flat or rod-like. Nevertheless, it is generally applicable.

It was in the process of implementing the comparison process described by Chau and Dean[2] that the algorithm described below was discovered. A key fact utilized by Chau and Dean is the symmetrical distribution of points on the surface of a sphere.

ICOSAHEDRAL MATCHING

The regular icosahedron has 12 vertices, all of which lie on the surface of a sphere, and there are 60 equivalent positions that the regular solid can adopt (there are six axes joining opposite vertices, each of which is a fivefold rotational axis, and each of which has a twofold axis at right angles to it). It follows that two arrays containing

Address reprint requests to Dr. Bladon at the Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, G1 1XL, UK.

Received 14 February 1989; accepted 14 March 1989

sets of values associated with two sets of regular icosahedral points can be compared by the simple expedient of permutation of the indices of one of the arrays. Thus, one calculation of the values at the points for each of the two sets (involving perhaps the computer-intensive calculation of square roots) can serve to search 60 positions in 3-space.

Only 12 points for the comparison process are provided by the icosahedron, but an additional 20 are provided by the regular dodecahedron, which has vertices at the centers of the triangular faces of the icosahedron. (The regular icosahedron and dodecahedron belong to the same symmetry group.) It is important to realize that the icosahedral points are permuted only among themselves, and the dodecahedral points are also permuted only among themselves; the two sets of points are treated quite separately.

The 60 equivalent orientations of the icosahedron do not span 3-space uniformly. The icosahedron is shown in Figure 1 with the axis joining vertices 1 and 12 aligned with the Cartesian y-axis; rotation about this axis gives five equivalent positions separated by 72 degrees (five orientations). Changing the axis to one joining two other vertices is equivalent to a step of 60 degrees on the Cartesian z-axis (5 × 6 orientations), and reversing each axis in the y-direction corresponds to a 180-degree rotation on the x-axis (total of 5 × 6 × 2 = 60 orientations). Matching two surfaces by searching the 60 orientations for the best match can determine the best fit to within ±36 degrees rotation on the y-axis, ±30 degrees on the z-axis, and ±90 degrees on the x-axis. The dodecahedron derived from the icosahedron, which is used to provide the second set of matching points, is shown in Figure 2.

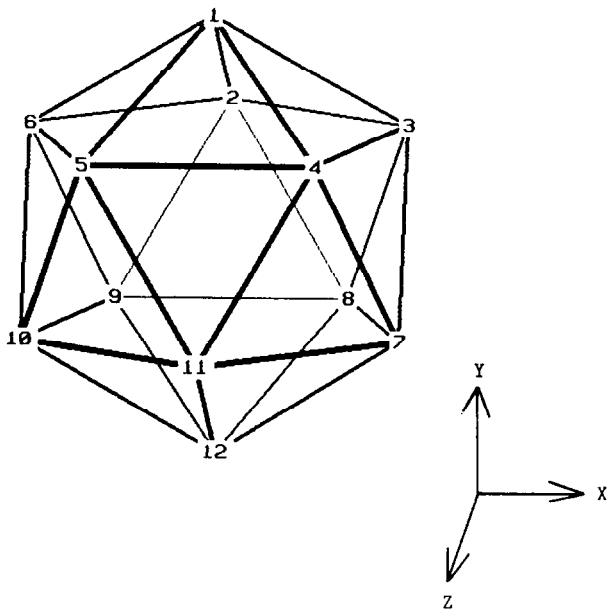

Figure 1. Standard form of the icosahedron showing numbering of vertices

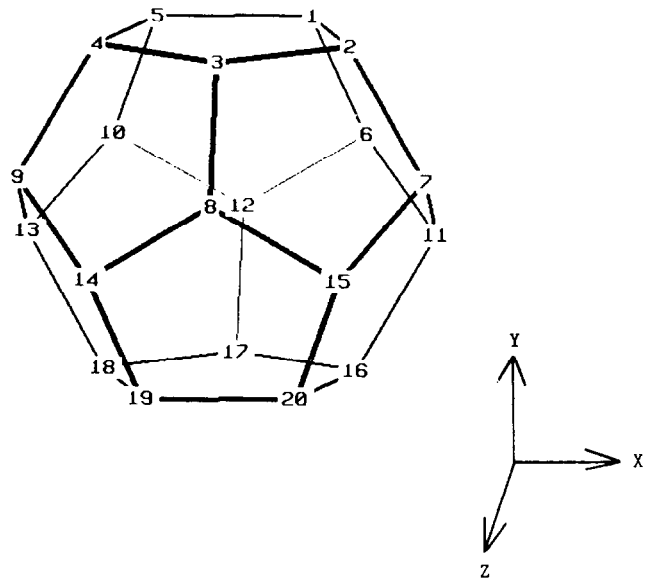

Figure 2. Standard form of the dodecahedron showing numbering of vertices

The scheme for searching all of 3-space consists of three nested loops that cause an initial orientation of one of the molecules to be rotated in, say, 10-degree increments on the x and z axes, and 9-degree increments on the z-axis; the limits on the rotations are −90 to +90, −30 to +30, and −36 to +36 degrees, respectively (see below). For each new orientation, the potentials are recalculated, and then the strategy of searching the 60 equivalent icosahedral positions, by index permutation, is reinvoked. In this way, a best match to within ±10 degrees on all three axes is obtained. Since at each stage in the triple loop, a total of 60 pairs of positions in 3-space are compared, the saving in computer time compared with the complete search of 3-space using conventional methods might be expected to approach a factor of 60. However, a brute-force search based on 360-degree rotations on all three Cartesian axes is inefficient, since this involves rotations that are improper for a rigid body. This shows up as a twofold degeneracy (i.e., every search position occurs twice in the complete process). So the maximum factor expected for computer time saving, in comparison to a properly designed brute-force method based on rotations on the Euler axes, is 30. (The author is indebted to a referee for pointing out this fact.)

In the process of matching, the criterion of maximum match is taken as being equivalent to the usual condition of a minimum in the sum of the squares of the differences in the matching parameter.

The matching procedure needs the specification of three arrays; the first (whose dimensions are 60 × 12) holds the permutations of the indices of 12 vertices of the icosahedron, the second (60 × 20) the corresponding permutations of the dodecahedron and the third (60 × 5) the five angles that are used to rotate the mobile molecule to a new orientation. Five angles are needed

because the permutations of the icosahedral system do not have simple correspondences to rotations in the Cartesian system. The contents of all these arrays are somewhat arbitrary (that is to say, alternative schemes could be devised), and they apply only if the particular alignments of the icosahedron and dodecahedron with the Cartesian axes, and the numbering of the vertices adopted here is followed. The angles in the array are used respectively to rotate the molecule on the x, y, x, y and y axes. Some small increase in efficiency can be achieved if, instead of the five angles, the corresponding elements in a rotation matrix are specified. In the appendix is shown a *pro forma* FORTRAN subroutine that, when called at every incremental step in the triple nested loop scheme, returns the values of the five angles.

In the implementation of the process in the molecular graphics program INTERCHEM,[10] as each improvement in the matching between the two molecules is found, the current best fit orientation of the mobile molecule is displayed alongside the fixed-orientation molecule. This necessitates the storage of two orientations of the mobile molecule—the original one and the current best match. At the end of this process, the original is replaced by the overall best match, and the 10-degree fitting can then be followed by a finer search with 2-degree steps.

If two (or more) successive 10-degree searches are performed using the previous best match as the starting orientation for the next search, an improvement in matching often occurs. As a development of this finding, starting each 10-degree search from an orientation obtained by random rotation on the three Cartesian axes was also shown to be beneficial. The overall effect in both cases is as if matching to finer tolerances than 10 degrees had been done.

The degeneracy noted above in connection with the brute-force method is only one possible source of redundancy in the calculation. Further redundancy occurs in both the brute-force and icosahedral matching procedures if the limits of the angle search are those specified above. However, if the limits are modified in the icosahedral method to $\pm 95°$, $-34°$ to $+36°$, and $-40°$ to $+41°$ on the z, x and y axes, respectively, this redundancy is removed. The angular step sizes remain at 10°, 10° and 9°. Extra steps are introduced, but tests have shown that a single 10-degree matching results in matching to an average level of less than $\pm 5°$. (The timings quoted in the results and discussion sections refer to the unmodified limits.)

CHOICE OF MATCHING PARAMETERS

Six schemes and associated parameters have been used in the present work for comparing molecules. They are as follows:

(1) A sphere is defined with a center coinciding with the centroid of the molecule and with a radius large enough to surround the molecule with a clearance of 10%. The 32 icosahedral/dodecahedral points are defined, and the distances from these points to the surfaces of the nearest atom are calculated. These distances are the parameters to be matched with distances similarly calculated for a second molecule. This method is referred to as *Distance Method 1.*

(2) This is identical to the foregoing method except that the reciprocals of the distances are used. This is referred to as *Distance Method 2.*

These two methods were used in the beginning in order to test out the matching methods. They have very little theoretical justification but serve to show that in the case of matching two instances of the same structures (or two very similar structures), almost any reasonable matching parameter will work! Because of time limitations imposed by administrators of the VAX cluster, these methods afforded the only functions capable of being used with large molecules such as proteins (see below).

(3) The sphere is defined as in the foregoing methods. For each of the 32 points, the electric potential is calculated due to contributions from all the atoms in the molecule. This is referred to as *Potential Method 1.*

(4) The sphere is defined as in the foregoing methods. For each of the 32 points on the sphere, a line is drawn to the center of the sphere. The point where this line cuts the surface of the molecule (i.e., the point of the van der Waals surface of the appropriate atom) is taken, and the potential is calculated for this point and assigned to the original point on the surface of the sphere. This method is referred to as *Potential Method 2.*

The two potential methods afford potential surfaces that are somewhat different; the former leads to a surface that is somewhat bland, while in the second method differences between regions of positive and negative potential are much more marked.

(5) For each of the icosahedral/dodecahedral points, the total potential is calculated using a Lennard–Jones 12–6 function plus the electric potential calculated using the *square* of the interatomic distance in the denominator to simulate a variable dielectric factor (and also to simplify the calculation). This is referred to as *Total Potential Method 1* or *Loose-Fitting Method.*

(6) This is a modification of method (5), in which the projection technique of method (4) is employed. This is referred to as *Total Potential Method 2* or *Close-Fitting Method.*

The last two methods are attempts to model the behavior of substrates in their approach to a binding site. *Total Potential Method 1* corresponds to a *loose-fitting* situation that might be expected to occur during the initial stages of the binding process, while *Total Potential Method 2* corresponds to the *close-fitting* situation when the substrate is tightly bound to the receptor.

All the models should be accepted with caution. Bear in mind that what is being compared is the similarity of two (or more) compounds of (relatively speaking) known structure in their binding to a receptor of (in general) unknown structure.

RESULTS

For testing the algorithm, structures were either generated using the program INTERCHEM or taken from

the literature in the form of X-ray crystallographic data. In the latter cases, hydrogen atoms were added if necessary. In suitable cases, the structures were submitted to the program MOPAC[11] using the option to perform a single self-consistent-field calculation. This resulted in the charges on the atoms being calculated without any structure optimization. Electric potentials were calculated in the normal way, assuming a dielectric constant of unity for the medium.

Tests were first performed on matching two instances of the same molecule, one of which (B) was in a randomly varied orientation with respect to the other (A). In all cases using all six matching parameters, rapid reorientation of the molecule (B) to near coincidence with (A) resulted. The time taken for a single 10-degree matching for typical small molecules is shown in Table 1. For comparison, times taken for the matching using a brute-force method without the icosahedral matching process are shown (search limits 0 to 350 degrees on all three Cartesian axes). Table 2 shows the timings for the matching of two nonidentical molecules using four of the matching parameters, using a single 10-degree matching followed by a 2-degree matching process.

With two nonidentical molecules, there may be more than one acceptable match position. Starting afresh from a random orientation of the mobile molecule and using single 10-degree and 2-degree passes yielded a limited number of acceptable arrangements. The results obtained by comparison of tetrodotoxin[12] (1) and saxitoxin[13] (2) using Potential Method 2 illustrate this. However, when up to three 10-degree passes were used with random reorientations after the first, then a consistent matching orientation of the mobile molecule (saxitoxin) with the fixed molecule (tetrodotoxin) was obtained in the majority of cases (five out of seven). These findings can be compared with the results reported extensively by Dean, Callow and Chau[5] for matchings using tessellations over a hemisphere.

Another example is afforded by the comparison of two conformers of the alkaloid lunarine (3). These were obtained from the X-ray crystal structures of the hydroiodide[14] and hydrobromide,[15] by removing the counterions and then adding hydrogens to the structures. Charges were then calculated by using MOPAC.[11] Because these structures differ only in the conformation of the macrocyclic ring, comparison always resulted in the near superposition of the rigid 6–5–6 rings (Table 2).

Table 1. Timings for matching two instances of the same molecule to the 10-degree level. Note that timings for Distance Method 2 are almost the same as those for Distance Method 1

Molecule	Parameter	Timings (sec.) for Icosahedral (I)	Brute-force (B)	Ratio (B/I)
Acetic acid	Distance 1	7.5	117.0	15.6
Ethylene oxide	Potential 1	9.0	173.0	19.2
Ethylene oxide	Potential 2	12.0	266.0	22.2
Tetrodotoxin	Potential 2	36.0	1209.0	33.5

Table 2. Timings for matching pairs of structures using a single 10-degree match followed by a single 2-degree match

Molecule A	Molecule B	Parameter	Time (sec.)	Notes
Tetrodotoxin	Tetrodotoxin	Distance 1	32	
Tetrodotoxin	Tetrodotoxin	Distance 2	33	
Tetrodotoxin	Tetrodotoxin	Potential 1	59	
Tetrodotoxin	Tetrodotoxin	Potential 2	73	
Tetrodotoxin	Tetrodotoxin	Close fitting	110	
Tetrodotoxin	Saxitoxin	Distance 1	31	
Tetrodotoxin	Saxitoxin	Distance 2	30	
Tetrodotoxin	Saxitoxin	Potential 1	62	
Tetrodotoxin	Saxitoxin	Potential 2	84	
Lunarine (HI)	Lunarine (HBr)	Distance 1	42	
Lunarine (HI)	Lunarine (HBr)	Distance 2	42	
Lunarine (HI)	Lunarine (HBr)	Potential 1	93	
Lunarine (HI)	Lunarine (HBr)	Potential 2	128	
Trypsin	Chymotrypsin	Distance 1	130	(1)
Trypsin	Chymotrypsin	Distance 1	907	(2)

(1) Alpha-carbons only for single chains
(2) Full structures (without hydrogens) for single chains

Finally, tests were performed, using Distance Method 1 only, on the matching of larger structures. The structures of the proteins trypsin[16,17] and chymotrypsin[18,19] were extracted from the Brookhaven Protein Data Bank[20,21] in the form of both alpha-carbon skeletons and full structures (without hydrogens). Matching of the alpha-carbon skeletons was relatively rapid (see Table 2) and resulted in an orientation of the mobile molecule (chymotrypsin), which was clearly well aligned with the fixed structure (trypsin). Matching of the full structures, with single 10-degree and 2-degree processes, took a somewhat longer but still acceptable time (about 15 minutes of CPU time). Because of the complexity of the structures, it was not immediately apparent that a meaningful alignment had been obtained; however, restricting the display to the cysteine residues (numbered 42, 58, 136, 168, 182, 191, 201, 220; these are key secondary structural features present in both proteins) and viewing the result in an overlay mode left no doubt that the alignment was good, as corresponding pairs of residues were coincident. This alignment, which is to be expected from the homology of these proteins, was obtained in two out of three experiments in which the starting orientation of chymotrypsin was varied; in a third attempt, a different alignment was achieved (the sum of squares of residuals was much higher), and this corresponds to the reverse alignment of the protein chains, consequent upon the well-known twofold screw axis present in the structures of the serine proteases.[22] The degree of alignment achieved in the case of these two proteins was measured as the average angular separation of the corresponding amino-nitrogen atoms in the cysteine residues mentioned above; after a single 10-degree match. It was 6.8 degrees.

DISCUSSION

The principles involved in the method described here can be applied in other situations where a rapid search of 3D space is required. The limited number of matching points (32) used in this implementation could be increased by, for example, using the midpoints on the edges of the icosahedron (30 extra points). This and other extensions that employ the vertices of tessellated icosahedra of varying degree require some care in use, since the points will not all belong to the same permutation set. For example, the midpoints of icosahedral edges form three overlapping sets of twenty points; each midpoint belongs to two sets. In the implementation described, the vertices of the regular dodecahedron, which, as explained above, are treated for comparison purposes quite from the icosahedral points, are one of the sets of points that would be involved in using tessellated icosahedra of even tessellation frequency. For each additional set of points, a separate permutation matrix needs to be constructed. (This can be done most conveniently by a separate computer program.) It would be possible to combine all the permutation matrices into one matrix, but modifications in the program would be made more difficult if this were done.

In the examples shown in Table 1, the savings in time using the icosahedral matching process as compared to the brute-force process (as represented by the reciprocal of the ratio of the relative times) increases as the size of the molecule increases. This is as expected, since as the molecular size increases, so does the proportion of the time taken for calculating the parameter surface. The time for the icosahedral comparison is constant and independent of the parameter being compared (it is approximately 5 seconds for the full 10-degree icosahedral search on the VAX 8650 computer). The question arises as to the ultimate limit of the ratio when comparing the molecular surfaces of large molecules. Several factors operate to reduce this ratio from the expected value of 60. In the present case, the step size on one of the axes is 9 instead of 10 degrees (largely for convenience of coding the algorithm); it can be shown that this reduces the expected limiting ratio to 54. Second, it is found in practice that in the icosahedral search algorithm, the limits of the angle searches within the loops need to be -90 to $+90$ in 19 steps, -30 to $+30$ in 7 steps, and -36 to $+36$ in 9 steps, rather than -90 to $+80$ in 18 steps, -30 to $+20$ in 6 steps, and -36 to $+27$ in 8 steps as might be expected. The total number of calculation steps in the triple loop is therefore:

$$19 \times 7 \times 9 = 1197$$

The brute-force method requires limits of 0 to 350 degrees in 10-degree steps on all axes; thus, a total number of steps equal to:

$$36 \times 36 \times 36 = 46{,}656$$

The limiting ratio is therefore:

$$46{,}656/1197 = 38.98$$

In the case of matching two instances of the molecule tetrodotoxin, the efficiency of the icosahedral search compared to the brute-force method is approaching this limit (ratio = 33.5).

Molecular recognition is now recognized as the key step in many guest–host interaction processes, and its study using computer graphics is now commonplace. The algorithm described here should be a useful addition to the tools available, particularly for the preliminary orientation of a pair of molecules. However, it should be emphasized that the algorithm is quite general; applications outside the field of molecular manipulations can be envisaged—for example, in the problems encountered in recognizing the shapes of arbitrary objects. Nor is the choice of the matching limited to metric parameters; the method could be applied to the alignment of two multicolored globes for maximum matching of colors.

ACKNOWLEDGEMENTS

The author is indebted to Professor Brian Ripley of the Department of Mathematics, and Professor Douglas McGregor of the Department of Computer Science, University of Strathclyde, for helpful comments on this paper. The help and advice of the members of the staff of the Computer Centre, and of Dr. David Pugh, Dr. David Armstrong of the Department of Pure and Applied Chemistry, University of Strathclyde, is also gratefully acknowledged.

PROGRAM DETAILS

Full listings of the data arrays, etc., can be obtained by writing to the author. Data can also be supplied on 51/4-inch double-sided, double-density disks if formatted disks are sent.

REFERENCES

1 Richards, W.G. and Hodgkin, E.E. *Chemistry in Britain* 1988, **24**, 1141
2 Chau, P.-L. and Dean, P.M. *J. Mol. Graphics* 1987, **5**, 97
3 Dean, P.M. and Chau, P.-L. *J. Mol. Graphics* 1987, **5**, 152
4 Dean, P.M. and Callow, P. *J. Mol. Graphics* 1987, **5**, 159
5 Dean, P.M., Callow, P. and Chau, P.-L. *J. Mol. Graphics* 1988, **6**, 28

6 Arteca, G.A. and Mezey, P.G. *J. Comput. Chem.* 1988, **9**, 554
7 Arteca, G.A., Jammal, V.B. and Mezey, P.G. *J. Comput. Chem.* 1988, **9**, 608
8 Arteca, G.A., Jammal, V.B., Mezey, P., Yadaf, J.S., Hermsmeier, M.A. and Gund, T.N. *J. Mol. Graphics* 1988, **6**, 45
9 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Ventkataraghavan, R. *J. Med. Chem.* 1988, **31**, 722
10 Breckenridge, R.J. and Bladon, P. Distributed by Interprobe Chemical Services Limited, University of Strathclyde, Glasgow, G1 1XL
11 Stewart, J.J.P. QCPE Program 455, and later versions distributed by the author from USAFA, Colorado Springs, Colorado, USA
12 Furusaki, A., Tomiie, Y. and Nitta, I. *Bull. Soc. Chim. Japan* 1970, **43**, 3332
13 Schantz, E.J., Ghazarossa, V.E., Schnoe, H.K., Strang, F.M., Springer, J.P., Pezzanita, A.O. and Clardy, J. *J. Amer. Chem. Soc.* 1975, **97**, 1238
14 Tamura, C. and Sim, G.A. *J. Chem. Soc. (B)* 1970, 991
15 Jeffreys, J.A.D. and Ferguson, G.A. *J. Chem. Soc. (B)* 1970, 826
16 Walker, J., Steigman, W., Singh, T.P., Bartunik, H. and Bode, W. *Acta Cryst. (B)* 1982, **38**, 1462
17 Entry 3PTN in Brookhaven Protein Data Bank
18 Tulinsky, A. and Brown, R.A. *J. Biol. Chem.* 1967, **262**, 7737
19 Entry 5CHA in Brookhaven Protein Data Bank
20 Bernstein F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr, E.F., Price, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. *J. Mol. Biol.* 1977, **112**, 535
21 Weng, J. in *Crystallographic Databases — Information Content, Software Systems, Scientific Applications* (Allen, F.H., Bergerhoff, S. and Sievers, R., Eds.). Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, 107–132
22 McLachlan, A.D. *J. Mol. Biol.* 1979, **128**, 49

APPENDIX

```
      BLOCK DATA ICOSAH
C                                   WARNING!!
C  This BLOCK DATA segment contains statements with more than the
C  maximum number of CONTINUATION LINES permitted by FORTRAN77.
C  It is usually possible to override this limitation by a compiler,
C  directive (e.g., /CONTINUATIONS=99 in VAX FORTRAN).

      COMMON/MATPER/ MATPER(60,12), ANGPER(60,5), MATDOD(60,20)

C  Permutation matrix. Each row gives the indices of the vertices of the
C  icosahedron, corresponding to one arrangement
      DATA ((MATPER(I,J), J=1,12), I=1,60)/

     * 1,  2, 3, 4, 5, 6,  7, 8, 9,10,11, 12,
     * 1,  3, 4, 5, 6, 2, 11, 7, 8, 9,10, 12,
     * 1,  4, 5, 6, 2, 3, 10,11, 7, 8, 9, 12,
     * 1,  5, 6, 2, 3, 4,  9,10,11, 7, 8, 12,
     * 1,  6, 2, 3, 4, 5,  8, 9,10,11, 7, 12,
                          .
                          .
                          .
                  (60 lines in all)
                          .
                          .
                          .
     * 7,  3, 8,12,11, 4,  9, 2, 1, 5,10,  6,
     * 7,  8,12,11, 4, 3, 10, 9, 2, 1, 5,  6,
     * 7, 12,11, 4, 3, 8,  5,10, 9, 2, 1,  6,
     * 7, 11, 4, 3, 8,12,  1, 5,10, 9, 2,  6,
     * 7,  4, 3, 8,12,11,  2, 1, 5,10, 9,  6/

C  Permutation matrix for the vertices of the dodecahedron.
      DATA ((MATDOD(I,J), J=1,20), I=1,60)/
     +  1, 2, 3, 4, 5, 6, 7, 8, 9,10,11,12,13,14,15,16,17,18,19,20,
     +  2, 3, 4, 5, 1, 7, 8, 9,10, 6,15,11,12,13,14,20,16,17,18,19,
     +  3, 4, 5, 1, 2, 8, 9,10, 6, 7,14,15,11,12,13,19,20,16,17,18,
```

```
     +    4, 5, 1, 2, 3, 9,10, 6, 7, 8,13,14,15,11,12,18,19,20,16,17,
     +    5, 1, 2, 3, 4,10, 6, 7, 8, 9,12,13,14,15,11,17,18,19,20,16,
                                        .
                                        .
                                  (60 lines in all)
                                        .
                                        .
     +   11,16,20,15, 7, 6,17,19, 8, 2,12, 1, 3,14,18,10, 5, 4, 9,13,
     +   16,20,15, 7,11,17,19, 8, 2, 6,18,12, 1, 3,14,13,10, 5, 4, 9,
     +   20,15, 7,11,16,19, 8, 2, 6,17,14,18,12, 1, 3, 9,13,10, 5, 4,
     +   15, 7,11,16,20, 8, 2, 6,17,19, 3,14,18,12, 1, 4, 9,13,10, 5,
     +    7,11,16,20,15, 2, 6,17,19, 8, 1, 3,14,18,12, 5, 4, 9,13,10/
C    Angle matrix. The angles in each group of five are to applied in the
C    order given on the x, y, x, y and y axes.
C    Each set of angles corresponds to one of the rows in the permutation
C    array, which are the angles that are to be applied to the first arrangement
C    (permutation) to obtain another arrangement.
C    Angles are specified in degrees and apply in the normal left-handed
C    Cartesian axis system. Positive angles denote anticlockwise
C    rotation when looking along the appropriate axis toward the origin.
      DATA ((ANGPER(I,J), J = 1,5), I = 1,60)/
     *    0.00,    0.00,    0.00,    0.00,    0.00,
     *    0.00,    0.00,    0.00,    0.00,   72.00,
     *    0.00,    0.00,    0.00,    0.00,  144.00,
     *    0.00,    0.00,    0.00,    0.00,  216.00,
     *    0.00,    0.00,    0.00,    0.00,  288.00,
                                        .
                                        .
                                  (60 lines in all)
                                        .
                                        .
     *  180.00,   72.00,   63.44,   36.00,    0.00,
     *  180.00,   72.00,   63.44,   36.00,   72.00,
     *  180.00,   72.00,   63.44,   36.00,  144.00,
     *  180.00,   72.00,   63.44,   36.00,  216.00,
     *  180.00,   72.00,   63.44,   36.00,  288.00/
      END
```

THE SUBROUTINE COMPAR

COMPAR is a routine that will make 60 comparisons of the pairs of arrays POTENA, POTENB and POTENC, POTEND. The index permutations are taken from the arrays MATPER and MATDOD. That comparison that gives the lowest sum of squares of residuals is noted, and the corresponding set of five angles is abstracted from the array ANGPER and returned as parameters. The lowest value of the sum of squares of residuals (POTDIF2) is also returned.

(Arrays POTENA and POTENC are for molecule –A–; POTENB and POTEND are for the mobile molecule –B–)

A suitable scheme for using this routine is as follows:

(code to assign values for potentials at the icosahedral and dodecahedral points for the fixed molecule –A–):

```
      POT2 = 1.0E+30   ! Set initial value of sum of squares
      DO 100   X = -30.0, 30.0, 10.0 ! Start of three nested
       DO 200  Y = -36.0, 36.0,  9.0 ! for rotation of molecule
        DO 300 Z = -90.0, 90.00, 10.0 ! -B-
         CALL RESTOR ! Restore initial orientation of -B-
         CALL ROTATE ("Y",Y)
         CALL ROTATE ("X",X)
         CALL ROTATE ("Z",Z)
         CALL FIXB ! Apply rotation matrix to molecule -B-
         (Code to (re)calculate potentials for molecule -B-)
         CALL COMPAR (POTENA,POTENB,POTENC,POTEND,A1,A2,A3,A4,A5,SSR)
         IF (SSR.LT.POT2) THEN
           CALL ROTATE ("X",A1) ! Four calls to a subroutine that
           CALL ROTATE ("Y",A2) ! modifies a rotation matrix that
           CALL ROTATE ("X",A3) ! is applicable to molecule -B-
           CALL ROTATE ("Y",A4+A5)
           CALL FIXB ! Call to a subroutine that applies the rotation
                     ! matrix to molecule -B-

            (Redisplay the reoriented molecule -B-

         ENDIF
300      CONTINUE
200     CONTINUE
100    CONTINUE

      SUBROUTINE COMPAR (POTENA,POTENB,POTENC,POTEND,
     1 ANG1,ANG2,ANG3,ANG4,ANG5,POTDF2)

      COMMON/MATPER/ MATPER(60,12), ANGPER(60,5), MATDOD(60,20)
      DIMENSION POTENA(12),POTENB(12),POTENC(20),POTEND(20)

      POTDF2 = 1.0E+30
      DO 100 I = 1,60
      DIF2 = 0.0
C   Compare first the icosahedral points
      DO 110 J = 1,12
      DIF = POTENA(J)—POTENB(MATPER(I,J))
      DIF2 = DIF2 + DIF*DIF
110   CONTINUE

C   and next the dodecahedral points
      DO 120 J = 1,20
      DIF = POTENC(J)–POTEND (MATDOD(I,J))
      DIF2 = DIF2 + DIF*DIF
120   CONTINUE
C   Now find that orientation of both icosahedron and dodecahedron that
C   corresponds to minimum energy difference
      IF (DIF2.LT.POTDF2) THEN
      POTDF2 = DIF2
      MINORI = I
      ENDIF
100   CONTINUE

C   Now return the angles by which structure -B- must be rotated to get
C   best potential agreement with -A-
      ANG1 = ANGPER (MINORI,1)
      ANG2 = ANGPER (MINORI,2)
      ANG3 = ANGPER (MINORI,3)
      ANG4 = ANGPER (MINORI,4)
      ANG5 = ANGPER (MINORI,5)
      RETURN
      END
```