

An investigation into the construction of molecular models by the template joining method

Andrew R. Leach, Keith Prout and Daniel P. Dolata*

Chemical Crystallography Laboratory, 9 Parks Road, Oxford OX1 3PD, U.K.

Received 15 February 1988

Accepted 30 March 1988

Key words: Artificial intelligence; Template joining method; Molecular modeling

SUMMARY

The results of a wide-ranging investigation into some of the different methods available for performing the 'joining' of templates to build molecular models show that the choice of algorithm can significantly affect the quality of the results obtained, and different algorithms are most suited to particular categories of join.

INTRODUCTION

There is considerable interest in the use of computers to generate chemically meaningful and interesting conformations of molecules, to which end a number of numerical methods have been developed (e.g. molecular mechanics and distance geometry). These methods can be extended to address the problem of searching the molecule's conformational space with techniques such as torsion angle driving and Monte Carlo search methods. In our work on the development of an expert system (WIZARD [1,2]) which utilises AI (artificial intelligence) techniques to search conformational space and identify conformations near energy minima we use a different method of producing Cartesian coordinates for the conformations, namely *template joining*.

This method of building molecular models is used in a number of other programs [3–8]. Such programs access a *library* which contains atomic (x,y,z) coordinates for conformations of distinct groups of connected atoms. In this paper these groups of connected atoms will be known as *conformational units* (typical examples being cyclohexane, cyclopentene, 1,3-butadiene and carboxyl), and each individual set of coordinates will be called a *template*. A conformation of the molecule is constructed by selecting templates from the library, and *joining* them together in a stepwise process. A simple example is the construction of decalin from two cyclohexane rings. In some programs the user must choose which templates to use, whereas in others this is done automatically.

*Present address to be used for correspondence: Department of Chemistry, University of Arizona, Tucson, AZ 85721, U.S.A.

Each stage of the construction involves two *fragments* of the molecule, a fragment being a group of connected atoms which contains one or more templates. Hence each join is akin to taking Dreiding models of the two fragments, and positioning them in space, after which they are merged together to give a combined structure. However, unlike Dreiding models, common atoms and bonds can be superimposed. The construction of conformations using Dreiding models involves the connection of 'half bonds', which leaves a degree of rotational freedom undefined, whereas the superimposition of two defined fragments is totally constrained (consider building butane by connecting two ethanes vs. overlapping two propanes). There are therefore two distinct operations performed in each join, as shown in Fig. 1 — the *fitting* of the fragments (A), and then the *assignment* of coordinates to form the resulting structure (B). The resulting model may then be minimized using a numerical method such as molecular mechanics or quantum mechanics. The advantages of this approach lie in the speed with which conformations of large, complicated molecules can be built, and the ease with which the conformations can be refined, as they are often very close to the actual energy minima.

As can be imagined, when constructing a conformation some templates fit exactly (for example the construction of *trans*-decalin from two chair cyclohexanes). In such cases all the different methods should produce the same model. However, when assembling two templates which do not show a perfect fit it is possible that different algorithms would afford a variety of structures, which although similar would have substantially different energies. A simple example is the formation of bicyclo-[3.2.1]-octane from cyclohexane and cyclopentane. Here, the difference between the torsion angles for the two main bonds of overlap is some 20° (60° for chair cyclohexane and 40° for cyclopentane). To illustrate the range of conformations which can be obtained, Fig. 2 shows two structures for this molecule produced from the same cyclohexane and cyclopentane templates, but using different fitting and assignment algorithms. The structures have been overlaid on the MM2 [9] minimized structure, and there is a substantial difference both in the shapes and energies of the conformations.

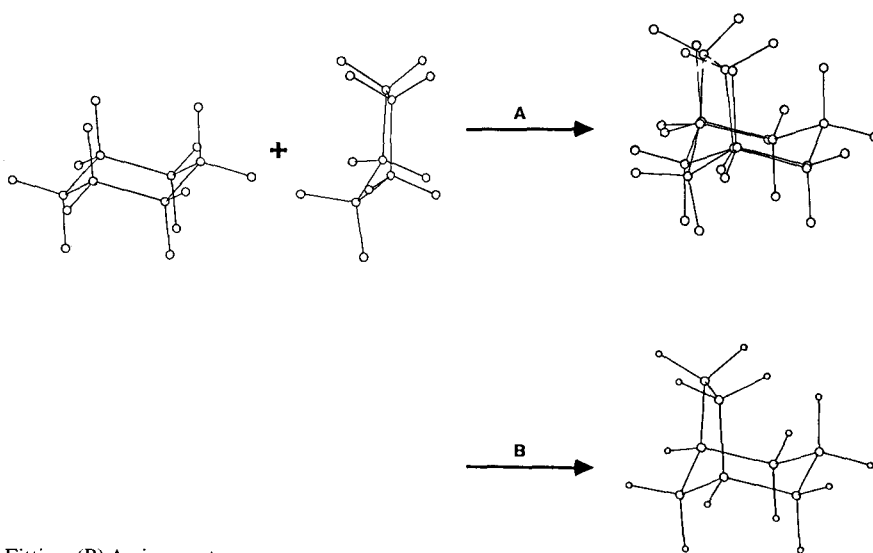


Fig. 1. (A) Fitting; (B) Assignment.

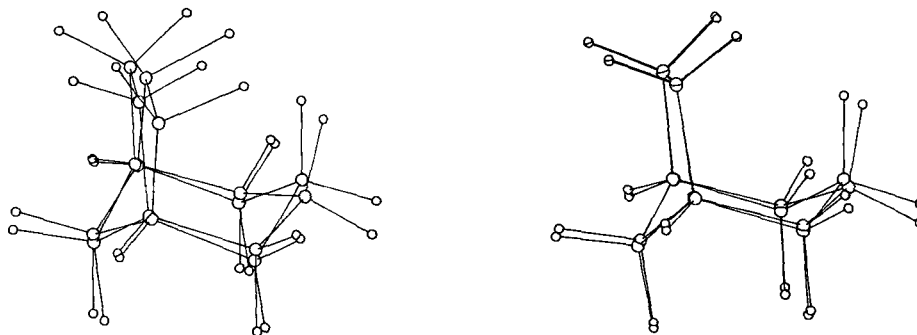


Fig. 2. Comparison of two structures, obtained using different fitting and assignment algorithms, shown as superimpositions on the MM2-minimized structure.

Thus it can be seen that the choice of algorithms for the two processes of fitting and assignment determines the quality of the results, and it is clearly of importance to us that the joins performed by WIZARD produce the best results possible from the library of templates available to the system. Our objective is to produce structures which are as close as possible to conformational energy minima (as determined by some numerical method of calculation) or to X-ray structures. The joining algorithms used in other programs have not been explicitly stated in most cases and so we have investigated a number of different methods to examine their scope and limitations. Since the joining algorithms contain two broadly independent steps (fitting and assignment), and since these steps also contain independently variable choices, we will describe these steps and choices, providing the results of each algorithm separately and then showing which combinations are best.

The fitting part of the join involves the movement of the fragments in Cartesian space until they are at the position of 'best fit'. We have restricted our investigations to fitting algorithms which do not change the individual fragment conformations and have identified two broad *classes* of algorithm which can be used for this step. The first class uses some numerical method to determine the position of best fit, such as a least-squares technique. The second class of fitting algorithm attempts to replicate in some fashion the method a chemist might use if faced with the same task. We also distinguish between different *categories* of join – for example, spiro joins may be treated in one way, single bond fusions in another, and bridging joins in a third. After fitting, each fragment retains its original three-dimensional structure, and an assignment algorithm must be used to produce a combined structure from the Cartesian coordinates of the two fragments. There are a number of ways in which this assignment algorithm could operate.

The various algorithms we tested were implemented to be compatible with the rest of the WIZARD system. Consequently, the detailed description of the algorithms given below requires a certain amount of knowledge about relevant parts of WIZARD, in particular about the structure of the conformational knowledge contained in the library (Fig. 3).

The knowledge frame about each conformational unit contains mapping information required to describe the unit (atom types, bond connectivity, and stereochemistry where appropriate). As we are currently using WIZARD to discover the minimum energy conformations of a molecule, its template library contains minimum energy templates, which have so far been obtained in one of two ways. In the first the Cambridge Structural Database is searched to find all the molecules

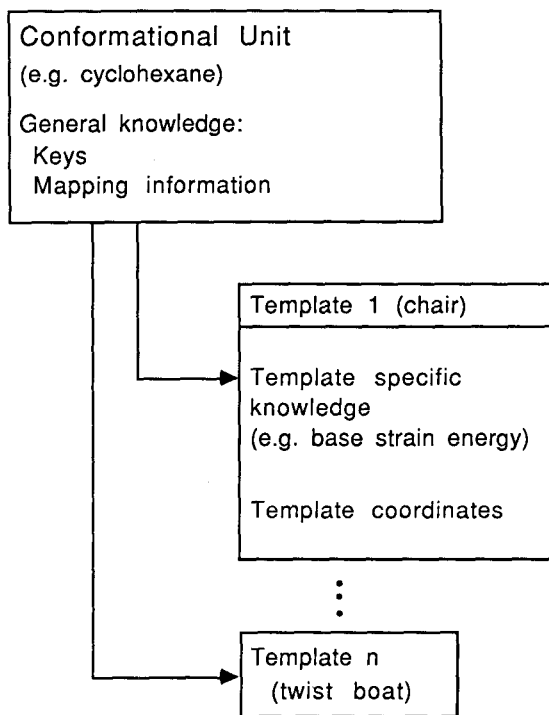


Fig. 3. Knowledge about conformational units.

in it which contain the conformational unit as a fragment. The resulting structures are then analyzed to find the set of minimum energy conformations, using statistical techniques [10–12] if necessary. The alternative approach is to use various numerical techniques (such as torsion angle driving, Monte Carlo search, etc.) to search the conformational space of the unit, and to identify all of its minimum energy conformations. The templates are then entered in the template library.

One piece of information contained about each unit which is used during the actual joining of the templates is a list of the so-called *vector* atoms in the unit. The Cartesian coordinates of such atoms in the template file indicate the *direction* but not the length of a particular bond in the molecule. During the actual joining procedure all vector bonds in each fragment are adjusted in length to give the correct bond distances according to the lengths found in the fully defined fragment. This is the first step illustrated in Fig. 4, which shows the construction of decalin from cyclohexane units in which each hydrogen in the cyclohexane is specified as a vector atom. Note that this process must be performed either explicitly (as in WIZARD) or implicitly in programs which obtain their data from other sources (e.g. from a crystal data base). The bond length for a given vector atom in a fragment is set equal to that of the corresponding bond in the other fragment if it is defined; otherwise an internal bond fragment is used. The atom labelled *i* corresponds to a vector atom in unit 1, but to a non-vector atom in unit 2. So the bond length to vector atom *i* in unit 1 is set to the bond length from unit 2. Atom *j*, however, is a vector atom in both unit 1 and unit 2, and so an additional C-H template is used to set the bond length in both fragments. Atom *k* is a vector atom in unit 1 alone, and its bond length is also set using the C-H template.

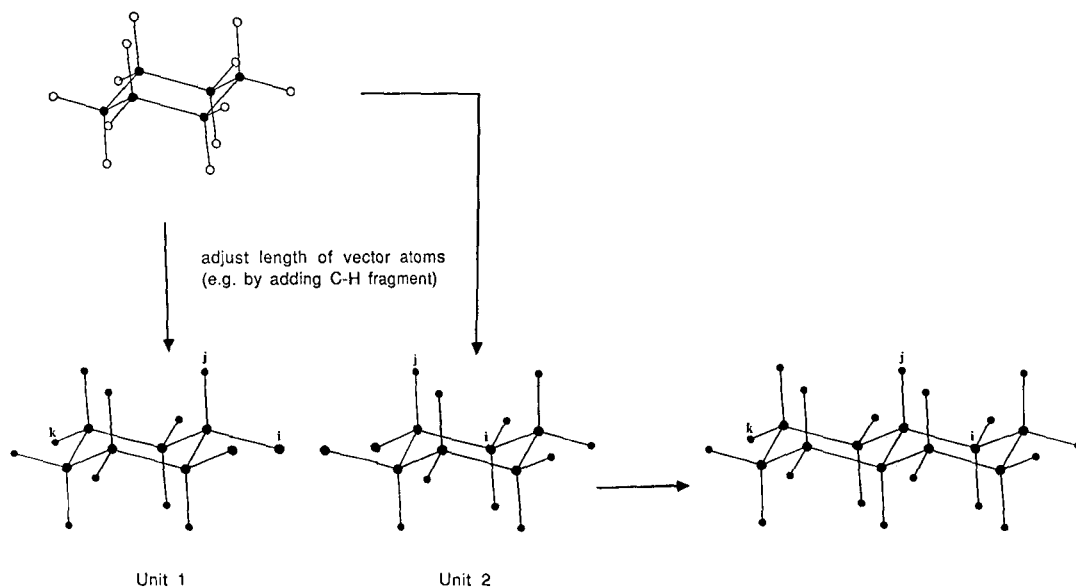


Fig. 4. Construction of decalin from two cyclohexane units. Coded as ○ vector atom, and ● non-vector atom.

DISCUSSION OF THE ALGORITHMS

Before any actual fitting occurs it is necessary to discover the *common atoms* involved in the join. These are the atoms in the molecule which have coordinate values defined in both fragments and are used to determine the subsequent operations performed on the two fragments. Figure 5 shows the common atoms involved in the construction of bicyclo-[3.2.1]-octane. The common atoms are classified into three types, which is a consequence of a distinction that is made between primary and non-primary atoms in the fragments. It should be noted that some template joining programs do not make such distinctions. In our studies, we explore the result of such an approach when the weighting schemes described below assign equal weights to these types (i.e., do not distinguish between the atoms). The first type of common atom comprises those which are non-primary atoms in both fragments; the second type those common atoms which are a non-primary atom in one fragment and a primary atom in the other; and the third type those common atoms

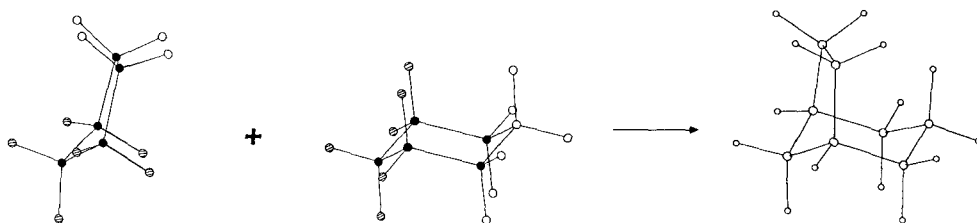


Fig. 5. Classification of common atoms. Coded as ● non-primary atom, and ⊙ primary atom.

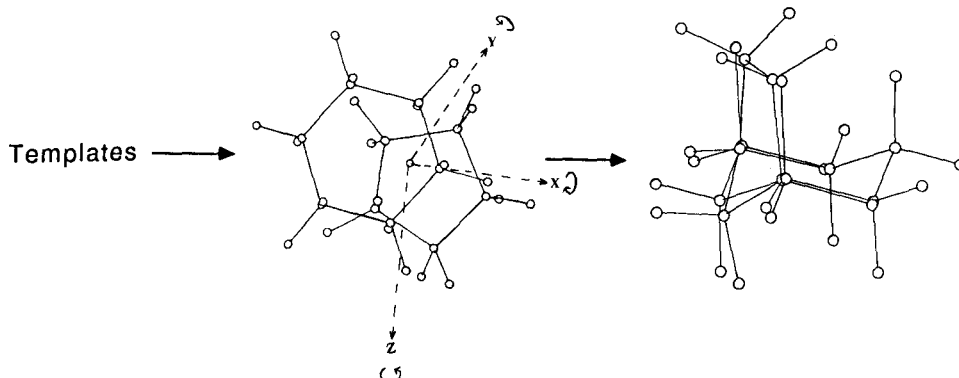


Fig. 6. Least-squares fitting.

which are primary atoms in both fragments. In Fig. 5 the non-primary and primary common atoms are distinguished for the two fragments using solid and striped filling.

As defined above, the first class of fitting algorithm uses a numerical method to determine the position of the 'best fit'. We have implemented two least-squares fitting algorithms, one using quaternions [13], and the other using an iterative search involving successive rotations as described by Nyberg [14]. Because comparative studies showed that the quaternion method gave no advantages when used to join fragments, and we encountered some problems with its use (e.g. how to deal with 180° rotations, more details of which can be found in the reference) we have concentrated our attention on the Nyberg algorithm. First, the centroids of the common atoms are calculated for each fragment. Both fragments are then translated so that these centroids are positioned at the origin. One of the fragments is then rotated in space until the sum of the squares of the separations of the common atoms is at a minimum (Fig. 6). This defines the position of 'best fit'. The method can be extended by giving different weighting coefficients to the atoms involved. Any weights are applied both to the calculation of the centroids and during the rotation. The same algorithm is used in all three categories of join. However, as discussed below, different weightings can be used for the different categories.

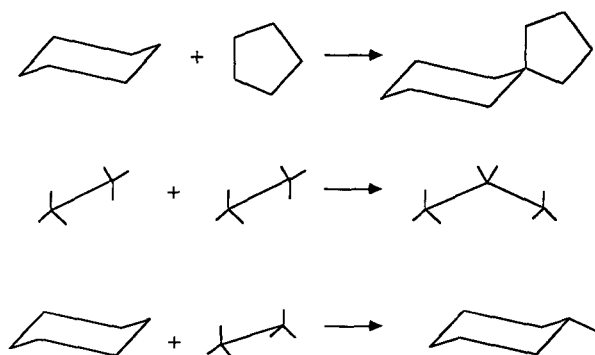


Fig. 7. Joins involving one type 1 common atom.

The second class of fitting algorithm designed to construct molecular models is used by the 'Rings' program [8], written at the University of Lund as part of a pilot study for an earlier version of WIZARD [1]. The success of this study led to its incorporation into the MOLBUILD program [7]. We have extended the technique to cover a wider range of possibilities and incorporated it into our current version of WIZARD. The method recognizes the three different categories of join described above (spiro, single bond fusion and bridging). These three categories are treated differently, as follows:

In the first category of join there is a single common atom of type 1 (non-primary/non-primary). Examples of such joins are the spiro ring join between cyclohexane and cyclopentane, the joining of two ethane fragments to produce propane or the construction of methyl cyclohexane from an ethyl fragment and a cyclohexane ring (Fig. 7). The two fragments are first translated so that in both the single type 1 common atom is positioned at the origin. Two adjoining atoms are then chosen, and a 'dummy' atom created at their midpoint. These atoms must be common to both fragments, and they are taken from the same ring (if appropriate). Each fragment is then rotated about the origin until its 'dummy' atom is positioned along the positive x-axis (Fig. 8).

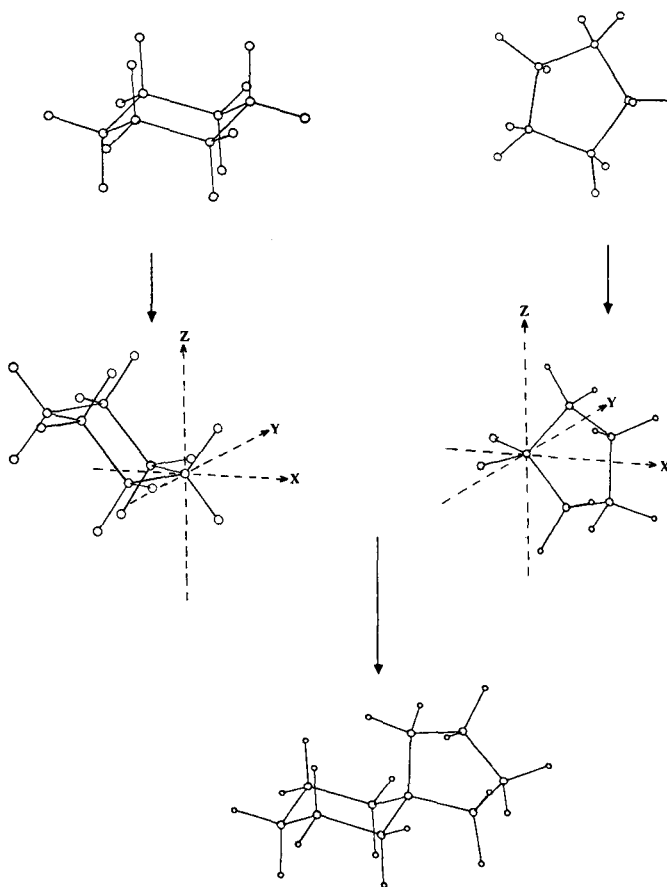


Fig. 8. 'Rings' method for a spiro join.

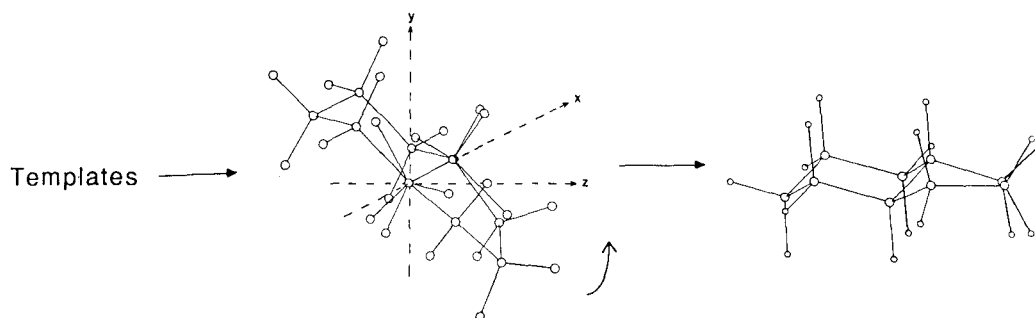


Fig. 9. 'Rings'-derived method for single bond fusions.

In the second category of join there are two type 1 common atoms, as exemplified by the construction of styrene from a benzene ring and a diene fragment or of bicyclo-[4.3.0]-nonane from cyclohexane and cyclopentane (Fig. 9). In such single bond fusions the fragments are first positioned so that the type 1 common atoms are overlaid, one pair at the origin and the other pair along the positive x-axis. One of the fragments is then rotated about the x-axis until the position of best fit for the remaining common atoms is reached. This is determined using a 'rotation function', the position of best fit corresponding to a minimum value in the function. Two such functions were tested, one based on a distance measurement and the other on an angular measure. The distance rotation function is given by the sum of the squares of the separations for the remaining common atoms, whereas the value of the angular rotation function is given by the sum of the squares of the residual angles for the remaining common atoms. A residual angle is defined by the position of an atom in one fragment, its adjoining non-primary atom, and the position of the corresponding atom in the other fragment. A binary search is used to rotate one of the fragments until the minimum value in the function is reached. In order to reduce the possibility of driving to a false minimum the fragment to be rotated is initially positioned so that one of the pairs of remaining common atoms is directly overlaid. It is also possible to apply different weights to the atoms involved in the calculation of the rotation function.

In a bridging join there are three or more type 1 common atoms, as in the construction of bicyclo-[3.2.1]-octane from cyclohexane and cyclopentane. In such cases the 'Rings'-derived method is somewhat similar to that used for the single bond fusion. The pair of type 1 common atoms which have the greatest interatomic separation are found, and both fragments positioned so that one of these atoms is at the origin, and the other along the positive x-axis. One of the fragments is then rotated about this axis, to the position of best fit (Fig. 10). A rotation function is used to determine this position; as with the single bond fusion both distance and angular variants have been tried. We have tried using four different sets of atoms in the evaluation of this function:

Set 1 contains all the remaining common atoms;

Set 2 contains the remaining type 1 common atoms;

Set 3 contains the common atoms which are bonded to the two which define the axis of rotation;

Set 4 contains those atoms in Set 3 but also any remaining type 1 common atoms.

For bridging joins in which there are three type 1 common atoms (as in the construction of bicyclo-[3.2.1]-octane) sets 3 and 4 are equivalent.

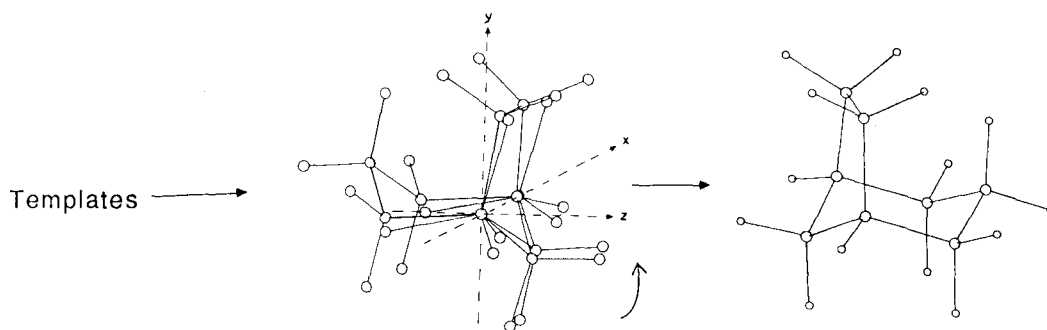


Fig. 10. 'Rings'-derived method for ring bridging joins.

Both the least-squares algorithm and the 'Rings'-derived algorithm allow *weighting schemes* to be used in the fitting process. The three different types of common atoms form the basis by which these weighting schemes are applied. Each weighting scheme is characterized by three numbers, e.g., (5:2:1); the first number gives the weight applied to the type 1 common atoms, the second number gives the weight applied to the type 2 common atoms, and the third number gives the weight of the type 3 common atoms.

The positioning of the two fragments thus completes the first part of the two-stage joining procedure. We now turn to the independent choice of algorithm for the assignment of coordinates in the second part to give a combined structure. This assignment of coordinates for each atom in the final structure can be performed in one or more ways, depending upon whether or not the atom is a common atom. For the atoms which are not common atoms, only one set of coordinates exists and so a straight assignment occurs from the appropriate fragment to the final structure. The common atoms, however, will have two sets of coordinates defined, one in each fragment, from which a combined value is to be derived. One possibility we have explored is to take an average value. An alternative approach uses the non-primary/primary distinction to preserve the skeletal structure of the two fragments. Here, average values are used for the type 1 and type 3 common atoms (i.e., non-primary/non-primary and primary/primary). For the type 2 common atoms (which are non-primary in one fragment but not in the other) however, the coordinates of the non-primary atom alone are transferred.

EXPERIMENTAL

There are therefore a large number of possible ways in which a given join can be performed, with the two fitting algorithms and the variations possible within them, two assignment algorithms, and an essentially infinite variety of weighting schemes. Initially, we investigated a small number (6) of test cases, joining some of the cycloalkane templates from the WIZARD template library in each of the three different categories of join (spiro, single bond fusion and bridging), using all the possible variants (weighted/unweighted, angular/distance rotation function, average coordinates assigned/non-primary atom structure preserved, etc.). Three weighting schemes were used in this initial trial; (1:1:1) (i.e., unweighted, corresponding to those programs which do not distinguish between types of common atoms), (2:1.5:1) and (10:5:1). The results are shown in Tables 1–4 with the joins performed illustrated in Fig. 11. In Table 1 are the results of performing

Joins performed:

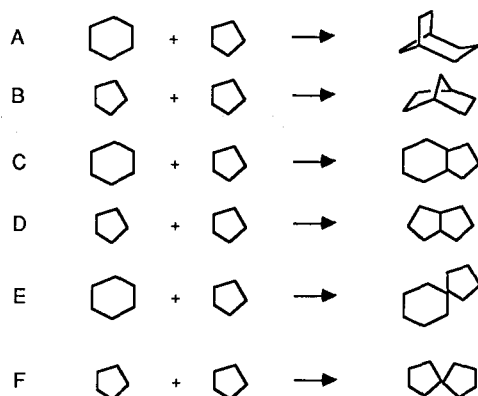


Fig. 11. Joins performed in initial trial.

the six joins using the least-squares fitting algorithm. Tables 2–4 give the results when the same joins are performed using the appropriate ‘Rings’-derived algorithm. Our aim in performing this initial trial was to see if any individual subscheme (e.g. distance rotation function) always gave high energy structures, irrespective of other option choices.

From the results shown in Tables 1–4 we drew the following conclusions:

(a) In a ‘Rings’-type fitting, an angular rotation function gives better results for bridging joins, and is essentially equivalent to a distance function for performing single bond fusions (as would be expected).

(b) As shown in Table 4, using the set 2 atoms to calculate the rotation function for bridging joins gives much higher energy structures than using sets 1 or 3.

(c) When assigning coordinates with either fitting algorithm, taking average coordinates for all

TABLE 1
LEAST-SQUARES FITTING ALGORITHM (ENERGIES IN KCAL MOL⁻¹)

Allocation of coordinates	Weighting scheme	Join performed					
		A	B	C	D	E	F
Average for all common atoms	(1:1:1)	*	*	35.9	78.0	23.3	25.7
	(2:1.5:1)	*	*	35.7	77.2	23.3	25.8
	(10:5:1)	*	*	35.5	77.4	23.3	25.9
Preferential assignment to non-primary	(1:1:1)	29.8	33.1	17.8	27.9	22.3	24.3
	(2:1.5:1)	29.1	32.4	17.9	28.2	22.3	24.2
	(10:5:1)	28.2	31.7	18.0	28.5	22.2	24.1

*Indicates an extremely deformed structure (e.g. planar carbons, penetrating atoms, etc.) for which an MM2 energy is meaningless.

TABLE 2
SPIRO JOINS ('RINGS'-DERIVED ALGORITHM). (ENERGIES IN KCAL MOL⁻¹)

Allocation of coordinates	Join performed	
	E	F
Average for all common atoms	24.2	29.0
Preferential assignment to non-primary	21.4	23.4

the common atoms always produces higher energy structures than the alternative method of preferentially assigning the coordinates of a non-primary atom over those of a primary atom.

(d) No single weighting scheme is consistently best for all joins performed.

These results were then used in an investigation of wider scope, in which all possible combinations of the cycloalkanes from cyclobutane to cycloheptane in four different types of join (3-atom bridging, *cis*- and *trans*-fusion and spiro) were examined, a total of 40 experiments. For cycloalkanes which have more than one template in the WIZARD template library there is more than one pairwise combination of templates. In such cases a prescreen was performed using the 'Rings'-derived fitting algorithm (with unweighted functions) to identify the pairwise combination of templates which gave the structure of lowest energy. For example, in joining 6 and 5 rings chair cyclohexane and envelope cyclopentane were thus chosen instead of using twist boat cyclohexane and planar cyclopentane. We also performed this prescreen for a number of critical cases using the

TABLE 3
SINGLE BOND FUSIONS ('RINGS'-DERIVED ALGORITHM). (ENERGIES IN KCAL MOL⁻¹)

Rotation function	Allocation of coordinates	Weighting scheme	Join performed	
			C	D
Distance	Average for all common atoms	(1:1:1)	38.2	75.8
		(2:1.5:1)	38.2	75.8
		(10:5:1)	38.2	75.8
	Preferential assignment to non-primary	(1:1:1)	17.9	27.6
		(2:1.5:1)	17.9	27.6
		(10:5:1)	17.9	27.6
Angular	Average for all common atoms	(1:1:1)	38.2	75.8
		(2:1.5:1)	38.2	75.8
		(10:5:1)	38.2	75.8
	Preferential assignment to non-primary	(1:1:1)	17.9	27.6
		(2:1.5:1)	17.9	27.6
		(10:5:1)	17.9	27.6

TABLE 4
MULTIPLE BOND FUSIONS ('RINGS'-DERIVED ALGORITHM). (ENERGIES IN KCAL MOL⁻¹)

Atom set used	Rotation function	Allocation of coordinates	Weighting scheme	Join performed	
				A	B
1	Distance	Average for all common atoms	(1:1:1)	44.5	*
			(2:1.5:1)	37.7	*
			(10:5:1)	37.7	*
		Preferential assignment to non-primary	(1:1:1)	33.9	36.0
			(2:1.5:1)	31.6	34.5
			(10:5:1)	31.6	35.2
	Angular	Average for all common atoms	(1:1:1)	30.6	*
			(2:1.5:1)	32.6	*
			(10:5:1)	37.7	*
		Preferential assignment to non-primary	(1:1:1)	28.7	31.8
			(2:1.5:1)	29.6	32.3
			(10:5:1)	31.6	35.2
2	Distance	Average for all common atoms	(1:1:1)	135.6	*
			(2:1.5:1)	135.6	*
			(10:5:1)	135.6	*
		Preferential assignment to non-primary	(1:1:1)	64.7	68.9
			(2:1.5:1)	64.7	68.9
			(10:5:1)	64.7	68.9
	Angular	Average for all common atoms	(1:1:1)	135.6	*
			(2:1.5:1)	135.6	*
			(10:5:1)	135.6	*
		Preferential assignment to non-primary	(1:1:1)	64.7	68.9
			(2:1.5:1)	64.7	68.9
			(10:5:1)	64.7	68.9
3	Distance	Average for all common atoms	(1:1:1)	26.9	*
			(2:1.5:1)	28.0	*
			(10:5:1)	37.7	*
		Preferential assignment to non-primary	(1:1:1)	26.6	30.6
			(2:1.5:1)	27.9	31.8
			(10:5:1)	31.6	34.5
	Angular	Average for all common atoms	(1:1:1)	26.3	*
			(2:1.5:1)	27.7	*
			(10:5:1)	37.7	*
		Preferential assignment to non-primary	(1:1:1)	25.5	30.1
			(2:1.5:1)	27.2	31.1
			(10:5:1)	31.6	34.5

*Indicates an extremely deformed structure (e.g. planar carbons, penetrating atoms, etc.) for which an MM2 energy is meaningless.

least-squares fitting algorithm. For these cases the same combinations of templates were chosen and so we are confident that by using the 'Rings' method for the prescreen we were not prejudicing the results in favour of this fitting algorithm. The 40 joins were then performed using the different methods, with the same three weighting schemes as before. The non-minimized energy of the final structure obtained in each case was calculated using MM2.

The results were analyzed to see which of the algorithms gave the lowest energy structure for each of the different joins, and whether any one algorithm consistently gave lower energy structures. The graph in Fig. 12 shows the difference in energy (in kcal mol⁻¹) between the best results obtained using the least-squares procedure and those obtained during the 'Rings'-derived method for the joins performed. For many of the joins the resulting structures from the two classes of fitting algorithm are close in energy. The major differences are found for bridging joins and for other joins which involve cyclobutane templates. In these cases the 'Rings'-derived method produces the lower energy structures.

It is, however, important to recognize the importance of the use of weighting schemes. In the study using just three weighting schemes only a small portion of the space of possible weighting coefficients was studied. In accordance with naive chemical intuition the type 1, 2 and 3 common

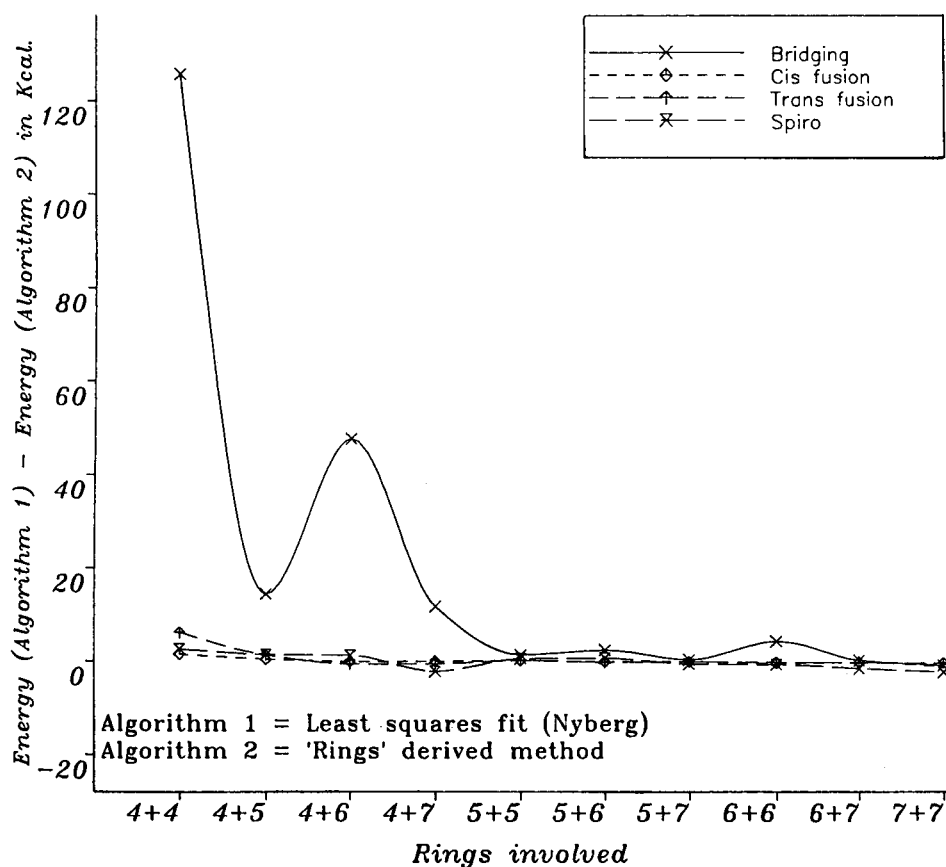


Fig. 12. Graph showing the difference between least-squares fitting and the 'Rings'-derived method.

atoms had been given descending priorities in the (2:1.5:1) and (10:5:1) weighting schemes. We thought we could see some pattern emerging about the type of weighting scheme to use in a particular type of join when using the least-squares method. Specifically, for spiro joins and *cis* single bond fusions the (10:5:1) scheme generally gave the best results; for *trans*-fusions the (2:1.5:1) scheme was pre-eminent and for bridging joins the (10:5:1) scheme was found to have given the best results in most cases. However, we thought it possible that as of then untried ratios might give better results.

We therefore decided to investigate more fully the scope of the least-squares fitting algorithm vis-à-vis the possible range of weighting coefficients, for several reasons. First, we wanted to see if we could produce any rules about the best weighting schemes to use with regard to join categories. Moreover, we were interested to see if the least-squares algorithm could produce structures comparable in energy to those obtained with the 'Rings'-derived algorithm for those cases in which the latter had given significantly better results. Finally, we wanted to make sure that the better performance of the 'Rings' algorithm shown in Fig. 12 was not simply an artifact of a badly chosen weighting scheme. Each of the 40 joins was repeatedly performed using the least-squares algorithm in a triply nested loop where each coefficient was allowed to vary from 1 to 100 in steps of 20. The ratio of any two coefficients thus varied from 1:100 to 100:1. The energy of each structure obtained was then calculated using MM2. With three such coefficients this gives a total of 216 different joins to be performed on each of the 40 pairwise combinations of templates. Furthermore, we analyzed several smaller regions of this space with much smaller step sizes. In none of these cases did we find sudden sharp local minima or maxima and are therefore fairly certain that the energy surface is of low curvature.

For each join the results were analyzed, using both graphical and numerical means. The results can be graphically illustrated using a diagram such as Fig. 13, in which the values along the three

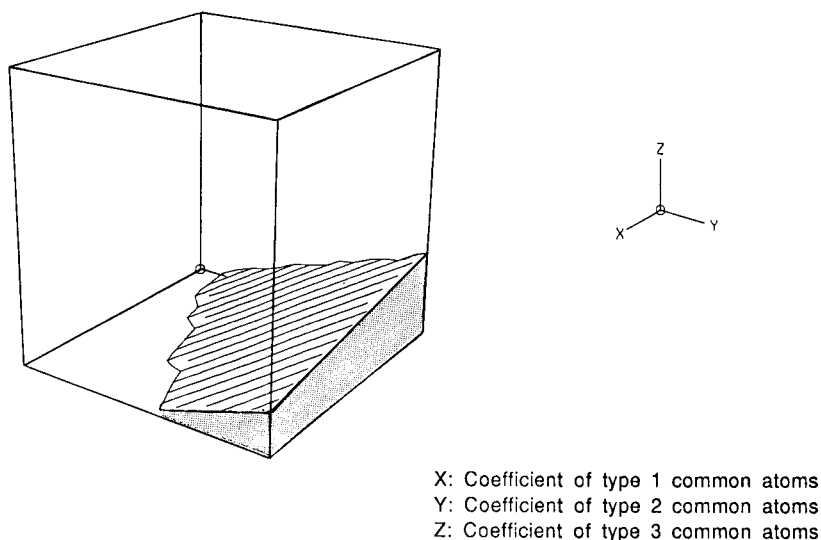


Fig. 13. Graphical representation of the variation in the energy of the final structure with the three weighting coefficients when using the least-squares method.

axes represent the weighting coefficients applied to the three types of common atom. The energy of the final combined structure is a function of the three weighting coefficients, which can be represented by drawing contour lines at a particular value on a series of parallel planes through the data, to produce a solid shape. All combinations of weighting coefficients which lie within the solid shape give rise to a structure which has an energy below the chosen level. The analysis revealed the combinations of coefficients which gave low energy structures, those which gave high energy structures, and the spread of energies from the best structure to the worst for each of the 40 joins performed.

For spiro joins the range of energies from the best structure obtained to the worst was small. In such joins there are usually only type 1 and type 2 common atoms, which makes the energy a function of two variables rather than three. We found that the lowest energy structures were obtained by giving a high weight to the single type 1 common atom, and a lower weight to the type 2 common atoms, the optimum ratio being of the order of 50:1. Using these weights all the spiro joins we investigated produced a final structure within 5% of the best result obtained overall. A high weight given to the type 1 common atom ensures that this atom will effectively be overlaid in the first step, and then the remaining common atoms are fitted during the subsequent rotation sequence. Indeed, for spiro joins the least-squares method with a suitable weighting scheme is probably more appropriate to use than the 'Rings' method, as the 'dummy' atom used in the latter does not really have any chemical basis. However, for the spiro joins which we investigated the difference between the algorithms was so small as to be insignificant.

For single bond fusions a weighting scheme of (80:20:1) produced final structures for both *cis*- and *trans*-fusions which were within 5% of the best result. With such a weighting scheme the high weighting coefficient given to the two atoms which form the common bond ensures that they are overlaid during the first step. The rotation then effectively occurs about this bond to fit the remaining common atoms. Consequently, the least-squares method and the 'Rings'-derived method can be rendered equivalent by a judicious choice of weighting scheme. The range of energies for the structures obtained is somewhat greater than for spiro joins, but not significantly so in most cases.

It is with bridging joins that the widest variation is observed in the resulting structures, and the choice of weighting scheme becomes most important. A general pattern which emerged was that the highest weighting coefficient should be given to the type 2 common atoms, and a very small one to the type 3 common atoms. The scheme (40:80:1) produces final structures for all the joins which we investigated in this category which are within 3% of the best value for all of the schemes tried.

For such joins it is not possible to directly reproduce the 'Rings'-derived method simply by choosing some suitable combination of weighting coefficients. It is interesting that we found the best combinations of coefficients for such joins had primarily higher weightings for the type 2 common atoms than for the type 1 common atoms. This can be understood if one examines the relative orientations of the two templates during the stages of the least-squares procedure, in the context of the way in which the coordinates are eventually allocated to give a combined structure. Figure 14 shows the final position of the two fragments for the bicyclo-[3.2.1]-octane case. The weighting scheme (40:80:1) has been used for Fig. 14A; Fig. 14B shows the effect of giving a higher weight to the type 1 common atoms. We had expected that the energy difference in the structures obtained would arise mainly from angle bending terms. However, an examination of the contribu-

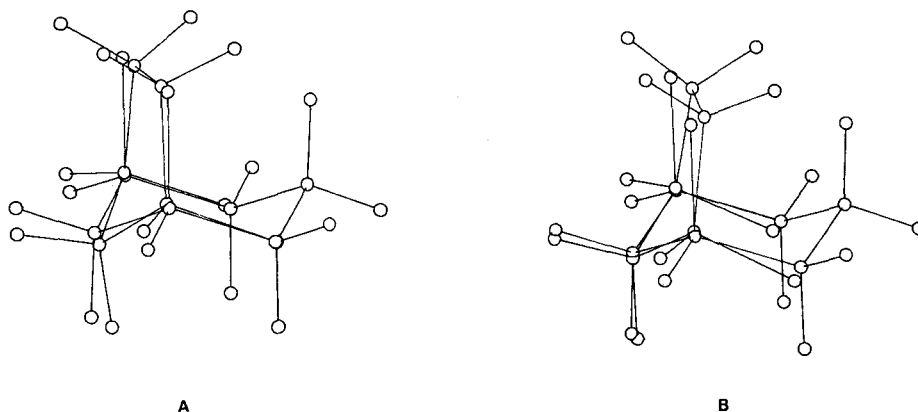


Fig. 14. (A) Least-squares fitting using (40:80:1) weighting scheme; (B) Least-squares fitting giving a higher weight to the type 1 common atoms.

tions to the total steric energy of the resulting structures revealed that the most significant energy term arises from higher H-H van der Waals interactions between the hydrogens on the 'bridge' and those on the other ring. The angle bending terms also make a (smaller) contribution to the energy difference. This is a general feature of the small ring joins we have investigated. In a more sophisticated joining algorithm it might be preferable to include some form of van der Waals calculation in recognition of this fact.

CONCLUSIONS

Thus it can be seen from the investigations described here that there is a wide variety of possible algorithms which can be used for the 'joining' of templates to build conformational models. Moreover, the choice of the basic algorithm and its variants can significantly affect the quality of the results obtained, and no one algorithm is superior for all cases. The comparison between a purely geometrical (RMS fit) method and one based on the way in which a chemist might perform the task shows that either method can produce acceptable results if the correct weighting schemes are used and the final coordinates are assigned selectively. However, we have found that no single weighting scheme produces uniformly good results when using the geometrical fitting algorithm. Thus we would suggest that programs which used such a basic algorithm recognize the type of join being performed and apply the appropriate weighting scheme. While our latest version of WIZARD utilizes a complex recognition algorithm to achieve the best joining (by taking into account factors not explored here, such as the cyclic vs. acyclic nature of the fragment), by simply counting the number of non-primary atoms and choosing a weighting scheme based on this number, results within 5% of the optimum in the majority of cases can be obtained.

ACKNOWLEDGEMENTS

ARL thanks the SERC for a quota award; DPD thanks Smith, Kline & French Research for financial assistance.

REFERENCES

- 1 Dolata, D.P. and Carter, R.E., *J. Chem. Inf. Comput. Sci.*, 27 (1987) 36–47.
- 2 Dolata, D.P., Leach, A.R. and Prout, C.K., *J. Comput.-Aided Mol. Design*, 1 (1987) 73–85.
- 3 CHEM-X, developed and distributed by Chemical Design Ltd., Oxford.
- 4 De Clerq, P.J., *Tetrahedron*, 40 (1984) 3717–3738.
- 5 Wipke, W.T. and Hahn, M.A., *ACS Symposium Series*, Vol. 306, American Chemical Society, Washington, D.C., 1986, pp. 136–146.
- 6 Cohen, N.C., Colin, P. and Lemoine, G., *Tetrahedron*, 37 (1981) 1711–1721.,
- 7 Liljefors, T., *J. Mol. Graph.*, 1 (1983) 111–117.
- 8 Van der Lieth, C.W., Carter, R.E., Dolata, D.P. and Liljefors, T., *J. Mol. Graph.*, 2 (1984) 117–123.
- 9 Allinger, N.L., *J. Am. Chem. Soc.*, 99 (1977) 8127–8134.
- 10 Murray-Rust, P. and Motherwell, S., *Acta Crystallogr., Sect. B*, 34 (1978) 2534–2546.
- 11 Murray-Rust, P. and Raftery, J., *J. Mol. Graph.*, 3 (1985) 50–56.
- 12 Taylor, R., *J. Mol. Graph.*, 4 (1986) 123–131.
- 13 Mackay, A.L., *Acta Crystallogr., Sect. A*, 40 (1984) 165–166.
- 14 Nyberg, S.C., *Acta Crystallogr., Sect. B*, 30 (1974) 251–253.