

Computer Manipulation of Central Chirality

By **Ashmeed Esack and Malcolm Bersohn**,* Chemistry Department, University of Toronto, Toronto, Canada
M5S 1A1

Any general method for ordering canonically the atoms of a molecule is also a method for defining the chirality of the atoms, just as the Cahn–Ingold–Prelog sequence rules can be used as a general method for numbering the atoms of a molecule. Use is made of this fact to devise techniques whereby a computer program dynamically determines the chirality of the atoms of a stereoselected product (reactant) knowing the stereochemistry of the reactant (product). These techniques are essential for noninteractive programs that generate synthetic routes to optically active molecules.

THE global problem of the canonical numbering of the atoms of a molecule is closely related to the local problem of describing chirality at a chiral centre. In both cases the atoms must be ordered according to various criteria. We have found that it is necessary to integrate these two problems fully in order to develop methods for a computer

program to use the concepts of *R* and *S*,¹ clockwise and counterclockwise, *etc.*

Looking over various conceivable synthetic routes in order to generate optimal ones, a non-interactive computer program may have to generate tens of thousands of intermediate molecules. In a backward search the structure of each of these molecules must be compared with the structure of any isomers which may be known to the program as available to determine if a successful route has been generated. In a forward search the structures of all intermediate molecules which are isomers of the goal molecule must be compared with that of the goal molecule to decide if a synthetic route has been generated. In either direction it is necessary that the chirality be correct at each stage for each molecule. Yet, in a non-interactive program the chemist can only indicate the chirality of the goal molecule and of the available substances. Somehow the computer program must be able to deduce the chirality of each molecule knowing only the chirality of the atoms of the previous molecule(s) and the description of the stereochemical changes required in each reaction. Our non-interactive synthesis program² now has routines in it which perform this task in both the forward and backward directions. This paper is devoted to a discussion of the concepts and techniques involved. To avoid clumsy verbiage and repetition we will assume herein that we are talking about deriving the chirality of the atoms of a product from the stereochemistry of the reactant(s) giving rise to this product. The backward case, *i.e.* deriving the chirality of the atoms of a reactant, knowing the stereochemistry of the product, is largely analogous. The first implementation of stereochemical manipulation in a computer program is that described by Wipke and Dyott.³ Their representation of the stereochemistry is extremely concise and saves storage. Our representation was designed to save time in retrieval.

Methods for the Canonical Ordering of the Atoms of a Molecule.—A canonical ordering of the atoms of a molecule is obtained by a general procedure the results of which are independent of the original numbering of the atoms. In order to attain the canonical ordering, a certain number must be stored for each non-hydrogen atom of the molecule. This number must be different for each pair of chemically non-equivalent atoms and equal for each pair of chemically equivalent atoms. Let us call these numbers 'canonical values.'

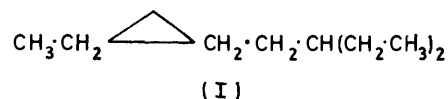
For a molecule like CHFIBr it would be sufficient to take the atomic number of each atom as its canonical value. In the Cahn-Ingold-Prelog rules for ordering the neighbouring atoms of a chiral atom,¹ the implicit canonical value is determined by the atomic numbers of nearest neighbours and remote neighbours. In their historic scheme for defining chirality, now universally accepted, two atoms are unequal if their atomic numbers are unequal or if those of their neighbours are unequal.

¹ R. S. Cahn, C. K. Ingold, and V. Prelog, *Angew. Chem. Internat. Edn.*, 1966, **5**, 385, 511.

² M. Bersohn, *Bull. Chem. Soc. Japan*, 1972, **45**, 1897.

This search for inequality can take us to the most remote atoms of the molecule. Hence the canonical values can become very long strings.

Morgan's method for obtaining canonical connection table representations of molecules⁴ involves among other things the computation of canonical values which depend not on atomic number but on the number of nearest neighbour non-hydrogen atoms that each atom possesses. There are an infinite number of valid procedures for finding a canonical order for the atoms of a molecule, differing as to which function of what combination of atomic properties is used. These procedures differ as to efficiency. If we are looking for a fast scheme the standard chirality sequence rules do not suggest themselves as a likely basis since the canonical values may be long strings, time consuming to store and compare. Also, other things being equal, it is faster to sum nearest neighbour's previous canonical values as a method for receiving information from further away in the molecule. This saves us time. In computing a canonical value for an atom 1, whose neighbours are atoms 2, 6, and 7, for example, by this method we do not have to identify the neighbours of atoms 2, 6, and 7. Finally, it would be advantageous to iterate as few times as possible. Consider the molecule (I). For most



canonicalization schemes four iterations would be necessary, the last serving merely to confirm the results of the third. If, however, members of three-atom rings were labelled differently from ordinary atoms not in such a ring then one less iteration is required. This illustrates the important idea that, other things being equal, the more information that is contained in the initial canonical value for each atom the sooner will the process terminate. The initial canonical values that we generate in our synthesis program are designed to be as rich in information as possible so as to make use of this principle and minimize the number of iterations. For example we label the 'central atoms' of functional groups⁵ differently. Consequently our canonicalization routing requires one less iteration for, *e.g.* the monomethyl ester of succinic acid than does Morgan's method.⁴ The time of our canonicalization routine for cholesterol is 2.4 ms on a computer (the IBM 370/165) with a store execution time of 0.32 μ s.

The representation of molecular structure currently implemented in our synthesis program is described in the Appendix in detail. Here we note the following. The information about each atom is divided into two discrete parts, the first containing all the connectivity information and the second only stereochemical information about the atom. The first part is an

³ W. T. Wipke and T. M. Dyott, *J. Amer. Chem. Soc.*, 1974, **96**, 4825, 4834.

⁴ H. L. Morgan, *J. Chem. Documentation*, 1965, **5**, 107.

⁵ A. Esack and M. Bersohn, *J.C.S. Perkin I*, 1974, 2463.

appropriate row in the connection table describing the molecule.⁶ Each atom of an element other than hydrogen has its own row in the connection table. Each row consists of two 32-bit computer words. The first word contains such features as the atomic number, the charge, and connectivity information such as the degree of unsaturation, the number of adjacent hydrogen atoms, and the number of atoms in the ring(s) of which the atom is a member. In the last two bits of this word *R*- or *S*-chirality is also indicated. *R*-Chirality means that the first three neighbours referred to in the connection list (second word of the row) are arranged in clockwise order when viewed from the side opposite the fourth neighbour. *S*-Chirality has the opposite meaning. The Appendix (Table 2) gives our canonical representation of a typical steroid. The connection table thus contains all the labelled graph properties of the molecule. The three-dimensional features of the molecule other than chirality are noted in a stereochemical table where the information is stored in considerably redundant fashion. The usual categories are 1, axial to some ring; 2, equatorial to some ring; 3, the atom in question bears a hydrogen atom which is *cis* to a certain specified atom; 4, the corresponding *trans* property; 5 and 6, *cis* and *trans* in the double bond sense; 7 and 8, *cis* and *trans* in the other sense; 9 and 10, *cis* and *trans* ring junctions; 11 and 12, *cis* and *trans* relationships between attached hydrogen atoms. Evidently this is not the irreducible minimum number of statements which can unambiguously describe the stereochemistry of a molecule. The redundancy is arranged so that if, for example, a reaction requires X and Y to be *cis* or *trans* to each other then the fact of this relationship will already be explicitly noted and the execution time of the program in performing simulated reactions is shortened correspondingly.

The canonical values in our system consist of two parts, the first being the constitutional or 'graph' part, which utilizes all the connectivity information as well as the atomic number and charge. In other words, it utilizes all the information of the first 30 bits of each row of our connection table. The chirality indication, stored in the last two bits of the 32-bit word, is not used in the computation of the canonical value since before canonicalization it refers to the chirality of the atom either when it was part of another molecule, a reactant, or when the molecule was canonicalized according to some different system. We say that atom A precedes atom B in the canonical ordering if the first 30 bits of the row of the connection table describing atom A have a higher numerical value than that of B. If the two values are equal then we examine the graph parts of the canonical values. If these are equal then the stereochemical parts of the canonical values determine the precedence. As will be evident later, it is normally unnecessary to calculate the stereochemical part of the canonical value. Only rather symmetrical molecules with constitutionally equivalent

double bonds or constitutionally equivalent chiral centres require the stereochemical canonical value.

The Parity Rule.—The list of one to four atoms numbers that appears in the right-most portion of each row of the connection table refers to the one to four non-hydrogen nearest neighbours of the atom. The canonical row numbers of these neighbours are to be arranged in ascending order. Consider the example of (*R*)-butan-2-ol. The connection table is Table 1. Inputs which arrange the nearest neighbours of atom 1 in the orders

TABLE 1
Connection table for (*R*)-butan-2-ol

8	0	3	0	00000000	1	3	1	2	4
6	0	3	7	00000010	0	2	3	CH ₃ ·CH·CH ₂ ·CH ₃	
6	0	2	0	00000000	1	4			
6	0	1	0	00000000	1	OH			
6	0	1	0	00000000	2				
						0			

3,0,2 or 2,3,0 are also acceptable descriptions of the *R*-enantiomer because these lists can be cyclically permuted into the correct order 0,2,3. We use the parity rule of Petrarca, Lynch, and Rush.⁷ They state that when four tetrahedral ligands are written out in linear order any even permutation of the linear list preserves the indicated chirality and any odd permutation of the ligands inverts the chirality. This principle is also used by Wipke and Dyott.³ The conventions in regard to the spatial relationships implied by the linear list may differ. Wipke and Dyott always understand that the first three ligands are grouped in clockwise fashion when viewed from the side opposite the last neighbouring atom. In our notation the first three atoms may have either a clockwise or a counterclockwise arrangement when viewed from the side opposite the last neighbouring atom. If the arrangement is clockwise the next to last (31st) bit of the first word of the connection table row describing the chiral atom is given the value 1. If the arrangement is counterclockwise the 32nd and last bit of this word is given the value 1. If there is no chirality both these bits are given the value zero.

The parity rule of Petrarca, Lynch, and Rush gives us a ready way of interconverting any two systems of chirality. If the canonicalization causes an odd permutation of the original list of adjacent atoms then the chirality of the atom concerned in the new system is opposite to that in the old system. Otherwise the chiralities are the same. Suppose for example that in one system the neighbours of atom 1, the chiral atom in butan-2-ol are given in the order 1,2,4, and we are told that this is in clockwise order. Suppose, further, that on recanonicalizing, *e.g.* to our system, the very same atoms are now referred to as 3,2,1. The permutation to our canonical ordering 1,2,3 is odd; hence we have to say that in this case the '*R*'-chirality of the other system is the same as '*S*'-chirality in our system.

In our program, inversion is accomplished by inverting the chirality bits (for a chiral atom either the *R* bit

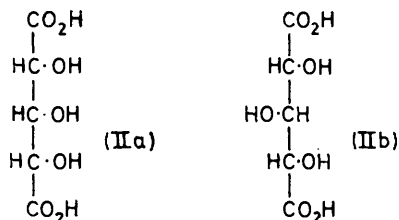
⁶ M. F. Lynch, J. M. Harrison, W. G. Town, and J. E. Ashm, 'Computer Handling of Chemical Structure Information,' Macdonald, London, 1971.

⁷ A. E. Petrarca, M. F. Lynch, and J. E. Rush, *J. Chem. Documentation*, 1967, 7, 154.

is 1 and the *S* bit has the value 0 or the *R* bit has the value 0 and the *S* bit the value 1) not by changing the order of presentation of the row numbers of the neighbouring atoms. In addition, the entering atom adopts the stereochemistry relations of the leaving atom except in reverse, *i.e.* axial becomes equatorial, *cis* becomes *trans*, *etc.* Furthermore, atoms which were *cis* to the leaving group now must be marked as *trans* to the entering group, *etc.* These latter changes all are made in the stereorelation table of the molecule.

Equivalent Doubly Bonded Atoms.—We consider the case where the graph parts (connectivity parts) of the canonical values of one pair of doubly bonded atoms are the same as those of another pair of doubly bonded atoms but the four atoms (three in the case of an oxime) attached to the first pair are, because of *cis-trans* isomerism, not equivalent to the corresponding four atoms attached to the other pair. In this case the double bond *cis-trans* values stored in the record of each atom concerned must be used in the calculation of canonical values. A separate stereochemical part of the canonical value is set up which is initially zero for all atoms not part of or adjacent to such double bond pairs. The stereochemical parts of the canonical values rank after the connectivity parts in the hierarchy of comparison for determining the canonical order of the atoms of the molecule.

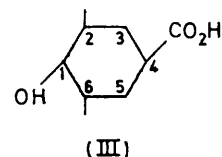
Equivalent Chiral Centres.—If two chiral centres have the same canonical value, as calculated up to this point, then they are constitutionally equivalent. Constitutionally equivalent chiral centres of opposite chirality should not be treated by the program as equivalent. Consider for example the case of the *meso*-glutaric acids (IIa and b) discussed by Mislow.⁸ If carbon atoms 2 and 4 are considered to be equivalent then the pseudo-asymmetric carbon atom 3 will be considered to be



achiral. The result would be a failure of the program to distinguish the two *meso*-isomers. Accordingly we use the following rule. If there is more than one constitutionally equivalent *RS*-pair then choose the pair with greatest canonical value. Then add one to a new quantity called the chiral part of the canonical value of the *S* atom of the pair. The chiral parts of the canonical values are all originally zero. We then resume the process of iterative generation of canonical values, computing the chiral part only, stopping when, as before, the number of equal signs does not decrease between two successive iterations. In the comparison for determining the canonical order of the atoms the chirality parts of the canonical values rank after the connectivity

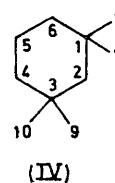
and stereochemical parts. We have here a two-step process. Chirality values are first assigned according to the constitutional sequence of the ligands. They are then revised in the special situation of equivalent chiral centres of opposite chirality.

If there remain any equivalent *RS*-pairs we select the pair with the highest canonical value and repeat the process of the preceding paragraph. After there are no more *RS*-pairs regarded as equivalent then the remaining *RR*- and *SS*-pairs still regarded as equivalent usually are truly equivalent. The program takes note of one exception to this statement. (Other exceptions will be incorporated when discovered.) If constitutionally equivalent chiral centres are members of the same ring then we must test for a situation like (III),



cited by Cahn, Ingold, and Prelog.¹ Here the constitutionally equivalent atoms 2 and 6 are actually different because their interchange at atom 1 produces a different isomer. In such a case the program uses all the stereochemical information for these atoms such as axial, equatorial, *cis* and *trans* in the ring, *etc.*, to calculate the stereochemical parts of the canonical value. If the inclusion of this information, which differentiates the two atoms, reveals at least two other atoms to be chiral (atoms 1 and 4) then the new stereochemical parts of the canonical values are adopted and the chirality of the other atoms is recognized. Otherwise we revert to the previous set of canonical values.

Chiral Centres Formed with the Direction of Other Chiral Centres.—Certain reactions involve a preferential approach from a relatively unhindered side of a ring. Some reactions give only axial or equatorial products specifically. If we use a chiral atom already present in the reactant as a reference atom then we have, as it were, a compass so that a program can tell what is meant by clockwise when it is examining a newly created chiral centre in the same ring. If we know that entering atom A must be *trans* to some angular methyl group attached to the same ring of a steroid, for example, the calculation of the chirality of the atom attached to A is straightforward. The general technique is as follows. (We use



a six-membered ring in the example but the technique can be used for any ring size.) Suppose that in the

⁸ K. Mislow, 'Introduction to Stereochemistry,' Benjamin, New York, 1966.

molecule (IV) the atoms numbered 7, 2, and 6 are arranged in clockwise fashion when viewed from the side opposite atom 8. We are told in the program that the entering atom, 10, must be *trans* to atom 7. The problem is to

program could find the same chirality result by inspection of its table for the six-membered ring case. The table tells the program that a 1,2-diaxial relationship implies a *trans* relationship, 1,3-equatorial-axial implies *trans*,

Representation of Molecular Structures

TABLE 2

(a) Format for a row of a connection table

First word

Column no.	1	2	3	4	5	6
	Atomic no.	Unsaturation	Four minus the number of attached hydrogen atoms	Serial no. of the functional group ⁵ if the atom is central to the group, otherwise, zero	Atom is in a ring of size other than 3, 5, or 6	Atom is in a three-membered ring
Column no.	7	8	9	10	11	12
	Atom is in a five-membered ring	Atom is in a six-membered ring	Plus charge on atom	Minus charge on atom	R-Chirality	S-Chirality

Unsaturation value is 0 for a saturated atom, 1 for an aromatic atom, 2 for a carbon atom doubly bonded to a carbon atom, 4 for a doubly bonded heteroatom or a carbon atom doubly bonded to a heteroatom, or 8 for a triply bonded atom. Columns 5–12 have the values 1 or 0 depending respectively on the truth or falsity of the statement applicable.

Second word

Column no.	13	14	15	16
	Row number of LIGAND 1	Row number of LIGAND 2 if it exists, else a blank	Row number of LIGAND 3 if it exists, else a blank	Row number of LIGAND 4 if it exists, else a blank

Doubly bonded ligands appear in two column. Triply bonded ligands appear in three columns. For butan-2-one, the connection table is:

Columns	1	2	3	4	5–12	13	14	15	16
	8	4	4	0	00000000	1	1		
	6	4	4	19	00000000	0	0	2	3
	6	0	2	0	00000000	1	4		
	6	0	1	0	00000000	1			
	6	0	1	0	00000000	2			

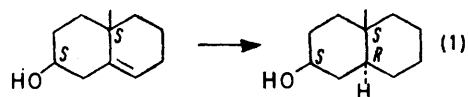
The atoms are numbered from 0 to 4. The number 19 indicates a ketone carbon.

(b) Format for a row of a stereochemical table

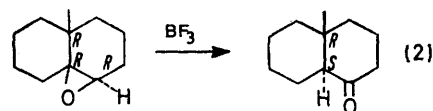
There are four columns. In special cases, this can be expanded to eight columns. Each entry is a pair of symbols. The first symbol is a number indicating a stereochemical relationship. The second symbol is the row number of the atom that is sterically related to the atom described in the row. If there are less than four stereochemical relationships to note, the rightmost columns are left blank. As an example, we give the row 1 : 5, 7 : 9, 8 : 13. This means that the atom in question is axial in the ring in which the atom with row number 5 is located. It is *cis*, *i.e.* 7, to atom 9 and *trans*, *i.e.* 8, to atom 13. The possible values of the first number of the pair and their meaning are as follows: 1, axial; 2, equatorial; 3, an attached hydrogen atom is *cis* to some other specified atom; 4, the same as 3 except that the meaning *trans* is understood; 5, *cis* in the double bond sense; 6, the corresponding *trans* relationship; 7 the atom is *cis* to a specified atom; 8, *trans*; 9, the atom is part of a *cis* ring junction with a specified atom; 10, a *trans* ring junction atom; 11 and 12, a hydrogen atom attached to the atom of the row is respectively *cis* or *trans* to the attached hydrogen atom of a specified atom.

find the chirality at atom 3. The program has previously made a sequential list of the atoms of the ring and it deduces from this list that the pair 4 and 2 have the same relationship to atom 3 that the pair 2 and 6 have to atom 1. Also since atom 10 is *trans* to atom 7 it must be in the position relative to 3 that atom 8 is to atom 1. If the nearest neighbours of atom 1 are listed as 2, 6, 7, and 8, then the corresponding atoms in the environment of atom 3 are respectively 4, 2, 9, and 10. Since an odd permutation is required to bring the list 4, 2, 9, and 10 into its canonical order 2, 4, 9, and 10, it follows that the arrangement of the atoms 2, 4, and 9 viewed from the side opposite atom 10 must be opposite to that of the known clockwise arrangement of the atoms 2, 6, and 7 viewed from the side opposite atom 8. Hence the chirality at atom 3 must be *S*. If instead of knowing that atom 10 was *trans* to atom 7 we simply knew that atom 7 was axial and atom 10 had to be equatorial, the ring being part of a fused polycyclic ring system, the

etc. Below are two actual examples [(1) and (2)] performed by our program.

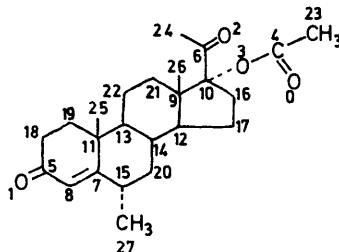


The chirality of the new centre is correctly calculated by the program to be *R* in our system. (It also happens to be *R* in the Cahn-Ingold-Prelog sense.) The program only knew that the hydrogen atoms were added on the more hindered side of the ring.



The chirality at three centres of the reactant was correctly deduced. The program only knew that the hydrogen atom entering has the stereochemistry opposite to that of the leaving oxygen atom.

TABLE 3

(a) Connection table of 17 α -acetoxy-6 α -methylpregn-4-ene-3,20-dione

Row no.	Column no.				5-12	13	14	15	16
0	8	4	4	0	00000000	4	4		
1	8	4	4	0	00000000	5	5		
2	8	4	4	0	00000000	6	6		
3	8	0	4	0	00000000	4	10		
4	6	4	4	21	00000000	0	0	3	23
5	6	4	4	19	00010000	1	1	8	18
6	6	4	4	19	00000000	2	2	10	24
7	6	2	4	0	00010000	8	8	11	15
8	6	2	3	0	00010000	5	7	7	
9	6	0	4	0	00110001	10	12	21	26
10	6	0	4	0	00100010	3	6	9	16
11	6	0	4	0	00010010	7	13	19	25
12	6	0	3	0	00110001	9	14	17	
13	6	0	3	0	00010001	11	14	22	
14	6	0	3	0	00010010	12	13	20	
15	6	0	3	0	00010001	7	20	27	
16	6	0	2	0	00100000	10	17		
17	6	0	2	0	00100000	12	16		
18	6	0	2	0	00010000	5	19		
19	6	0	2	0	00010000	11	18		
20	6	0	2	0	00010000	14	15		

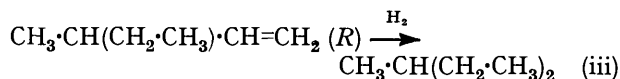
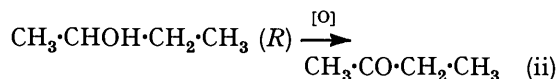
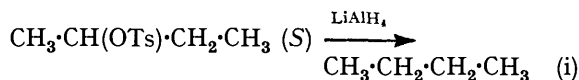
Row no.	Column no.				5-12	13	14	15	16
21	6	0	2	0	00010000	9	22		
22	6	0	2	0	00010000	13	21		
23	6	0	1	0	00000000	4			
24	6	0	1	0	00000000	6			
25	6	0	1	0	00000000	11			
26	6	0	1	0	00000000	9			
27	6	0	1	0	00000000	15			

(b) Nonblank rows of the stereocorrelation table of the molecule

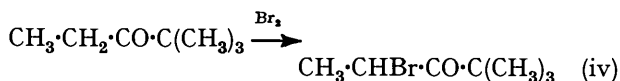
Row no.									
3	8	26							
3	5	11							
6	7	26							
9	10	12							
11	5	5							
12	10	9	12	14					
13	4	25	10	14			12	14	
14	4	27	10	13			12	12	
15	3	25						12	13
25	1	11	8	27					
26	1	9	7	6			8	3	
27	2	15	8	25					

Atoms 5 and 6 are ketone central atoms; 4 is an ester central atom.

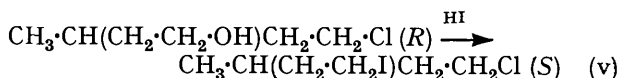
The Removal and Recognition of Chirality.—We now give some examples [(i)—(v)] of the program's treatment of chirality removal and chirality recognition. In cases



(i)—(iii) the program removes the chirality indicated. In (iv), since no stereospecificity is indicated the program



recognizes an enantiomeric pair and cuts the yield in half. In (v) the program labels the chirality of the chiral atom of



the product as S because the canonical order of the two

methylene carbon atoms adjacent to the chiral atom has been reversed by the chemical reaction.

Discussion.—Our representation of stereochemistry requires redundant notes, e.g. A is *cis* to B and B is *cis* to A. Redundancy is used to add to the speed of execution because the various steric relationships of interest in considering the feasibility of reactions are calculated just once and are available thereafter for as long as the program is concerned with the molecule. Speed is a *sine qua non* to a non-interactive synthetic program, so it is unacceptable to derive the information that A is *cis* to B from other information more than once.

We expect wide application of the technique described in this paper for using reference chiral centres unaffected by the reaction to determine the chirality of newly created centres in the product, or, in the backward search, to determine the chirality of chiral centres of the reactant which are made achiral by the reaction. In addition the segregation of the canonical values into connectivity and stereochemical parts should be a helpful aid for programs to use in 'thinking' about stereochemical problems.

APPENDIX

We thank the National Research Council of Canada for financial support.

[4/2176 Received, 21st October, 1974]