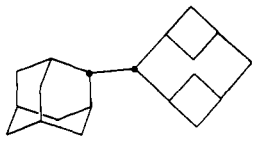


V

number. If the Q's really were equivalent, the choice would be immaterial, but as they are not, one may define two different "unique" numberings based on this initial arbitrary decision. The "proof" of the algorithm offered by Jochum and Gasteiger contains the implicit assumption that their ad hoc method of symmetry perception is accurate. If this were true, I believe their canonicalization algorithm could be proven correct, though a more detailed and precise chain of reasoning would need to be given.

CONCLUSIONS

The algorithm of Shelley and Munk fails in some cases because it does not consider the complete topological environment (i.e., the whole molecule) of each atom in computing scores. The algorithm of Jochum and Gasteiger does consider the whole molecule as "viewed" from a given atom, but in cutting the molecule into shells about that atom, it discards important connectivity information which may be needed to distinguish between nonequivalent atoms. It is my belief that a provably reliable algorithm for identifying symmetrically equivalent atoms must either explicitly compute the total symmetry group of the molecule or carry out a full atom-by-atom, bond-by-bond comparison of the total topological environments of atoms being compared.

ACKNOWLEDGMENT

I wish to thank both the Science Research Council and the National Institutes of Health (RR 00612) for their support of my current research.

REFERENCES AND NOTES

(1) Author's address until May 1, 1978.
(2) C. A. Shelley and M. E. Munk, *J. Chem. Inf. Comput. Sci.*, **17**, 110 (1977).
(3) C. Jochum and J. Gasteiger, *J. Chem. Inf. Comput. Sci.*, **17**, 113(1977).
(4) The term "symmetry" is used throughout this paper to refer to the topological symmetry of chemical graphs, rather than the geometrical symmetry of three-dimensional molecules. Topological symmetry is derived from the connectivity between atoms without reference to their spatial relationships.
(5) D. H. Smith and R. E. Carhart, *Tetrahedron*, **32**, 2513(1976), and previous papers in the series.
(6) R. E. Carhart, D. H. Smith, H. Brown, and N. S. Sridharan, *J. Chem. Inf. Comput. Sci.*, **15**, 124(1975).
(7) H. L. Morgan, *J. Chem. Doc.*, **5**, 107(1965).
(8) The accuracy of the Morgan canonicalization scheme is not in question here; it uses the atom-scoring algorithm in a theoretically sound manner to limit the number of potentially canonical representations which need be considered for a given structure. Rather than assuming that atoms with the same final score are symmetrically equivalent, Morgan's approach uses the fact that atoms with *distinct* final scores *cannot* be symmetrically equivalent.
(9) C. A. Shelley, private communication.

Computer Design of Synthesis in Phosphorus Chemistry: Automatic Treatment of Stereochemistry

F. CHOPLIN,* R. MARC, and G. KAUFMANN

Laboratoire de Modèles Informatiques appliqués à la Synthèse (ERA 671) ULP, Institut Le Bel, 67000 Strasbourg, France

W. T. WIPKE*

Department of Chemistry, University of California, Santa Cruz, California 95064

Received January 9, 1978

Computer treatment of the stereochemistry of phosphorus compounds is studied in the aim of building a program for the design of synthesis involving this type of chemistry. A naming algorithm has been developed to differentiate the isomers of pentacoordinated species. Techniques have been developed to conveniently and unambiguously describe the stereochemical changes which may occur at a pentacoordinated species during a reaction. Relative stabilities and possible intramolecular rearrangements between stereoisomers are also considered.

INTRODUCTION

The last several years have witnessed very rapid and outstanding progress in the field of computer design of synthesis. All previously reported work has been confined exclusively to the field of organic synthesis.[1a-i] However, some other areas of synthetic chemistry, such as those relevant to organophosphorus compounds, exhibit features which would expectedly permit their treatment by currently employed computer programs. Indeed, the theoretical and practical importance of these compounds makes application of computer-assisted synthesis techniques desirable. The fact that their synthesis can be designed with key steps involving general reactions makes a computer treatment feasible. If a program designed for organic synthesis could accomodate organophosphorus compounds, it would demonstrate the versatility of such a program. Moreover, many of the problems involved in organophosphorus chemistry are different from those encountered in organic chemistry; thus algorithms and strategies designed to solve these new problems will lead to a broader approach of the whole field. As a part of a project aimed at the development of the SECS program[2] for the design of organophosphorus synthesis, this paper describes the computer treatment of phosphorus stereochemistry. New problems arise owing to the different possible configurations about this atom, and to intramolecular rearrangements, which reduce the number of stable isomers. We shall describe first the un-

0095-2338/78/1618-0110$01.00/0

ambiguous graphic representation of any molecule containing a phosphorus atom and then the computer analysis of such structures in order to obtain the internal representation. The SEMA naming algorithm of the SECS program[3] has been extended so that it can treat the configurations of a pentacoordinated atom. Specific routines and commands have been developed to handle and display all the stereochemical changes which occur in reactions involving pentacoordinated phosphorus. Finally, in order to recognize identical or interconvertible structures, the possible intramolecular rearrangements between isomers are evaluated.

TECHNICAL DETAILS

The version of the program presently being developed has been adapted from version 2.0 of SECS.[3] It has been implemented on a high-speed UNIVAC 1110 machine (128 K of 36 bits words, basic cycle = 0.75 μs), the graphic input and output being made via a DEC PDP 11/10 minicomputer (8 K of 16 bits words), monitoring a graphic display (GT 40). A standard teletype, a light pen, and the switch register of the PDP11 are used as interactive devices. The graphic software was initially written at the Computer Center of the Copenhagen University,[4] and adapted to the UNIVAC available at the Centre de Calcul de Strasbourg-Cronenbourg.[5] It includes Fortran routines which reside in the UNIVAC for the creation of graphic commands, PDP11 routines to interpret the commands, and an ALGOL compiler for the translation of PDP11 assembly language via the UNIVAC 1110.

GRAPHIC REPRESENTATION OF MOLECULES

The first problem to be solved is that of providing the structural data to the computer. As previously seen,[6] the graphic input of structural diagrams is very convenient, and the specifications of UP and DOWN for a bond have been found sufficient to accurately describe an sp^3 carbon. The external representation of tri- and tetracoordinated phosphorus is achieved in the same way. For trivalent species, care must be taken to accept secondary phosphorus atoms as potential stereocenters. In tetracoordinated ones, the specification UP or DOWN must be given to a P=X double bond.

For a pentacoordinated phosphorus atom, different possible spatial environments are encountered, corresponding to a trigonal-bipyramid (TBP) (which can be distorted), a square-basis pyramid (SBP), or a X°TR geometry.[7] The latter is still open to discussion. Although the SBP geometry occurs in some cases,[8] stereochemical problems involved in reactions dealing with pentacoordinated phosphorus can generally be solved using the TBP hypothesis;[9] consequently, we use a TBP notation for input. The program will accept a SBP input structure,[10] but currently it will not analyze it, and the spatial environment of the atom will be considered as unimportant in the next steps. As very few phosphorus compounds exhibit a hexacoordinated center,[11] octahedral geometry is not considered in this work. However, as we shall show later, the algorithms developed for the treatment of the TBP can be extended easily to any kind of geometry.

The accepted drawings of a pentacoordinated atom are represented in Figure 1 (upper part). Once these representations have been analyzed by the program, the apical and equatorial bonds are recognized (see next section), and the structure is displayed again, the apical bonds being shown by dashed lines as indicated by Figure 1 (lower level). The chemist, when entering the structures, is required to draw the two axial bonds roughly on the same straight line (the angle between the two bonds must be greater than 165°). If no stereochemical specification is given in the input, the program concludes that stereochemistry is unimportant, and it will not be taken into account in later synthetic steps.

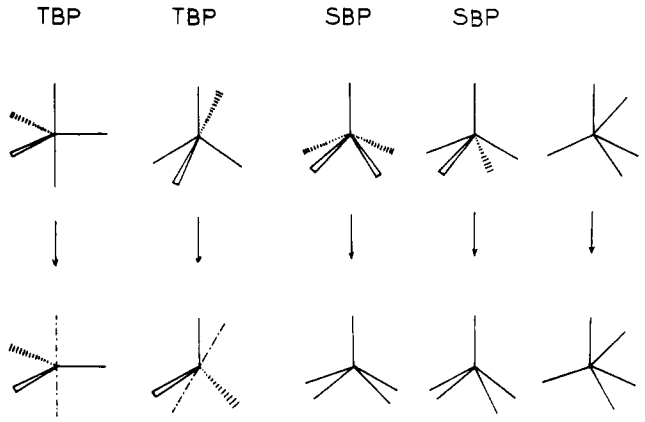


Figure 1. The upper level shows the drawings entered by the chemist for a pentacoordinated structure. The lower level shows the same structure displayed by the program after perception of the stereochemistry.

COMPUTER REPRESENTATION

Once the graphic conventions necessary to draw unambiguously a structure on the display have been defined, the program must be able to interpret this representation in order to store all the revelant information. As has been described previously,[6] a structural diagram is internally represented in the computer by a connection table (CT) which contains all the atomic, topological, and relevant bond information. As the structure is entered graphically, each atom is given in sequence an arbitrary index which we call *input number*. The input numbers of the attachments to an atom are stored in the form of an ordered list which represents the spatial configuration around this atom by mapping these attachments onto positions of a specified reference skeleton. Implicit H's bonded to a carbon atom are not stored but can be imagined to be at the end of the attachment list.[6] For a secondary phosphine group, the lone pair similarly is not explicitly represented but can be imagined to be at the end of the list, which thus contains only two entries. The phosphorus atom of this group is considered as representing a potential stereocenter. Many representations using either matrices or ordered lists have been proposed for a TBP center.[12] The convention described by Ramirez[13] was selected as our model for a TBP: the three equatorial atoms are arranged in a clockwise order when seen from one apical atom looking up toward the second apical atom. In the program, the list attachments for a TBP center is arranged so that the first three entries represent equatorial positions and last two entries represent the two apical positions. Since phosphorus chemistry often involves apical attack or departure of an atom, this organization of the list simplifies the symbolic operations required to describe such reactions, and is consistent with SECS stereochemical notations.[6]

Although we permit implicit hydrogens when attached to tri- and tetracoordinated atoms, for pentacoordinated species we require any attached hydrogens to be represented explicitly. We prefer not to have "holes" in the middle of the ordered list which would result from an implicit equatorial hydrogen (generally the favored position for hydrogen[14]). Except in a few cases such as H_2PF_3, pentacoordinated phosphoranes containing a P–H bond have only one such bond, so explicit representation is not inconvenient.

Recognition of the apical and equatorial orientation of the ligands from the input diagram of a TBP is achieved in the following way (see Figure 2): the program first looks for the largest dot product $Ti \cdot Tj$ (where i and j designate any atom bonded to the TBP center T). This gives the apical atoms 4 and 5, which are stored at the end of the attachment list of T. Then the cross product $Ta \times Tb$ is obtained (where a and

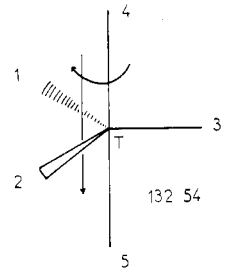


Figure 2. Ordering of the attachments list of a pentacoordinated atom (TBP).

b are the first equatorial atoms in the list, e.g., 1 and 2) and the dot product (T*a* × T*b*)·T4 is calculated. According to the sign of this quantity, the order of the two axial atoms (and bonds) is either inverted or not inverted in order to correctly represent the configuration.

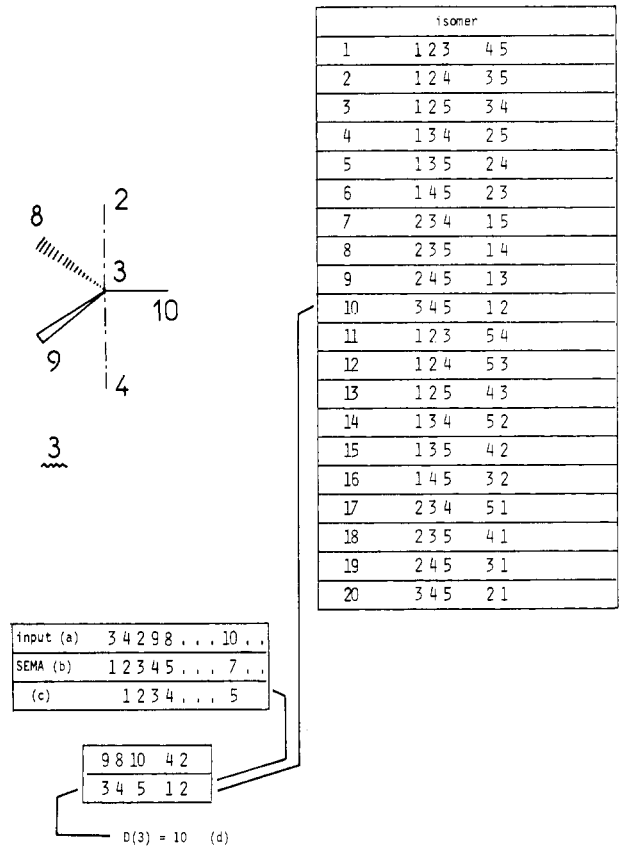
Thus, in Figure 2 the unordered attachment list for atom T(5,4,3,2,1) is ordered during perception[6] into ⟨1,3,2 5,4⟩. The three equatorial atoms are in a clockwise order when viewed from the second toward the first apical atom. A specific stereodescriptor is assigned to bonds T4 and T5, so that they are displayed in a special manner to indicate to the user that they have been perceived as apical.

At this point the CT is in agreement with the displayed structure. It is not necessary that the attachment list ordering be canonical and in fact there are many equivalent representations of the configuration in Figure 2: ⟨1,2,3 4,5⟩, ⟨2,3,1 4,5⟩, ⟨3,2,1 5,4⟩, ⟨3,1,2 4,5⟩, etc. A pairwise interchange of equatorial ligands requires interchange of apical ligands to maintain the configuration. A cyclic permutation of order three on the equatorial ligands does not require interchange of apical ligands.

STEREOCHEMICAL CANONICAL NAME

A canonical name for a molecule is desirable because it provides a compact description of the structure and allows efficient determination of equivalence or isomeric relationship between structures without graph matching. The SEMA naming algorithm,[3] a stereochemical extension of an algorithm first described by Morgan,[15] includes the configuration of carbon–carbon double bonds and saturated carbon stereocenters. We have extended it to description of a TBP center and, in fact, any kind of ligand-skeleton environment. Instead of having only two possible configurations of ligands as does sp^3 carbon, these complex skeletons have many possible configurations. The assignment of a configuration descriptor therefore involves an explicit comparison of the ordered attachment list of ligands to a reference list of all possible orderings of ligands about the given skeleton. This has already been briefly pointed out by Wipke et al.,[3] and used for another purpose in a computer program designed to search all the isomers in octahedral complexes.[16]

The first problem in creating a canonical name for a structure is to establish an ordering on the atoms of the structure which is invariant with respect to the original input numbers. SEMA[17] uses the following features in order of decreasing priority: extended connectivity (EC), spanning tree connectivity, ring closures, atom types, bond types, sp^2 stereodescriptors, sp^3 stereodescriptors, TBP stereodescriptors. If two different orderings of the atoms appear equivalent on the basis of EC values and connectivity, then atom types are used to break the equivalence. If they are still equivalent, the next lower priority feature is examined until one ordering can be selected as "preferred". Ties up to, but not including, TBP stereodescriptors indicate equivalencies among atoms; these are stored on an equivalence list.[16] The unique ordering placed on the atoms of the structure also serves as our precedence



isomer		
1	123	45
2	124	35
3	125	34
4	134	25
5	135	24
6	145	23
7	234	15
8	235	14
9	245	13
10	345	12
11	123	54
12	124	53
13	125	43
14	134	52
15	135	42
16	145	32
17	234	51
18	235	41
19	245	31
20	345	21

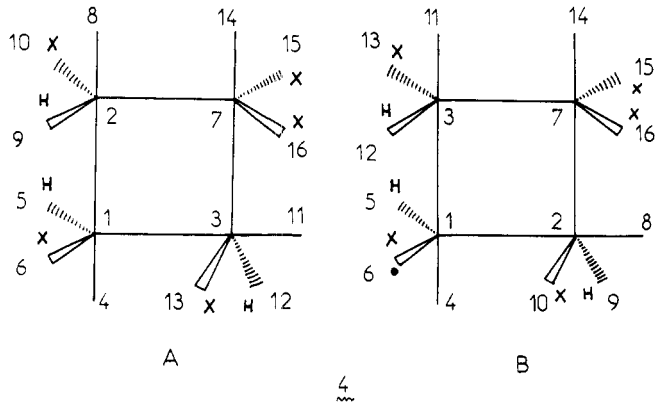
Figure 3. Assignment of a conformation descriptor to a TBP center: (a) input numbers, (b) sequential ordering of the atoms when the SEMA name is obtained, (c) SEMA relative ordering for the attachments to atom 3 (input number), (d) conformation descriptor for atom 3.

rule in describing stereochemical configurations—trigonal, tetrahedral, and TBP. We now describe how the TBP configuration description is obtained.

TBP CONFIGURATION DESCRIPTOR

For each TBP center T, the program starts from the ordered attachment list of T, and builds an identical ordered list containing the relative precedence of each ligand. Thus, in Figure 3, attachment list for atom 3 is ⟨8,9,10 2,4⟩. Replacing each atom number by its SEMA numbering gives ⟨5,4,7 3,2⟩, and if we replace each SEMA number by its relative rank in the list, we obtain ⟨4,3,5 2,1⟩, which describes the configuration at atom 3 in a notation system which is independent of the input ligand numbering. This list is reordered so that the first three indices (equatorial) are in ascending order, the configuration being maintained by a contingent permutation of the last two indices (axial): ⟨4,3,5 2,1⟩ ≡ ⟨3,4,5 1,2⟩. The list is then matched against a reference array of the 20 possible TBP configurations, each configuration being described by a list having the first three indices in ascending order (see Figure 3). The index of the matching configuration in the reference array is assigned as a configuration descriptor for the TBP under consideration. The reference array itself is ordered so that the difference between the descriptors of two enantiomeric configurations is equal to 10.

If all the ligands attached to a TBP are different, then there will be only one SEMA numbering for the ligands and that precedence leads to one unique configuration descriptor for the TBP. But if some of the ligands are identical, then there will be several different SEMA numberings, each leading to an identical SEMA name with the exception of the TBP configuration descriptor. In this case the SEMA name having the lowest TBP configuration descriptor will be selected as the



4

Assignment	SEMA numbering 1	2	3	7
A	9	9	9	6
A'	9	9	9	16
B	6	9	9	9
B'	6	9	9	19

Figure 4. Example of assigning a canonical name to a complex structure containing several TBP centers.

canonical name. This is analogous to our treatment of double bond and sp^3 configurations as described earlier.[3] It is necessary that TBP descriptor be used to select the "lowest" numbering. When there is more than one TBP, then as usual, there is a vector of configuration descriptors,[3] ordered by SEMA number of the phosphorus atoms involved. The TBP having the lowest SEMA numbered phosphorus atom is considered most significant and comparison of this vector with another vector is made lexicographically beginning with the most significant TBP. This algorithm can derive a canonical name for very complex cyclic structures having many TBPs and carbon stereocenters where the configurations are mutually interdependent.

The complex hypothetical structure shown in Figure 4 serves as an illustration of how simply the algorithm works with multiple TBP centers. For the purpose of naming a structure, it is irrelevant whether the structure is stable or can even exist, and we have been careful to *exclude concepts of strain and stability from the naming process*. Figure 4 shows the four trial numberings of structure **4** (A,A′,B,B′ where the primed assignments have SEMA numbers 15 and 16 reversed), all of which are equivalent with the exception of the TBP configuration vector. The algorithm for generating these trial numberings is described elsewhere.[3] The preferred SEMA numbering for **4** is B since it leads to the lowest TBP configuration vector, (6,9,9,9), representing the configurations for atoms with SEMA numbers 1, 2, 3, and 7, respectively. The configuration of atom 7 is least significant, only being considered when configurations for atoms 1, 2, and 3 are tied, e.g., between A and A′ or between B and B′. One can see in Figure 4 that the choice of configuration description for atom 7 is partially dependent on the choice made for atom 1. A proof of uniqueness and one-to-one correspondence between SEMA name and molecular structure was given earlier.[3] This extension to higher coordination numbers does not alter that proof since the proof was a graph theoretical one, not dependent on coordination number.

ENANTIOMERIC RELATIONSHIPS

Our earlier work described how from the SEMA name for a structure one can directly create the SEMA name for the mirror image structure.[3] If the two names are identical, then the structure is *configurationally achiral* (*excludes possible conformational chirality*); otherwise, the structure is chiral and the two names represent a pair of enantiomeric structures. This same powerful capability exists when SEMA is extended to higher coordination numbers.

For structure **4** (Figure 4), its mirror image leads to four mirror image numberings with mirror configuration descriptors:

$\bar{A}$ 19,19,19,16
$\bar{A}'$ 19,19,19,6
$\bar{B}$ 16,19,19,19
$\bar{B}'$ 16,19,19,9

where the mirror $\bar{C}$ of configuration C is given by expressions 1 or 2.

$$\bar{C} = (C + 9) \text{ modulo } 20 + 1 \quad (1)$$

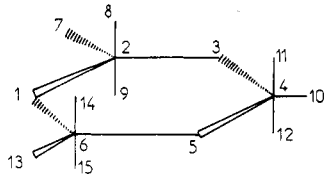
$$\bar{C} = \text{if } C \leqslant 10 \textit{ then } C + 10 \textit{ else } C - 10 \quad (2)$$

The preferred name then for the mirror image of **4** is $\bar{B}'$ having TBP descriptors (16,19,19,9). Thus as SEMA calculates trial names for the structure it also calculates trial names for the mirror image structure (by simply inverting all configurations). By comparing the preferred SEMA descriptors for each, (6,9,9,9) for **4**, and (16,19,19,9) for the mirror image of **4**, SEMA creates a set S_{RC} containing those atoms whose configurations are inverted in the preferred name for the mirror image. $S_{RC} = \{1,2,3\}$ and $7 \notin S_{RC}$. Information about which stereocenters belongs to S_{RC} is stored as part of the SEMA name. Therefore, to generate the SEMA name for a mirror image structure one simply copies the constitutional part and inverts every stereocenter belonging to S_{RC}. If $S_{RC} = \phi$, then there are no stereocenters to invert, meaning the mirror image structure is configurationally superimposable on the original; hence the structure is configurationally achiral. If $S_{RC} \neq \phi$ then the structure is chiral. Thus one can easily determine the relationship between any two structures by generating the SEMA name for each and comparing the names. Enantiomeric structures will have names which differ only in the configuration of stereocenters $\in S_{RC}$ and the difference must be a mirror relationship which for a TBP is defined by expression 1 or 2.

EFFECT OF SYMMETRY AND EQUIVALENT LIGANDS

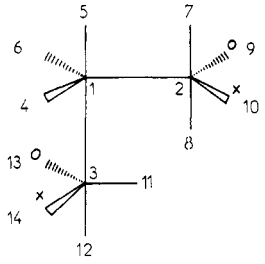
Since the algorithm generates all the trial numberings, it is quite general and it leads to the best list of descriptors. However, the presence of symmetry and equivalent ligands greatly increases the number of possible SEMA number assignments which must be explored; e.g., the presence of 2 X's on atom 7 of **4** (Figure 4) doubled the number of possible assignments. A simple shortcut which minimizes effort in such cases has been implemented and tested on structures containing up to six TBP centers. The basic principle involved is that one can separate the equivalent attachments to TBP centers into two sets SE1 and SE2. Equivalent here means that interchange among the attachments does not result in a different molecule. SEMA determines equivalencies as it generates equivalent names.[3] The set SE1 contains all attachments for which a different numbering modifies only the stereodescriptor of the TBP atom to which they are bonded: for the hypothetical structure I below, SE1 = {7,8,9,10,11,12,13,14,15}. Set SE2 contains the other attachments (1,3,5). Different numberings of atoms in SE2 may alter the stereodescriptors of several TBP's. The algorithm seeks the best numbering for the TBP centers and the atoms in SE2 first. In order to

accomplish this, for each SEMA number assignment it carries out all possible permutations of the equivalent ligands attached to each TBP center. During this operation, each TBP is considered independently of the others. Then, the program computes a set of equivalent stereodescriptors for each TBP and selects the lowest number of this set as the configuration descriptor for the atom under scrutiny.



I

For example, structure II receives the SEMA numbers shown, (4 ≡ 5 ≡ 6), (7 ≡ 8), and (11 ≡ 12), and SE1 = {4,5,6,7,8,9,10,11,12,13,14}. The numbering within sets of equivalent atoms is arbitrary if one does not consider configuration. The attachment list for atom 1 is ⟨2,4,6 3,5⟩ or ⟨1,3,5 2,4⟩ when converted to relative rank form. To find the lowest descriptor corresponding to this list, we replace numbers of equivalent atoms by the lowest number in the equivalence class, e.g. (4 ≡ 5 ≡ 6) is (3 ≡ 4 ≡ 5) in relative rank, converting ⟨1,3,5 2,4⟩ to ⟨1,3,3 2,3⟩. The reference array (Figure 3) is



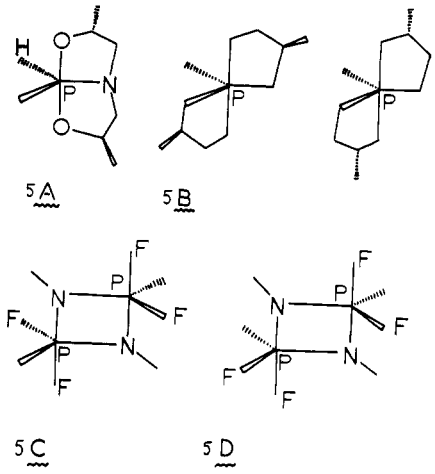
II

similarly modified: 123 33, 123 33, 123 33, 133 23, 133 23, ... By comparing ⟨1,3,3 2,3⟩ against the modified reference array, we find a set of configuration descriptors, each of which describe the configuration of atom 1. Treating atom 2 and 3 in a similar manner we obtain:

Atom	Configuration
1	{4,5,6,14,15,16}
2	{6,16}
3	{9,10}

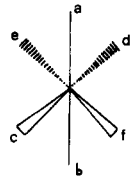
Picking the lowest descriptor in each set as preferred, the final SEMA descriptor list becomes (4,6,9) for atoms 1, 2, and 3, respectively. An atom is a member of S_{RC} if the complement $\bar{c}$ of its configuration c is not contained in its set of configurations. Thus, in this example S_{RC} = {3}; therefore, the enantiomer of II has TBP descriptors (4,6,19) and II is chiral. As a result of dividing up the naming problem into a global part (SE2) which handles interactions between configurations and a local part (SE1) in which interaction between configurations is impossible, the number of trial numberings is reduced in these examples by 2^5.

In order to check the correctness of the program, configuration descriptors obtained for more than 50 molecules have been successfully compared to those calculated manually. Figure 5 shows some simple applications of the naming algorithm. The two meso and one *dl* pair of the bicyclophosphoranes **5a**[18] were properly differentiated. The phosphorus atom of spirophosphoranes **5b** received descriptions 13 and 20, respectively. The trans and gauche isomers **5c** and **5d** of 2,4-dialkyl-2,2,4,4-tetrafluoro-1,3,2,4-diaza-



5A 5B 5C 5D

Figure 5. Hydrogen atoms are not explicitly represented (except in structure 5A for H atom bonded to the phosphorus). Unlabeled atoms are carbons.



6

Figure 6.

diphosphetidine[19] are easily differentiated by the program.

The SEMA algorithm assigns configuration descriptors to whatever structure it is given; it treats a TBP center as if it were rigid. Since many pentacoordinated phosphoranes have a fluxional character,[20] provision is made elsewhere in the program for pseudorotation and stability criteria.[21]

The extension of the algorithm to any kind of 3D skeleton appears straightforward. Thus, for an octahedral structure, one simply needs to define an ordered list of six atoms such as (a b c d e f) (see Figure 6), provide the program with the list of 30 reference isomers, and some symmetry information describing how to manipulate the ordered list. One also needs to define unambiguous conventions for graphical representation of the 3D skeleton for input to and output from the computer. Such notation should be stereochemically pure and should not depend on unrelated concepts such as ring strain to resolve ambiguities.

STEREOCHEMICAL MODIFICATIONS

After perceiving and naming the target molecule, the program contains a full and accurate representation of the molecular structure. The next step involves applying chemical transformations to this structure in order to obtain synthetic precursors. For this purpose, the program has access to a library of chemical reactions which are written backwards and called transforms. They are written by the chemist in an English-like language called ALCHEM,[2,22] which allows a correct description of all the features and conditions relevant to any reaction. In a very general way, one can represent a reaction at any atom by the changes incurred by three parameters: coordination number, oxidation number, and spatial arrangement of the attachments to this atom. These parameters are in fact interrelated, but their knowledge provides a full description of the molecular topology at any moment.

For brevity we will denote a chemical reaction taking place at a phosphorus atom by P(ij/kl), where i and k are the initial and final oxidation numbers (valences) and j and l are the

corresponding coordination numbers. A transform will be denoted P(kl ==> ij).

These atomic parameters have been included in the CT and are manipulated by ALCHEM statements. For example, a modification of the oxidation number is keyed by the following instruction:

MODIFY VALENCE FOR ATOM m TO n

which changes the oxidation number of ATOM m into the new value n. This is especially useful when implicit hydrogens do not allow a direct calculation of the valence. In standard cases, the valence modification is automatically made according to the variation of the coordination number. This holds also for a stereochemical change. Generally, a specific coordination number corresponds to one type of spatial environment. For pentacoordinated phosphorus, the most common structure is a TBP. A statement of the following type can be used to indicate that a SBP geometry is obtained instead:

DEFINE STRUCTURE SBP FOR ATOM n

In this case, the program considers the five attachments list of ATOM n as representing a square-basis pyramidal configuration. However, this statement does not provide any information on the spatial positions of the attachments to ATOM n. This can be done by using MOVE instructions in the transform description (vide infra). If no statement DEFINE STRUCTURE is used, then the program assumes a TBP configuration for a phosphorus atom.

The coordination number is modified in the usual way[6] by manipulation instructions which break bonds or add atoms. For reaction types P($i3/j3$) or P($i4/j4$) involving tetrahedral centers, the stereochemical changes are handled in the same way as for carbon atoms. For reactions whose stereochemistry is dictated by a pentacoordinated intermediate, such as those involving phosphine oxide reduction or a nucleophilic substitution at a phosphonium ion,[23] empirical rules derived from the stability of the intermediate appear sufficient to describe many cases. If the reaction is stereospecific, queries are made in the transform about the presence and size of rings, which dictate the configuration of the precursor. If the reaction leads to a mixture, then an instruction

LOSE STEREOCHEMISTRY AT ATOM n

is used. To handle these cases more correctly, it is necessary to have ALCHEM queries on electronegativities of groups and atoms. This can be done by using statements such as:

IF ATOM n IS ELECTRONEGATIVE THEN. . .

Groups electronegativities can also be tested. It is necessary to store the group atoms in a set register[2,22] (n):

IF (n) IS ELECTROPOSITIVE THEN. . .

Group electronegativities are simply computed by averaging the individual electronegativity values of the group atoms.[24] These values have been checked with those obtained by a more complex treatment,[25] and the agreement is quite satisfactory: for 72 groups out of 87, the discrepancy is within 5%. ELECTRONEGATIVE means that the computed electronegativity is greater than that of carbon; ELECTROPOSITIVE means the reverse.

In a reaction of type P($i4/j3$), an atom is added to the target in the retrosynthetic direction. In the computer, this atom is added symbolically at the end of the ordered attachments list representing the configuration of the target phosphorus atom. The atom added replaces the lone pair with retention of stereochemistry at P. An inversion can be carried out by the instruction: INVERT ATOM n. A phosphine oxidation (Type P33/54) is similarly handled in the retrosynthetic direction. The leaving atom is replaced by a dummy which is moved to the end of the attachment list by an even number of permutations,[6] again with retention of stereochemistry.

The reactions relevant to the formation of pentacoordinated phosphoranes are more complex. We considered in a very general way the stereochemical changes involved in reactions dealing with these compounds. Although today the synthetic chemistry of phosphoranes does not really require such sophisticated treatment, this makes the program more versatile in meeting future demands. A problem arises in the definition of the ALCHEM statements which allow an accurate description of the passage from any TBP conformation to another, or from a TBP geometry to tetrahedral one, and vice versa. For this purpose, the five atoms bonded to a TBP center define five fixed positions in space with respect to this center (see Figure 2). The numbering shown corresponds also to the entries in the attachment list. Let us consider, for example, a transform which requires a substructure containing at least two atoms (ATOM 2 and ATOM 3) bonded to a pentacoordinated phosphorus (ATOM 1). Such a pattern will be provided under the linear form P(-A)-A, where A means any atom. The program will try any set of three atoms fulfilling this requirement, but the user does not know whether or not the selected atoms 2 and 3 are in axial or equatorial positions, and therefore cannot unambiguously describe any stereochemical change.

By the following queries in the transform:

IF ATOM 2 IS NOT ATOM IN POSITION 1 / ATOM 1 THEN KILL (3)

IF ATOM 3 IS NOT ATOM IN POSITION 4 / ATOM 1 THEN KILL (4)

where / ATOM 1 means with respect to ATOM 1, the chemist makes sure that the TBP is in a well-defined position. When the first query (3) is encountered, a positive answer can be corrected by checking if there is an available rotation to put ATOM 2 in position 1. If the query (4) is not immediately fulfilled, the program examines the possible rotation leading to a correct situation, and leaving ATOM 2 in its required position. If no such operation is feasible, the program tries another match for the required substructure. When found, and when queries (3) and (4) are successfully passed, manipulation instructions can be performed. A modification of the configuration at a TBP center is keyed by the following instructions:

MOVE ATOM 2 INTO POSITION 5 / ATOM 1 (5)

MOVE ATOM 3 INTO POSITION 2 / ATOM 1 (6)

For example, the first statement (5) puts ATOM 2 in the fifth entry of the attachment list of ATOM 1, and the atom previously occupying this entry is put in place of ATOM 2. Thus, if the target input numbers of ATOM 2 = 1, and ATOM 3 = 9, the queries and statements (3)–(6) will give:

$$\langle 3,1,5\ 9,2\rangle \xrightarrow{(3)} \langle 1,3,5\ 2,9\rangle \xrightarrow{(4)} \langle 1,5,3\ 9,2\rangle \xrightarrow{(5)}$$

$$\langle 2,5,3\ 9,1\rangle \xrightarrow{(6)} \langle 2,9,3\ 5,1\rangle$$

Note that parity was preserved during (3) and (4), but not during (5) and (6).

In a transform P(54 ==> 55), such as that corresponding to the reaction **7b** ==> **7a**, atoms are added to the target at the end of the attachment list, which corresponds in the synthetic sense to an axial departure from a TBP (see Figure 7). This situation is generally considered as the most favorable,[23] although equatorial departure can also be invoked when good leaving groups are present.[26] In such a case, the correct representation can be obtained by inserting a MOVE

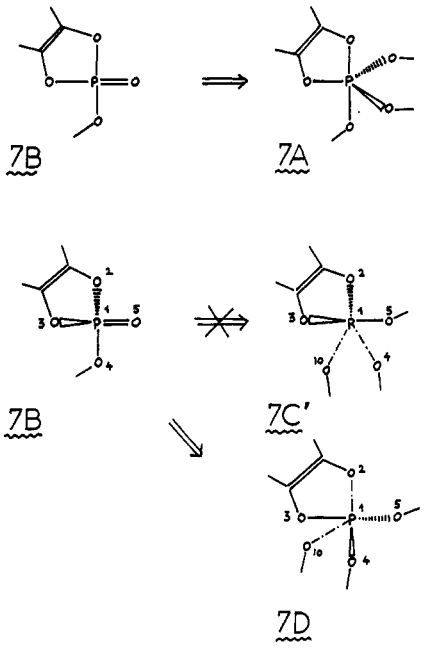


Figure 7. Computer display of the stereochemical modifications involved in the hydrolysis of a pentaoxyphosphorane.

instruction in the transform description.

This description is not necessary in many cases because we have built into the program standard default manipulations for common stereochemical changes occurring in reactions involving pentacoordinated phosphorus species. In a transform of type P(54 ==> 55), such as the hydrolysis of a pentaoxyphosphorane **7a**,[27] an atom is added to the target (atom number = 10). Normally this should lead to precursor **7c'**, since the atom is added at the end of the attachment list and is thus in axial position:

$$P\langle 3,2,5,4\rangle \Rightarrow \langle 3,2,5\ 4,10\rangle \qquad (7)$$

Since isomerization occurs easily at the phosphorus atom, the only condition imposed by the program is that structure **7a**, when displayed, meets the requirements of a TBP geometry.[21] If the second list in (5) is directly interpreted and displayed, it would lead to diagram **7c'**, in which a five-membered ring spans two equatorial positions. To avoid such a situation, the list P(3,2,5,4) is automatically rearranged so that the last entry is occupied by either a ring atom, or the most electronegative atom of the list, leading to the following operations:

$$P\langle 3,2,5,4\rangle \equiv P\langle 3,5,4,2\rangle \Rightarrow P\langle 3,5,4\ 2,10\rangle$$

and thus to the display of diagram **7d**.

The displays **7d** and **8a** are difficult to understand because the projected diagrams no longer conform to the conventions for the TBP configuration (see Figure 8). Automatically rearranging the drawing is for the general polycyclic case a difficult layout problem. Instead SECS allows the user to manually reposition the atoms using a tablet or light pen, but the program does assure that the diagram stays in agreement with the CT representation of the configuration. If the chemist repositions the atoms as in **8b**, the diagram is in agreement with the CT. However, if the chemist repositions the atom as in **8c**, the diagram is incorrect, and the program immediately modifies the hashing and wedging to **8d** which is correct and equivalent to **8b**. Structure **8d** just represents a view of the TBP from below the plane of equatorial substituents.

Some of the important features of ALCHEM for phosphorus chemistry are: (i) stereospecific and stereoselective reactions can be represented, and specific stereochemical relations can be examined in the target or created in the

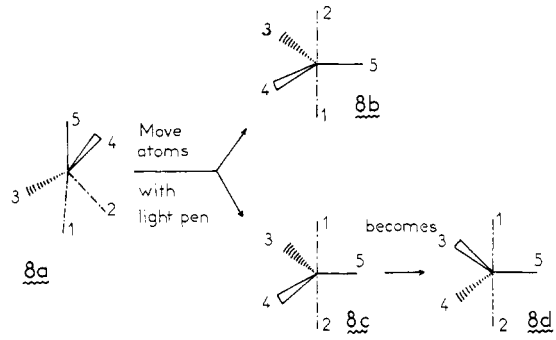


Figure 8. Manual modification of a display TBP structure. In **8a**, atoms 1 and 2 are moved manually with the light pen. The first way gives **8b**, which is correct. However, since the user does not know the exact representation of **8a**, he can move 1 and 2 so that **8c** is obtained instead. In this case, the program corrects automatically the displayed structure and gives **8d**.

precursor, (ii) the default for attack on a tetracoordinated center is face attack with groups entering or leaving the TBP from apical positions, (iii) electronegativity of groups can be examined, (iv) changes in coordination and configuration can be controlled, and (v) ALCHEM automatically pseudorotates precursors which have unstable configurations into stable configurations.

EVALUATION OF INTRAMOLECULAR EXCHANGES

The fact that a TBP can undergo intramolecular exchanges has already been taken into account in the previous section by ALCHEM to generate only stable TBP structures.[21] After the precursor is generated the EVAL module of SECS[1g] determines if the precursor is a duplicate of any other structure already existing in the synthesis tree. The exact action taken when a duplicate is encountered depends on many factors discussed elsewhere.[1a] The interesting problem to be discussed here is that it is not sufficient to determine if the precursor is identical with or enantiomeric to another structure in the tree, because for TBP structures the precursor may be easily interconvertible by pseudorotation to another structure in the tree.

The solution to this problem can still be obtained by comparison of SEMA names of the structures, which is an efficient way of comparing molecular topologies, but additional logic is required for TBP pseudorotations. A flowchart of the comparison logic is shown in Figure 9.

If the SEMA names for molecules A and B are identical up to the TBP descriptor, then the comparison proceeds to determine if the respective TBP configurations in A and B, Ca and Cb, are either identical or interconvertible. To determine which configurations can be interconverted for this molecular skeleton, the program begins with a table X containing, for each of the 20 reference isomers, a set σ of three descriptors corresponding to the configurations which can be reached by a pseudorotation process (BPR).[28] An example for spirophosphoranes is given in Figure 10. The BPR operation (Figure 11) is internally represented as:

$$\langle a,b,c\ d,e\rangle \xrightarrow{\text{pivot } a} \langle a,d,e\ c,b\rangle$$

The initial σ's, being structure independent, are simply constant. The stability of each configuration is then evaluated on the grounds of ring strain and electronegativity, and the unstable configurations are removed from the σ sets. This prevents pseudorotation through high-energy intermediates.

Now the sets[29] of stable interconvertible configurations {SINC} are obtained in the following way. Let σ_i be any set belonging to the table X, (i is stable) $\sigma_i = (\delta_{i1}, \delta_{i2}, \delta_{i3})$. One replaces each element δ_{ij} by its corresponding σ and the

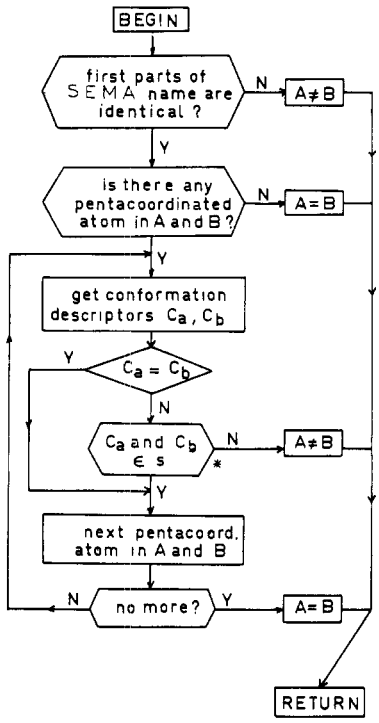


Figure 9. Flowchart of the topological comparison between two molecules A and B, which can include pentacoordinated atoms. *See text and Figure 10.

RANK	(a)	ATTACHMENTS (b)	GIVES (c)
1	U	1 2 3 4 5	16 20
2	U	1 2 4 3 5	8 10
3	S	1 2 5 3 4	20
4	U	1 3 4 2 5	13 18
5	U	1 3 5 2 4	
6	S	1 4 5 2 3	18
7	U	2 3 4 1 5	3 6
8	S	2 3 5 1 4	16
9	U	2 4 5 1 3	
10	S	3 4 5 1 2	13
11	U	1 2 3 5 4	6 10
12	U	1 2 4 5 3	18 20
13	S	1 2 5 4 3	10
14	U	1 3 4 5 2	3 8
15	U	1 3 5 4 2	
16	S	1 4 5 3 2	8
17	U	2 3 4 5 1	13 16
18	S	2 3 5 4 1	6
19	U	2 4 5 3 1	
20	S	3 4 5 2 1	3

{SINC}(e)	{SIDENT}(f)		{S}(g)
(3,20)	(1,12)	(7,14)	(3,20)
(6,18)	(2,11)	(10)	(6,8,16,18)
(8,16)	(3)	(13)	(10,13)
(10,13)	(4,17)	(15,19)	
	(5,9)	(16,18)	
	(6,8)	(20)	

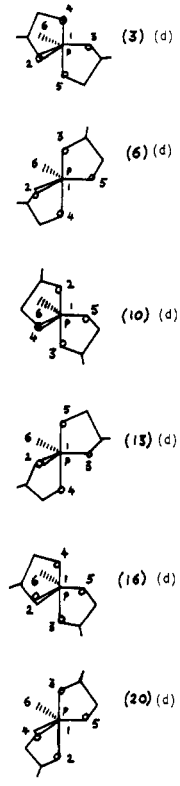


Figure 10. Determination of the sets of stable interconvertible configurations for a spirophosphorane: (a) indicates the stability of the configuration, S for stable, U for instable; (b) attachment list for atom 1 (relative ordering of the ligands); (c) stable configurations (3 at most) reached by exchange involving on pseudorotation (BPR);[28] (d) configuration descriptor of atom 1; (e) sets of stable interconvertible configurations, without taking account of the equivalent attachments; (f) sets of stable identical configurations; (g) final sets of stable interconvertible configurations.

operation is repeated until the set σ_i becomes invariant. Now,

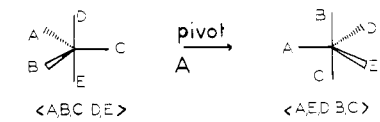


Figure 11. Internal representation of BPR process.

it is also necessary to take account of the sets of equivalent configuration descriptors, which are stored in {SIDENT}: they were built during application of the SEMA algorithm and were saved for this moment. The final sets of interconvertible configurations {S} can be deduced from {SINC} and {SIDENT} as follows. Let Sident_i and Sinc_j be any member of {SIDENT} and {SINC}: if $\text{Sr} = \text{Sident}_i \cap \text{Sinc}_j \neq \phi$, let us replace Sinc_j by $\text{Sinc}_j \cup \text{Sident}_i$. Thus, in Figure 10, for $\text{Sinc}_2 = (6,18)$, and $\text{Sident}_6 = (6,8)$, $\text{Sr} = (6)$, and we replace Sinc_2 by (6,8,18). If Sr is empty, then Sic_j is kept unchanged. If $\text{Sinc}_j \subset \text{Sinc}_k$ ($k < j$), then Sinc_j is deleted; i.e., redundant subsets are removed. The scanning over i and j is resumed and leads to the final sets of interconvertible configurations {S}. It is thus seen that configurations 3 and 20 can be exchanged, but this is not possible for configurations 3 and 6. This technique does not take into account the number of steps necessary to perform this exchange,[30] but does assure that there is no high-energy intermediate on the path. Note that there is a different {S} for each TBP in the molecule A. These sets for precursor A are created only once and are used in comparing A to all other structures in the synthesis tree. After structure A is established as being unique, the sets {S} can be discarded since for later comparison of a new precursor with A, there will be {S}'s for the new precursor, and it is only necessary to have {S}'s for one of two structures in a comparison.

CONCLUSION

We have developed a series of algorithms and conventions which allow a comprehensive treatment of the stereochemistry of phosphorus compounds. Conformations of pentacoordinated atoms are easily recognized, and the extension of the ALCHEM language permits handling all the stereochemical queries and manipulations which arise in reactions of these compounds. This represents a new direction of the SECS program and extends the scope of the field of computer design of synthesis. Work is now in progress for the writing of files of specific transforms, and for the development of strategies relevant to this area of synthetic chemistry. Moreover, the SEMA algorithm extension described in this paper can be adapted rather easily to any kind of spatial environment, and can thus allow the treatment of hexacoordinated species as well.

ACKNOWLEDGMENT

The authors thank M. Sabourin for the development of the graphic VERONICA software and Dr. F. Mathey for helpful discussions. They thank also the CNRS for financial support (R. M. and F. C.) and for a grant of computer time, and they (T. W.) thank the National Institutes of Health, Division of Research Resources, Biotechnicology Branch for support of SECS through the SUMEX Resource at Stanford and through Grant RR1059. Information requests about SECS are to be sent directly to W. T. Wipke.

REFERENCES AND NOTES

(1) (a) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969); (b) E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, and G. Petersson, *J. Am. Chem. Soc.*, **97**, 6116 (1975); (c) E. J. Corey, W. J. Howe, and D. A. Pensak, *ibid.*, **96**, 7724 (1974); (d) E. J. Corey and W. L. Jorgensen, *ibid.*, **98**, 189 (1976); (e) *ibid.*, **98**, 203 (1976); (f) E. J. Corey, R. D. Cramer III, and D. A. Pensak, *ibid.*, **98**, 210 (1976). (g) W. T. Wipke in "Computer Representation and Manipulation of Chemical Information", Wiley-Interscience, New York, N.Y., 1974, p 147; (h) E. J. Corey, W. T. Wipke, R. D. Cramer, and W. J. Howe, *J. Am. Chem.*

Soc., 94, 421, 431 (1972); (i) P. E. Blower Jr., and H. W. Whitlock Jr., ibid., 98, 1499 (1976).
(2) T. M. Dyott, program "Simulation and Evaluation of Chemical Synthesis", Ph.D. Thesis, Princeton University, 1973.
(3) W. T. Wipke and T. M. Dyott, J. Am. Chem. Soc., 96, 4834 (1974).
(4) Software VERONICA, Regional Computer Center of the University of Copenhagen (Denmark), RECKU, Publication No. 10, 1972.
(5) F. Sabourin and G. Kaufmann, UNIVAC 1110 version of the VERONICA software, Laboratoire de Spectrochimie Moléculaire, Internal Report, 1975.
(6) W. T. Wipke and T. M. Dyott, J. Am. Chem. Soc., 96, 4825 (1974).
(7) R. Sarma, F. Ramirez, B. McKeever, J. F. Marecek, and S. Lee, J. Am. Chem. Soc., 98, 581 (1976).
(8) J. I. G. Cadogan, R. O. Gould, and N. J. Tweeddle, Chem. Commun., 773 (1975).
(9) R. Luckenbach, "Dynamic Stereochemistry of Pentacoordinated Phosphorus and Related Elements", Georg Thieme, Stuttgart, 1973.
(10) The X°TR configuration is not accepted as an input by the program, as it appears very infrequently, and as it requires more complex data input and output conventions to be adequately displayed.
(11) G. M. Kosolapoff and L. Maier, "Organic Phosphorus Compounds", Vol. 3, Wiley-Interscience, New York, N.Y., 1972, p 303.
(12) R. Luckenbach, ref 9 pp 214–220, and references therein.
(13) P. C. Lauterbur and F. Ramirez, J. Am. Chem. Soc., 90, 6722 (1968); J. D. Dunitz and V. Prelog, Angew. Chem., 80, 700 (1968).
(14) M. G. Newton, J. E. Collier, and R. Wolf, J. Am. Chem. Soc., 96, 6888 (1974).
(15) H. L. Morgan, J. Chem. Doc., 5, 107 (1965).
(16) W. E. Bennett, Inorg. Chem., 8, 1325 (1969).
(17) For a detailed description of SEMA, see ref 3.
(18) D. Houalla, T. Mouheich, M. Sanchez, and R. Wolf, Phosphorus, 5, 229 (1975).
(19) R. K. Harris, M. I. M. Wazeer, O. Schlak, and R. Schmutzler, J. Chem. Soc., Dalton Trans., 17 (1976).
(20) K. I. The and R. G. Cavell, Chem. Commun., 716 (1975).
(21) Throughout this work, the following standard rules are used to describe a stable TBP configuration: (i) Rings containing a TBP center span an apical and an equatorial position. In such a ring, the apical position is occupied by the more electronegative atom. (ii) If there is no such ring, apical positions are occupied by the more electronegative atoms bonded to the TBP.
These rules appear rather general. See, for example, R. Luckenbach, ref 9, p 10. Ring strain is generally considered preponderant over electronegativity, and few counterexamples are given in the literature. In such an example, a four-membered ring is forced to span two equatorial positions in a difluorophosphorane, owing to the presence of fluorines: N. J. De'ath, D. Z. Denney, D. B. Denney, and C. D. Hall, Phosphorus, 3, 205 (1974); see also: H. A. E. Aly, J. H. Barlow, D. R. Russell, D. J. H. Smith, M. Swindles, and S. Trippett, J. Chem. Soc., Chem. Commun., 449 (1976). However, although these rules are sufficient for our purpose, apicophilicity does not depend entirely on these factors, and the situation is more complex. See, for example, R. G. Cavell, D. D. Poulin, K. I. The, and A. J. Tomlinson, Chem. Commun., 19 (1974); R. K. Oram and S. Trippett, Chem. Commun., 554 (1972); S. Bone, S. Trippett, and P. J. Whittle, J. Chem. Soc., Perkin Trans. 1, 2125 (1974).
(22) T. M. Dyott, P. Friedland, and C. Still, W. T. Wipke, submitted for publication.
(23) K. Mislow, Acc. Chem. Res., 3, 321 (1970); F. H. Westheimer, Acc. Chem. Res., 1, 70 (1968).
(24) A F. Clifford, J. Phys. Chem., 63, 1227 (1959).
(25) J. E. Huheey, J. Phys. Chem., 69, 3284 (1965).
(26) W. Hawes and S. Trippett, Chem. Commun., 295 (1968).
(27) F. Ramirez, O. P. Madan, and C. P. Smith, J. Am. Chem. Soc., 87, 670 (1965); F. Ramirez, Synthesis, 2, 90 (1974).
(28) R. S. Berry, J. Chem. Phys., 32, 933 (1960). The fact that we consider only BPR in this work does not mean that this is the only process able to describe this exchange. If a new process appears to be more likely and general than this one, very few changes have to be made in the program. However, at the present time, BPR appears to be the most favorable process, at least in structurally flexible phosphoranes. See A. Strich and A. Veillard, J. Am. Chem. Soc., 95, 5574 (1973); J. A. Altmann, K. Yates, and J. G. Csizmadia, ibid., 98, 1450 (1976).
(29) A set is internally represented in the computer by 72 (or 64) bits. For details, see ref 1g.
(30) M. J. Clark and S. F. A. Kettle, Inorg. Chim. Acta, 14, 201 (1975).

NEWS AND NOTES

MEETINGS

1978 Gordon Research Conference: Scientific Information Problems in Research

Plymouth, State College, Plymouth, New Hampshire
Chemical and Biochemical Information Problems

10 July. Emerging information activities in the physical, chemical, and biological sciences. The effect on decision makers of new and projected technologies.

11 July. Future trends for information capture and analysis systems for the physical, chemical, and biological sciences.

12 July. Netting and interfacing biological, chemical, and physical data and information systems.

13 July. Artificial intelligence—framework for future chemical and biological information systems.

14 July. Voluntary and mandatory data depositories—impact on science and science/administrators.

SLA Annual Conference

Special Libraries Association will hold its 69th Annual Conference at the Radisson Muehlebach Hotel and the Kansas City Convention Center in Kansas City, Missouri, June 10–15, 1978. The theme of the Conference is "Managing for Change".

Six Continuing Education courses are scheduled on Sat. June 10. Substantive and timely programs are planned by the Association's 29 Divisions and by Association Committees. Programs planned by external organizations include the National Federation of Abstracting and Indexing Services and the American Society for Information Science. CLENE (Continuing Library Education Network and Exchange) has scheduled its two-day meeting on June 10–11.

National Computer Conference

The 1978 National Computer Conference, to be held June 5–8 in Anaheim, will feature the largest exhibit of computer hardware, software, systems, and services ever held. According to an announcement by the American Federation of Information Processing Societies, Inc. (AFIPS), sponsor of the annual NCC, more than 330 organizations reserved 1,382 booths.

More than 4,000 industry representatives will be available during the four-day exhibit program to demonstrate their latest products and services, provide technical and commercial data, and help attendees find solutions to their information processing needs. The exhibit program will include competitive offerings in a wide range of areas including components, data communications equipment, education and training materials, minicomputers, microcomputers, microprocessor systems, mainframes, memory systems, software systems, test equipment, terminals, and other computer peripherals.

In addition to the conference exhibit program in the Anaheim Convention Center, NCC '78 will also feature a separate exhibit of consumer computing products and services as part of the Personal Computing Festival to be held June 6–8 at the nearby Disneyland Hotel Convention Center.

Winter Simulation Conference

The 1978 Winter Simulation Conference (WSC 78) will be held December 4–6, 1978, in Miami Beach, Florida.

Cosponsoring WSC 78 with NBS are the American Institute of Industrial Engineers; Systems, Man, and Cybernetics