Pure and Appl. Chem., Vol. 49, pp. 1291-1306. Pergamon Press, 1977. Printed in Great Britain.

MOLECULAR FORMULA, CONFIGURATION AND CONFORMATION BY X-RAY ANALYSIS

Isabella L. Karle

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375, U.S.A.

Abstract - In the realm of natural products, the more important objectives of crystallography are the elucidation of the structural formula, particularly when many asymmetric centers are present or when unusual linkages occur, and the determination of the preferred conformation of flexible molecules. The former objective is illustrated by the unusual formulas established by single crystal x-ray diffraction analysis for the potent toxins secreted in the skins of tropical American frogs. Some of these toxins are steroidal alkaloids, while others contain novel allenic and acetylenic substituents on a spiro ring system or on cis-decahydroquinoline. For the latter objective, studies on cyclic peptides have established the existence of stable multiple conformations, a variety of intramolecular hydrogen bonds, and intricate conformational changes upon complexation.

The analysis of crystal structures by x-ray crystallography has become indispensable in a large number of diverse scientific disciplines ranging from metallurgy and mineralogy to molecular biology. The objectives extend from the very practical, i.e. establishing which atom is bonded to which in a particular substance, to the very fundamental, e.g., the distribution of electron density in covalent bonds. For the chemist dealing with natural products and other organic substances, one of the more important uses of crystallographic investigations is the elucidation of the structural formula of a natural product, a reaction intermediate or a photorearrangement product, especially when the substance has unusual or previously unknown chemical linkages or a large number of asymmetric centers. In such instances, other means of analyses often cannot establish the structural formula or distinguish among several equally plausible possibilities.

Another objective of crystal structure analysis is the determination of the preferred conformation of flexible molecules such as oligopeptides. Conformation is often the key to reaction mechanisms and reaction rates. As a bonus, from a single crystal structure analysis, in addition to the structural formula, stereoconfiguration and conformation, fairly exact values are obtained for the bond lengths, bond angles and torsion angles with standard deviations of 0.01A, 0.5° and 1.0°, respectively. Such values serve as norms for postulating models and conformations for amorphous substances, substances in solution and otherwise non-crystalline materials.

Some 14,000 crystal structures of organic and metal organic molecules have been described in the literature. Most of these structures have been reported in the last few years (Ref.1). The great proliferation is due mainly to four factors: the introduction of high speed computers, the development of automatic diffractometers for rapid data collection, the direct method for phase determination particularly for substances containing only light atoms, or many heavy atoms, and the appreciation that vast amounts of chemical and physical information become available with the expenditure of relatively little effort.

A primary requirement for a crystal structure analysis is a single crystal of sufficiently good quality to diffract a beam of x-rays to reasonably high scattering angles so that a large number of reflections can be measured. Each reflection is characterized by four values, h, k, l and I, where h, k and l are the indices of the plane in the crystal from which the diffraction occurs and I is the intensity associated with the reflection. A typical diffraction pattern recorded photographically is illustrated in Fig.1. Currently most x-ray data are collected on automatic diffractometers with counters rather than film. These data are used in the electron density function

$$\rho(XYZ) = \frac{1}{V}\sum\sum\sum F_{hkl}\, e^{-2\pi i(hX + kY + lZ)}, \qquad (1)$$

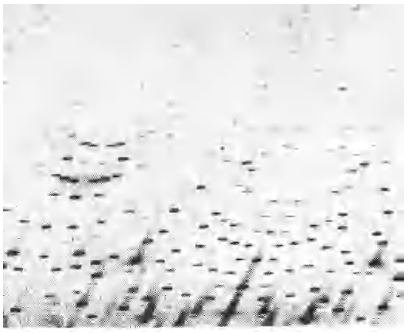


Fig. 1. An x-ray diffraction pattern of a crystal of an organic substance.

a function that is essentially zero everywhere except at particular values of X, Y and Z where atoms occur in the unit cell. The maxima of this three-dimensional function represent graphically the crystal structure where the coordinates X, Y and Z are measured as fractions of the crystallographic axes, the quantity V is the volume of the unit cell, and h, k, l are the indices of the planes from which reflections were recorded. The coefficients F_{hkl} are, in general, complex quantities, $|F_{hkl}| \exp(-i\phi_{hkl})$, where $|F_{hkl}|$ is proportional to the square root of the corresponding measured intensity. If the values of the phases ϕ_{hkl} were also known, then the $\rho(XYZ)$ function could be computed directly to immediately reveal the positions of all the atoms.

In general, three different procedures are used to obtain phase information: the heavy atom method, isomorphous replacement, and direct phase determination. The position of an atom considerably heavier than the remaining atoms, can be located in a vector map (Ref.2) and the phases calculated from the heavy atom alone are usually sufficiently correct to lead to the complete structure (Ref.3). In isomorphous replacement, used primarily in the crystal structure analyses of proteins, very heavy atoms or complex ions, are introduced into the cell without disturbing the cell contents. Several derivatives are used in which the heavy constituent occupies different sites in the cell. The positions of the heavy atoms are found in vector maps and lead to phase information (Ref.4).

In the third method, phase information is obtained directly from the measured intensities. It is based on the concept that the number of data collected is very much larger than the number of unknowns, namely, the three coordinates defining the positions of each atom. In 1950, J. Karle and H. Hauptman (Ref.5) constructed a complete set of inequalities showing the relationships between the structure factors F_{hkl} in any crystal. The formulas for direct phase determination (Ref.6) stem from these inequalities. In addition to the formulas, the inequalities provide the probability functions that measure the correctness of the phase determination. One of the simplest relationships is

$$\phi_{\vec{h}} \approx \phi_{\vec{k}} + \phi_{\vec{h}-\vec{k}} \qquad (2)$$

(where $\vec{h} = h_1k_1l_1$ and $\vec{k} = h_2k_2l_2$). This relationship is valid for those reflections with strong intensities. In applying this relationship, the indices are added and the values of the phases are added. An important step was the discovery that very few key phases are needed in order for the process to cascade (Ref.7). Initially only a few phases are known: those assigned (within prescribed rules) to select an origin and enantiomorph in the crystal and several others whose phase is designated by an unknown symbol. The values of the phases for other strong reflections are obtained from the initial phases in the cascading process. The values for some or all of the unknown symbols are indicated by relationships that occur in this process, Fig. 2. For those symbols that are not evaluated, there need be only a small number of alternative values to be considered (Ref.8)

The approximate values for the phases as obtained from Eq.(2) are refined by recycling in the tangent formula (Ref.9):

$$\tan \phi_{\vec{h}} = \frac{\sum |E_{\vec{k}} E_{\vec{h}-\vec{k}}| \sin(\phi_{\vec{k}} + \phi_{\vec{h}-\vec{k}})}{\sum |E_{\vec{k}} E_{\vec{h}-\vec{k}}| \cos(\phi_{\vec{k}} + \phi_{\vec{h}-\vec{k}})} \qquad (3)$$

where the $E_{\vec{h}}$ are normalized values for the $F_{\vec{h}}$.

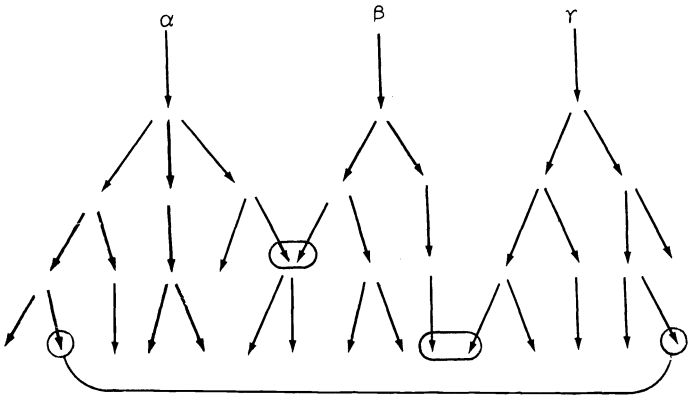
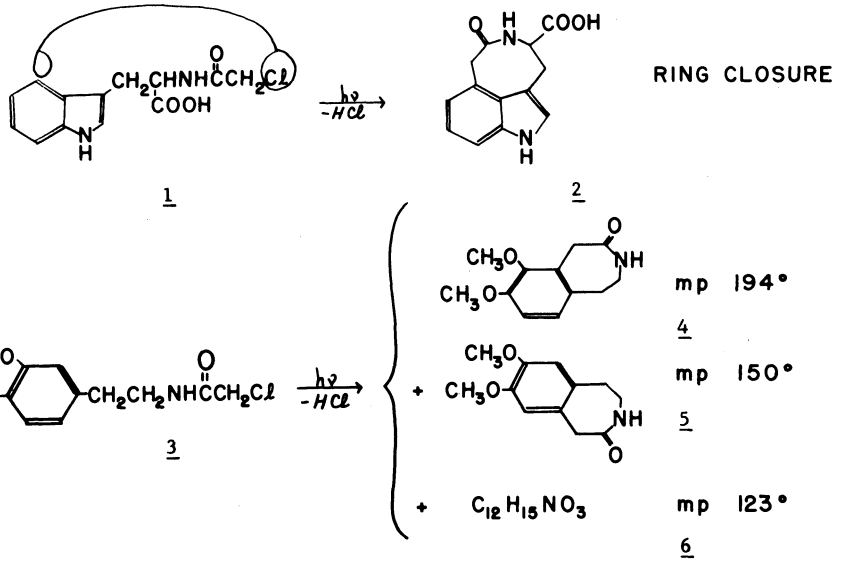


Fig. 2. Schematic representation of the cascading process in phase determination. The pairs of arrowheads enclosed within a circle represent a reflection whose phase value has been determined from two different paths. Such occurrences lead to algebraic relations among the symbols.

Let us now consider a variety of problems to which x-ray diffraction analysis has been applied. The examples are taken mostly from the research program at the Naval Research Laboratory.

PHOTOREARRANGEMENTS

Ring closure had been effected in N-chloroacetyl tryptophane 1 by eliminating hydrogen chloride as a result of exposure to ultraviolet light (Ref.10). This preparative method had been applied to other aromatic amino acids and pharmacodynamic amines without changing the original chromophore. In contrast, upon UV irradiation of N-chloroacetyl 3,4-dimethoxyphenethylamine 3, ring closure occurred to yield two isomeric compounds 4 and 5 and a third isomeric compound 6 of completely unknown structure which was unyielding to n.m.r., mass spectra and other spectroscopic means of analysis. Hence an x-ray diffraction analysis of a single crystal of 6 was carried out.



Phases for the F_{hkl} values were obtained directly from the measured x-ray intensities by means of relationships (2) and (3). The calculated electron density function (1) contained 16 maxima which are plotted in the left portion of Fig. 3. The numbers next to the dots indicate the third coordinate. Connections were made between dots separated by 1.2 to 1.5A to reveal the geometry and structural formula of the molecule (right portion of Fig. 3) (Ref.11). This unusual product has two five-membered rings fused to two four-membered rings. The atoms can be identified as C, N or O either by chemical considerations or by weight and thermal factor analysis from the x-ray data. An electron density map computed from the differences between the observed structure factors and those calculated on the

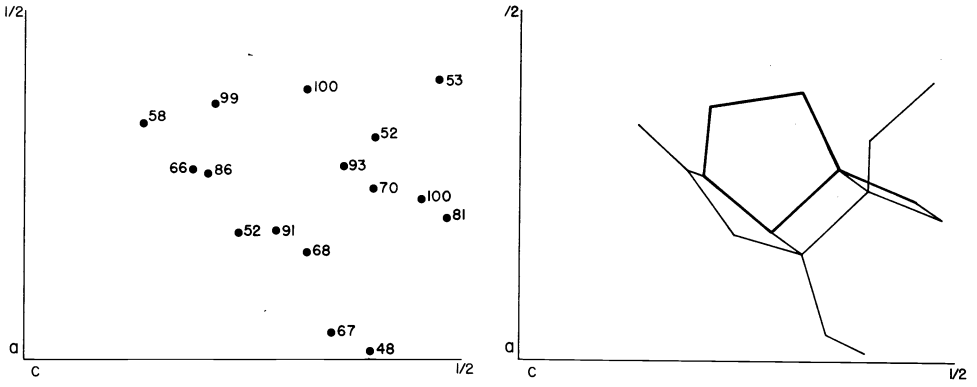


Fig. 3. Left portion represents the positions of the maxima of a three-dimensional electron density function calculated from experimental data. Numbers adjacent to the dots represent the third coordinate (X100). In the right portion, dots separated by 1.2-1.6A have been connected and the structure of the molecule is revealed.

basis of the known C, N and O atoms reveals the positions of the hydrogen atoms, Fig. 4.

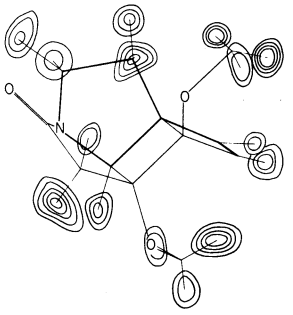
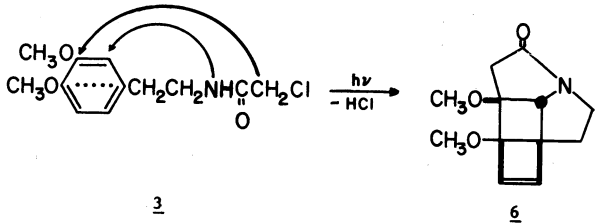


Fig. 4. Electron density contours, obtained from a difference map, for hydrogen atoms for molecule in Fig. 3 (Ref.11).

Thus the structural formula of a completely unknown substance has been established quite directly without any assumptions or prior knowledge and without the introduction of a heavy atom marker. At the same time, much other information is obtained at no extra effort, such as the stereoconfiguration about the four asymmetric carbon atoms, the bond lengths, bond angles, conformational angles, hydrogen bonds, if they exist, packing modes, etc.

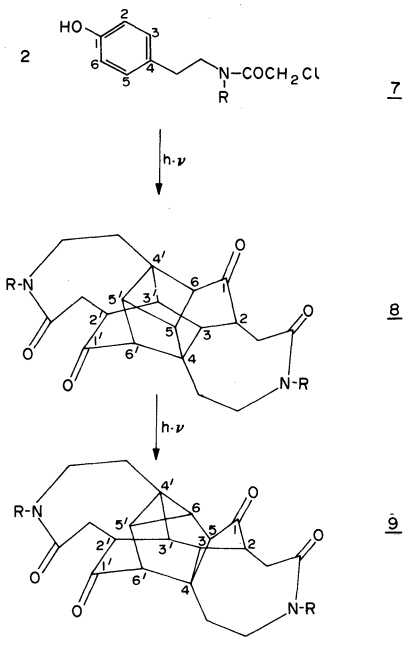
The changes that took place upon irradiation are shown schematically as follows:



3 6

From such information it is possible to postulate reaction mechanisms.

Substances differing only in the number of methoxy groups on the phenethylamine produce drastically different products upon irradiation under identical conditions (Ref.12 & 13). If a hydroxyl group is present instead of methoxy, two different dimeric substances 8 and 9 are produced in addition to simple ring closure (Ref.14):

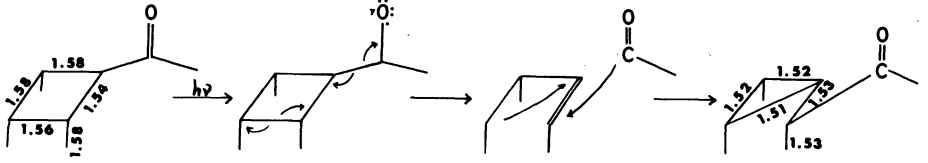


7

8

9

Again, the structural formulas of both 8 and 9 were established by crystal structure analyses (Ref.15). As indicated, dimer 9 is more stable than dimer 8 since the reaction is not reversible, although the cage in dimer 8 is more symmetrical than the cage in dimer 9. A possible mechanism for the rearrangement from the four-membered ring in 8 to the three-membered ring in 9 is as follows:



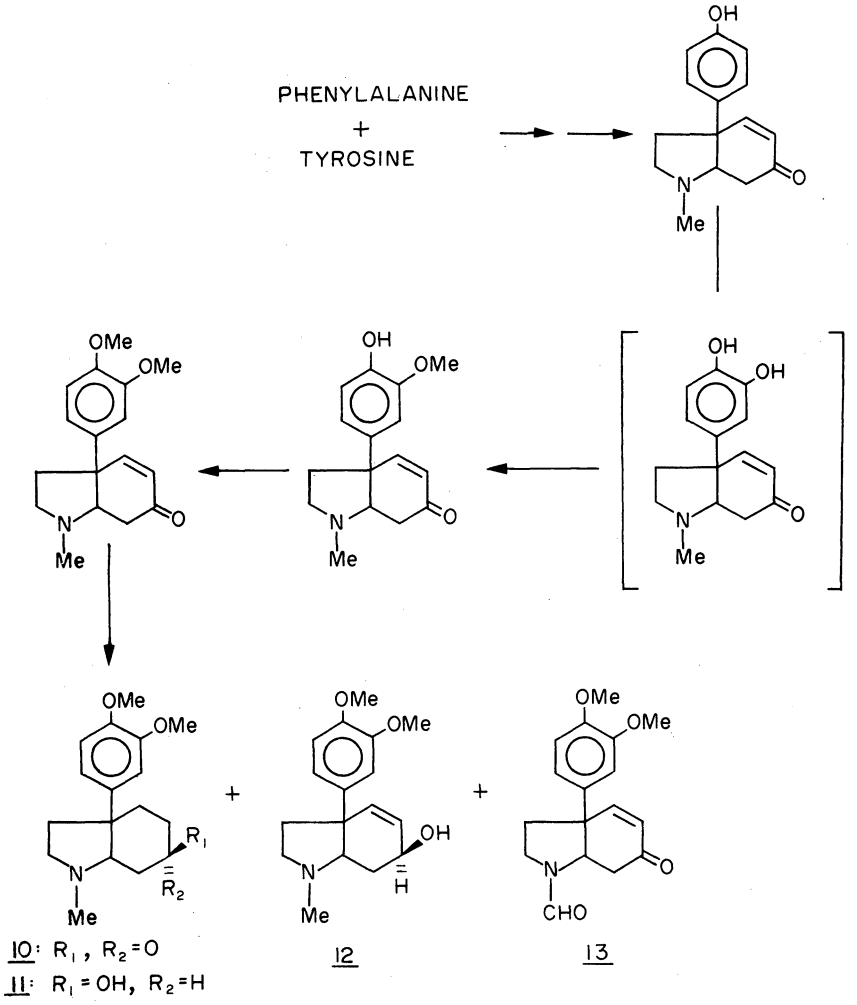
$_{\sigma}2_{a} + {}_{\sigma}2_{a}$ CYCLOREVERSION, OR

NORRISH TYPE I CLEAVAGE

An interesting observation is that the bond lengths in the area of rearrangement are much larger than normal in the intermediate, less stable dimer and are reduced to more normal values of 1.52-1.53A after rearrangement.

BIOSYNTHESIS

Biosynthesis of the mesembrine alkaloids occurring in Sceletium strictum, a South African succulent, begins with phenylalanine and tyrosine. It proceeds through many steps, some of which are indicated here, and finally produces the four congeners 10, 11, 12 and 13 (Ref.16). All the formulae were established by chemical and spectroscopic means except for product 13 which occurred in a very small amount. The formula remained in doubt until a minute single crystal was grown and subjected to x-ray diffraction analysis (Ref.17). As a bonus, the orientation of the dimethoxyphenyl moiety with respect to the fused rings was revealed, as well as the mode of puckering in the five- and six-membered rings, as shown in Fig. 5. The five-membered ring is in the classic envelope conformation with four coplanar atoms and the fifth atom 0.62A out of the plane, while the six-membered ring which contains one unsaturated bond and a carbonyl carbon also assumes an envelope conformation with five atoms nearly in a plane and the sixth atom out of the plane.



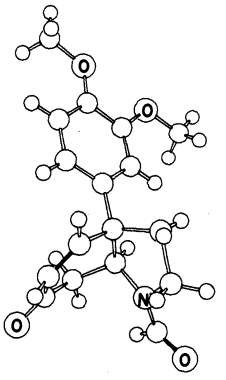


Fig. 5. Structure and conformation of the mesembrine alkaloid N-dimethyl-N-formylmesembrenone, 13 (Ref. 17).

FROG TOXINS

The brightly colored frogs occurring in Central and South America produce extremely potent toxins in their skins that the native Indians have used as arrow tip poisons. These venoms have very exciting pharmacological properties in that they cause very specific neuromuscular blocks (Ref.18 & 19). The quantity of venom in each frog is very small, thus necessitating the collection of many thousands of frogs in order to produce a minute amount of purified material (Ref.20). Thus the elucidation of the structural formula of each toxin by single crystal x-ray analysis was a vital step.

These arrow poison frogs belong to the genera Phyllobates and Dendrobates and some of their characteristics are summarized below:

DENDROBATIDAE

Phyllobates	Dendrobates	Colostethus
brightly colored arrow poison frogs		less strikingly colored; non-toxic alkaloids

Phyllobates aurotaenia (Colombia)

1. batrachotoxin, $C_{31}H_{42}N_2O_6$
2. homobatrachotoxin, $C_{32}H_{44}N_2O_6$
3. pseudobatrachotoxin
4. batrachotoxinin A, $C_{24}H_{35}NO_5$

Dendrobates histrionicus (Colombia)

1. histrionicotoxin, $C_{19}H_{25}NO$
2. dihydrohistrionicotoxin, $C_{19}H_{27}NO$
3. histrionicotoxin D, $C_{19}H_{29}NO$

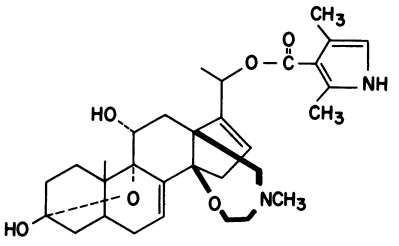
Dendrobates pumilio (Panama)

1. pumiliotoxin A, $C_{19}H_{33}NO_2$
2. pumiliotoxin B, $C_{19}H_{33}NO_3$
3. pumiliotoxin C, $C_{13}H_{25}N$

Dendrobates auratus

pumiliotoxins A, B, and C.

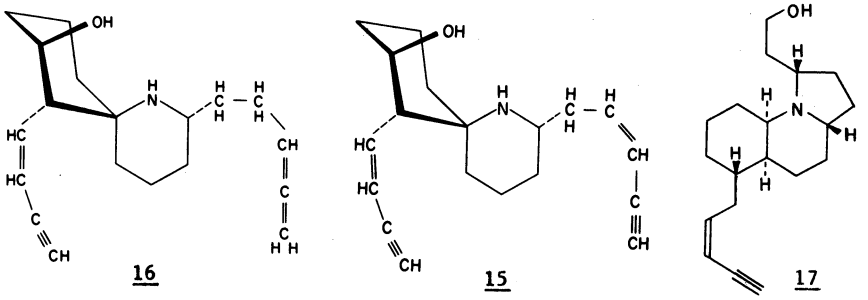
Batrachotoxin was originally extracted from the skins of 5000 frogs. The final purified products consisted of 11 mg batrachotoxin, 1 mg pseudobatrachotoxin and 7-8 mg of batrachotoxinin A. The crystal structure analysis was performed on a derivative of batrachotoxinin A (Ref.21) at which time its absolute stereoconfiguration was also determined by diffraction means (Ref.22). Batrachotoxin consists of batrachotoxinin A with a pyrrole ester on the 20α hydroxyl group. Its formula was deduced from that of batrachotoxinin A plus spectral data and eventually verified by synthesis (Ref.23 & 24):



14

This natural substance presented a number of novel features that were observed for the first time; namely, the 9α-OH, the 3α-9α oxide, the 3β-hemiketal, the 7-membered 13β-14β heterocyclic ring encompassing N-methylaminoethanol (a precursor of choline) and the Δ^{16} unsaturation. The stereochemistry of the cis A/B and cis C/D ring junctions is the same as in the cardenolide glycosides and bufodienolides.

Unlike the steroidal alkaloids present in the Phyllobates species, the alkaloids from the Dendrobates species are quite different. Unprecedented structural features have occurred in the histrionicotoxin 15 (Ref.25), dihydrohistrionicotoxin 16 (Ref.26) and histrionicotoxin D 17 (Ref.27):



16 15 17

These are the first examples of a spiro-ring system and of acetylenic and allenic bonding in the animal kingdom. The absolute stereoconfiguration and bond lengths and angles in 16 are shown in Figs. 6 and 7.

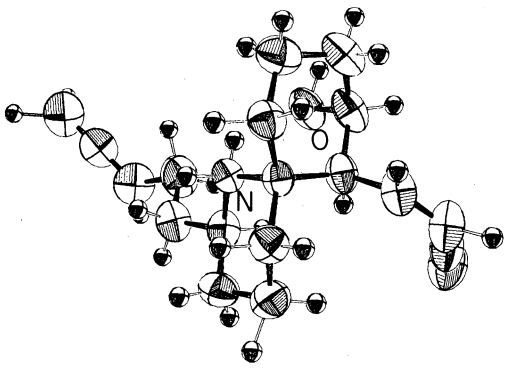


Fig. 6. Conformation and absolute configuration of dihydrohistrionicotoxin (Ref.26).

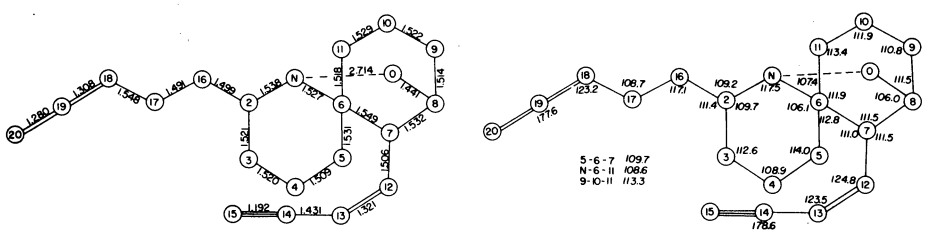
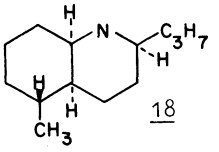


Fig. 7. Bond lengths and bond angles in dihydrohistrionicotoxin (Ref.26).

The formula of pumiliotoxin C 18 (Ref.28), reminiscent of conine (from a plant source),



18

contains a cis ring junction, as has been also found in 17. The structural formulas of pumiliotoxins A and B have not yet been elucidated and thus far it has not been possible to prepare crystals of these substances. Since some species of Dendrobates contain both histionicotoxins and pumiliotoxins, it is believed that both classes of alkaloids have the same precursor.

These alkaloids from the poison arrow frogs have become valuable tools for studying neurotransmission events in synapse and muscle. The batrachotoxin is associated with an irreversible increaseof membrane permeability to Na^+ ions and is an antagonist to tetrodotoxin activity (Ref.18).

CYCLIC PEPTIDES

The emphasis in this lecture thus far has been upon elucidating the structural formula and configuration. For many natural products, the molecular formula is readily derived from spectral data and/or chemical activity. The challenge lies in establishing the conformation, particularly for flexible molecules such as the cyclic peptides. These substances occur in nature and many of them have antibiotic, ion transport and other useful physiological properties that are associated with the manner of folding of the peptide backbone and side chains.

Coincidentally, cyclic hexaglycyl, the first cyclic peptide for which the crystal structure and conformation were determined (Ref.7), was also one of the first substances whose structure was determined by the direct method of phase determination. This crystal proved to be particularly interesting since there were four independent molecules in the cell, not related by symmetry. Thus, there were 98 atoms (aside from hydrogen atoms) that needed to be located. It proved to be a very unusual crystal because the molecules assumed four different conformations, side-by-side, in the same unit cell. Figure 8 illustrates the packing of the molecules in the unit cell while the four different conformers are shown individually in Fig. 9. This determination established that molecules can exist in more than one stable conformation, thus indicating that there is very little difference in energy between a number of different conformers.

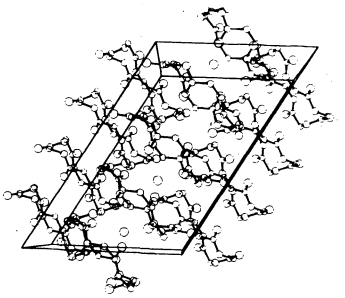


Fig. 8. Packing of cyclic hexaglycyl molecules in a unit cell (Ref.7).

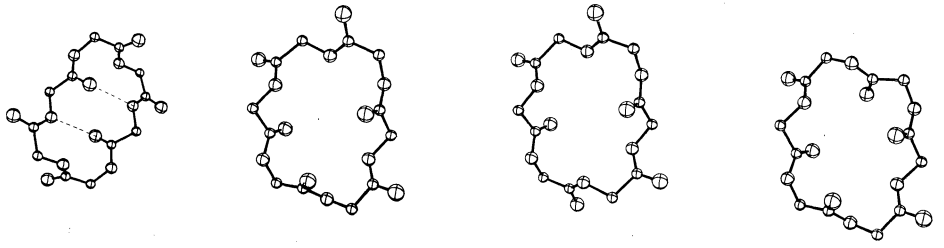


Fig.9. Four conformers of cyclic hexaglycyl occurring in the unit cell in Fig. 8.

The conformer illustrated at the left of Fig. 9 has two intramolecular hydrogen bonds, often called β-turns, β-bends or 4→1 bonds. This type of folding is a common feature in many linear peptides and proteins and this determination provided the parameters needed for model building, minimum energy calculations, comparison with spectroscopic data, etc. Figure 10 contains another view of hexaglycyl with intramolecular hydrogen bonds where it is being compared with the conformations of a linear tetrapeptide (Ref.29), a linear

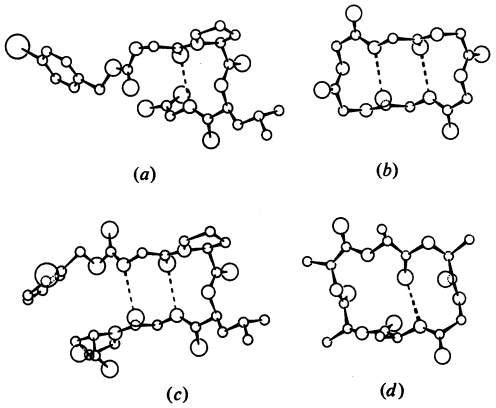


Fig. 10. Comparison of the conformations of: a) bromobenzoate of a linear tetrapeptide, b) cyclic hexaglycyl, c) bromobenzoate of a linear pentapeptide, d) backbone of the ring in Ferrichrome A. The dotted lines indicate intramolecular NH···OC bonds (Ref.30).

pentapeptide (Ref.30) and the cyclic hexapeptide ring of Ferrichrome A (Ref.31). The hydrogen bonds in the two linear molecules have almost identical conformational parameters to those of c-hexaglycyl, while the 4→1 bond in Ferrichrome A is different, particularly with respect to the middle peptide unit where the C=O group is directed upward instead of downward. It has been established that the c-hexaglycyl form occurs for peptide residues that are both levo, both dextro, or both Gly at the corners of the bend (called Type I) and that the Ferrichrome A form (called Type II) occurs for D,L or L,D or L,Gly sequences (Ref.32 & 33).

Scale drawings of the three types of β-bends found in oligopeptides, i.e. types I and II, already mentioned, and a bend in which the middle peptide groups has a cis conformation (Ref.34) are shown in Fig. 11. The numbers adjacent to the corner bonds indicate typical values for the torsional twist about the bonds in order to make the fold. (For a fully extended chain in the trans conformation, the angular values would be 180°.) The 4→1 type of hydrogen bond contains 10 atoms in the ring.

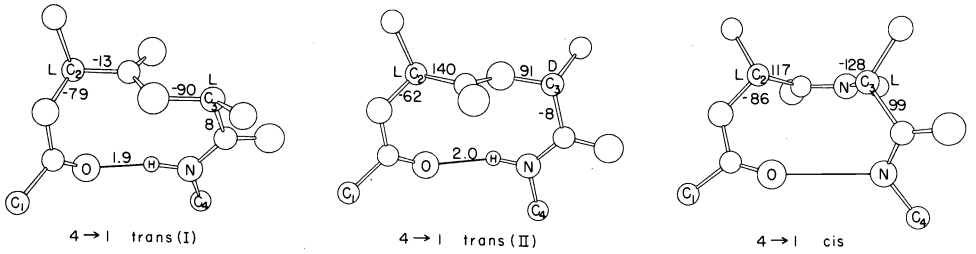


Fig. 11. Three types of 4→1 intramolecular hydrogen bonds. The numbers shown near atoms C_2^{α} and C_3^{α} indicate typical experimental values found for the rotations about the bonds to C_2^{α} and C_3^{α}. An extended chain in the trans conformation would have angular values of 180°.

Intramolecular hydrogen bonds of a 3→1 type, containing only 7 atoms in the ring, have been discussed in the literature for a number of years but have been observed in molecules in the crystalline state only recently. The conformation of dihydrochlamydocin (Ref.35), a metabolite from Diheterospora chlamydosporia (Ref.36), is illustrated in Fig. 12. This cyclic tetrapeptide contains two intramolecular 3→1 hydrogen bonds and the details for one of these bonds are shown in Fig. 13. The N...O separations are of the order of 2.8 - 2.9A and the H...O separations are ~2.1 - 2.2A. The all trans tetrapeptide ring in dihydrochlamydocin is of further interest since it has been demonstrated geometrically that four planar peptide units, each in the trans conformation, cannot form a closed ring (Ref.37). Although peptide units, $C^{\alpha}-C'(=O)-N-C^{\alpha}$, whether cis or trans, are

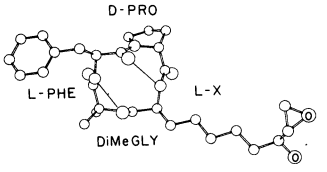


Fig. 12. Conformation of dihydrochlamydocin. The fine lines indicate the two 3→1 intramolecular NH···OC bonds (Ref.35).

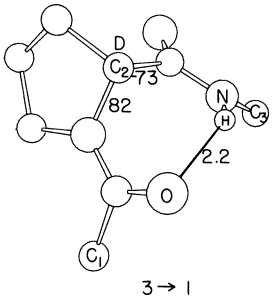


Fig. 13. Details of a 3→1 intramolecular hydrogen bond. The numerical values near C_2^{α} indicate the angular twists about $N_2—C_2^{\alpha}$ and $C_2^{\alpha}—C'$.

generally quite planar, an accommodation has been made in this substance so that ring closure could be effected. A rotation has taken place about each C'-N bond from 180° (trans, planar) to values of +162°, -166° +156° and -164°.

Another example of a 3→1 hydrogen bond has occurred in the synthetic cyclic pentapeptide c-Gly-Pro-Gly-D-Ala-Pro (Ref.37) shown in Fig. 14. Four of the peptide units are trans

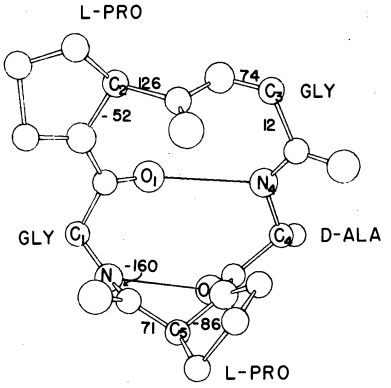


Fig. 14. Conformation of a cyclic pentapeptide containing both a 3→1 and a 4→1 intramolecular hydrogen bond (Ref.38).

and planar while the fifth one is trans and non-planar with a rotation of 160° about the C_5'-N_1 bond, comparable to the non-planar peptide units in dihydrochlamydocin. The c-pentapeptide also contains a 4→1, type II, hydrogen bond.

In the folding of the backbone of cyclic peptides, 5→1 intramolecular hydrogen bonds, encompassing 13 atoms, have also been observed. Figure 15 shows the details of such bonding in complexed valinomycin (Ref.39 & 40), where the peptide units are L, L and D. A comparable 5→1 hydrogen bond, where all the residues are levo has also been observed in uncomplexed antamanide, which will be discussed later in the manuscript. Meanwhile, Fig. 16 shows a view of the conformation of valinomycin. The molecule contains a 36-membered ring that loops up and down in a sinusoidal fashion and that is stabilized by four 4→1 and two 5→1 hydrogen bonds. The latter two bonds transverse the flattened ring.

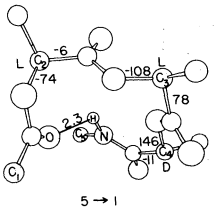


Fig. 15. Details of a 5→1 intramolecular hydrogen bond occurring in valinomycin (Ref.40).

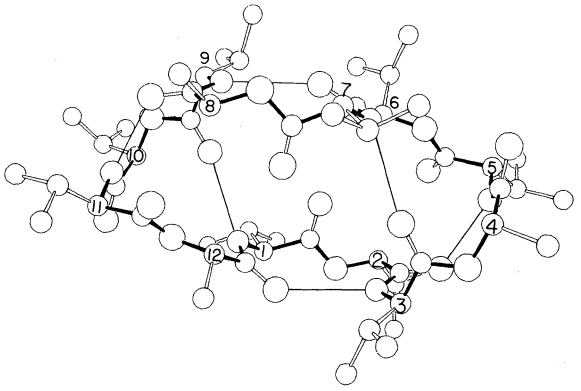


Fig. 16. Conformation of valinomycin looking into the cavity. The molecule contains four 4→1 hydrogen bonds and two 5→1 hydrogen bonds (across the flattened ring) (Ref.39).

A number of different arrangements for intramolecular hydrogen bonds that stabilize folds in peptide backbones have already been demonstrated. There is still another possibility, and that is the absence of intramolecular hydrogen bonding. The conformation of cyclic-L-Leu-L-Tyr-ô-Avaler-ô-Avaler (Ref.41), a synthetic inhibitor of chymotrypsin (Ref.42), Fig. 17, is very similar to one of the four conformers of cyclic hexaglycyl, Fig. 9. The particular point of interest is not only the absence of a hydrogen bond across the ring, the N(3)...O(4) distance is 4.59A, but also that the proton on N(3) does not participate in hydrogen bond formation, even though it is quite accessible and not blocked by side chains. The lack of hydrogen bond formation by NH groups is of fairly frequent occurrence in cyclic oligopeptides.

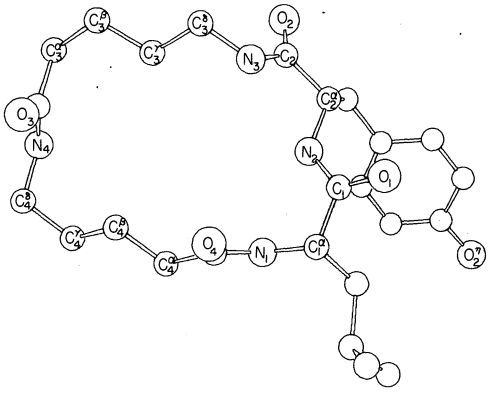


Fig. 17. Conformation of cyclic-L-Leu-L-Tyr-ô-Avaler-ô-Avaler, a molecule without any intramolecular hydrogen bonds (Ref.41).

It would be useful to establish certain norms of conformational behaviour for peptide molecules. Thus far, however, each new structure determination reveals some unexpected features such as a new configuration for internal hydrogen bonds, nonplanar peptide units, cis peptide groups or unusual folding or unfolding of the peptide chain.

COMPLEXED AND UNCOMPLEXED ANTAMANIDE

Antamanide, a cyclic decapeptide found in the poisonous green mushroom Amanita phalloides, has the following sequence:

```
 8   9   10  1   2
Pro-Phe-Phe-Val-Pro
 |                 |      (all L-).
Pro-Phe-Phe-Ala-Pro
```

Wieland and coworkers have isolated, sequenced, synthesized, and characterized the biological activity of this peptide (Ref.43). It acts as an antidote to the toxin phalloidine that occurs in much larger quantities in the same mushroom. The mode of action of antamanide is not yet clearly understood but there are strong indications that the antamanide acts upon the membranes of liver cells to prevent the toxin from entering the liver.

Antamanide forms alkali-metal complexes with a strong preference for Na^+ over K^+ ions. The alkali metal complex is most stable in a lipophilic environment. The chief objective of the present research was to establish the conformation of the uncomplexed antamanide and the changes in the folding or unfolding of the peptide backbone that accompany the alkali complex formation. Knowledge of the conformation and changes in conformation accompanying complexation or change in the polarity of the solvent may aid in the understanding of the binding to receptor sites and may serve as a model for ion carriers.

The conformation of the Na^+ complex of a biologically active synthetic analog of antamanide is shown in Fig. 18 (Ref.44). There are a number of unique features to notice: (1) The peptide backbond (darkened bonds) is folded into a shape resembling the periphery of a saddle. (2) Peptide units have usually been found to be planar and to have the trans conformation. Here, eight of the peptide units are trans, but two of them have the cis conformation (see e.g. the cis amide bond between the $C^\alpha(7)$ and $C^\alpha(8)$; similarly for $C^\alpha(2)$ and $C^\alpha(3)$). A survey of peptides containing cis amide bonds (Ref.32) indicates that the cis bond occurs only in those peptide units that have a substituted N atom, e.g., in sarcosine or proline. However, peptide units with a substituted N atom do not necessarily have the cis conformation, as is illustrated in this molecule, where two prolyl residues occur in the cis conformation and two other prolyl residues occur in the trans conformation. (3) The Na^+ resides in a cup formed by the folding of the peptide chain. It is liganded to four carbonyl oxygen atoms of the peptide. A fifth ligand is formed with the O atom of a solvent molecule, ethyl alcohol in this case. Thus the Na^+ is pentacoordinate. The polar cavity is plugged with a solvent molecule and the lipophilic end of the solvent molecule completes the lipophilic exterior surface. (4) The side groups, particularly the phenyl groups, are folded up against the molecule, to make a compact spheroidal shape. Two phenyl groups on two phenylalanine residues are omitted from this drawing for clarity. They fold up against the molecule in front and in back. (5) There are six NH groups available for hydrogen bonding. Two of them enter into intramolecular hydrogen bonds (indicated by the dashed lines) of the 4→1 type previously described. Two others participate in hydrogen bonds between molecules, and two NH groups do not participate in any hydrogen bonds at all.

Antamanide forms complexes with Li^+ as well as Na^+. Fig.18 also shows a comparison of the two different alkali complexes, Li^+ antamanide.CH_3CN (Ref.45 & 46) and Na^+[Phe^4Val^6]antamanide.C_2H_5OH (Ref.44). Even though the alkali ions are different, the solvent molecules are different, side chains in residues 4 and 6 are different, and the crystallographic packing is different, the two complexes are isostructural, thus indicating that the conformation of the alkali complex of antamanide is independent of its environment.

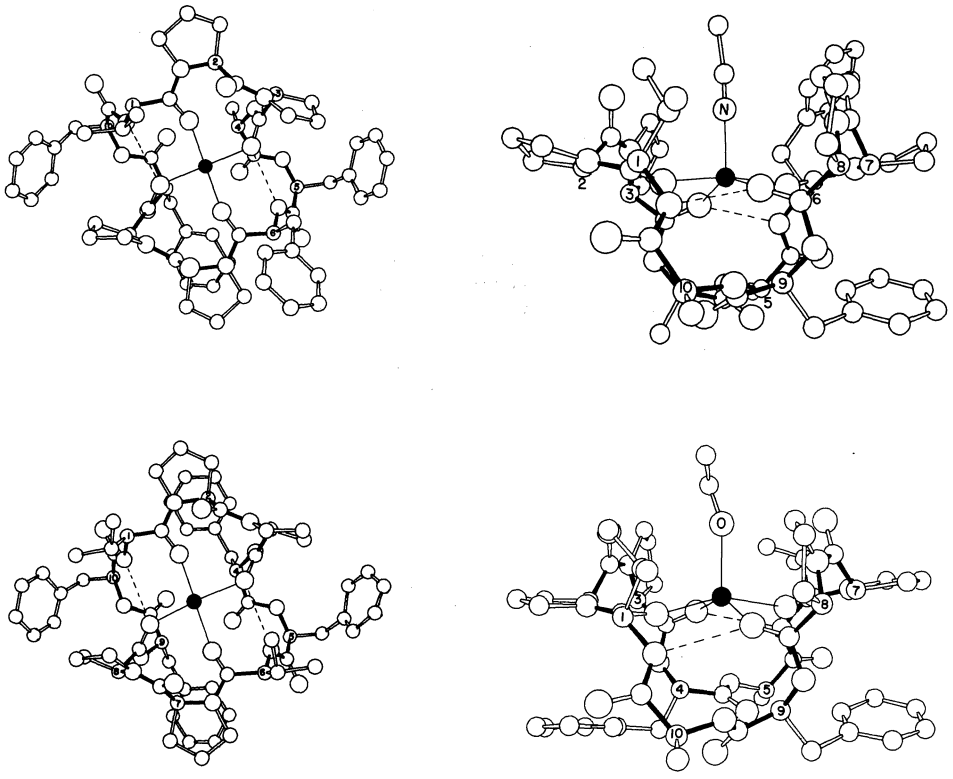


Fig.18. Upper diagrams: Two views of Li^+antamanide.CH_3CN. Lower diagrams: Two views of Na^+[Phe^4Val^6]antamanide.C_2H_5OH. The alkali metal ion is represented by a black dot. For clarity the solvent molecules have been omitted from the views on the left and two phenyl groups that fold up against the middle of the molecule have been omitted from the views on the right (Ref.44, 45 & 46).

Now we turn our attention to uncomplexed antamanide. From spectroscopic evidence in solution, it is known that the conformation is quite sensitive to the polarity of the solvent (Ref. 47). The crystal used for the x-ray diffraction analysis was grown from a mixture of n-hexane and methyl acetate, a very nonpolar solvent. In Fig.19 (Ref.48), it is immediately obvious that the ring in antamanide has become unfolded; in fact, it is completely elongated. The only features of the conformation that are similar to the alkali complex are the pairs of adjacent prolyl groups, one <u>trans</u>, one <u>cis</u>.

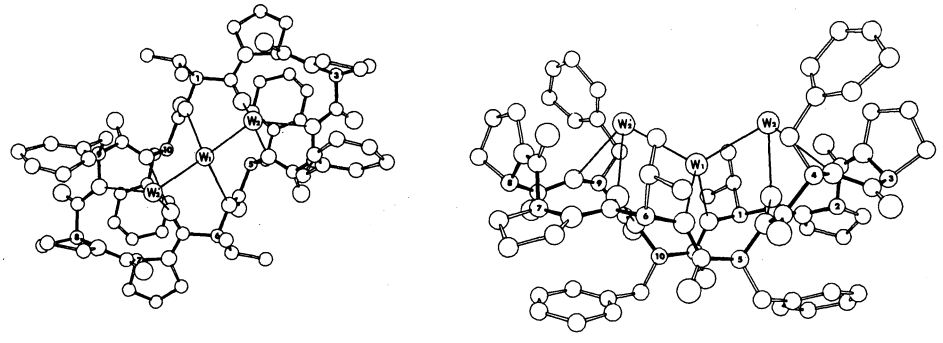


Fig. 19. Two views of the conformation of antamanide crystallized from nonpolar solvents. Three molecules of water, designated by W, are contained in the molecular cavity (Ref.48).

Two NH groups make hydrogen bonds with carbonyl oxygens across the ring. Again, this is a different mode of hydrogen bonding than any that has been observed earlier. It is a 5→1 type with all L-residues with the middle peptide unit in the cis conformation. A very interesting feature of this structure are the three H_2O molecules residing in the interior. They form hydrogen bonds with four of the six NH groups of the molecule and account for the relative rigidity of the 30-membered peptide ring. These water molecules appear to be an integral part of the peptide molecule since they have been retained even after extensive drying of the antamanide and the use of an extremely nonpolar solvent. This molecule can indeed be considered as an H_2O-complex.

The three water molecules reside in a shallow cup with the hydrophobic side chains forming a wall around them. Again all the side groups fold up against the peptide ring to make a very compact globular entity. It seems unlikely that the uncomplexed molecule in this conformation recognizes a Na^+ ion, rapidly releases its H_2O and folds up to engulf the Na^+ ion. Antamanide crystals have now been prepared from various polar solvents for an x-ray diffraction examination. We shall have to wait for a while to see whether the conformation obtained from polar solvents will be closer to that of the alkali metal complex.

REFERENCES

1. O. Kennard and D.G. Watson, Molecular Structures and Dimensions, Bibliography, Vols.1-7, Oosthoek, Scheltema and Holkema, Utrecht (1970-1975).
2. A.L. Patterson, Phys. Rev. 46, 372-376(1934); Zeits. F. Kristallog. 90, 517-542 (1935).
3. See e.g. G.H. Stout and L.H. Jensen, X-Ray Structure Determination, pp.270-299, Macmillan, London (1968).
4. See e.g. B.W. Matthews, "X-Ray Structure of Proteins" in The Proteins, 3rd Ed.Vol.III, Eds. H. Neurath and R.L. Hill, Academic Press, New York (in press).
5. J. Karle and H. Hauptman, Acta Cryst. 3, 181-187 (1950).
6. See e.g. J. Karle and I.L. Karle, Acta Cryst. 21, 849-859 (1966).
7. I.L. Karle and J. Karle, Acta Cryst. 16, 969-975 (1963).
8. I.L. Karle and J. Karle, Acta Cryst. 17, 835-841 (1964).
9. J. Karle and H. Hauptman, Acta Cryst. 9, 635-651 (1956).
10. O. Yonemitsu, P. Cerutti and B. Witkop, J. Amer. Chem. Soc. 88, 3941-3945 (1966).
11. I.L. Karle, J.W. Gibson and J. Karle, Acta Cryst. B25, 2034-2039 (1969); O. Yonemitsu, Y. Okuno, Y. Kanaoka, I.L. Karle and B. Witkop, J. Amer. Chem. Soc. 90, 6522-6523 (1968).
12. O. Yonemitsu, B. Witkop and I.L. Karle, J. Amer. Chem. Soc. 89, 1039-1040 (1967); I.L. Karle, J. Karle and J.A. Estlin, Acta Cryst. 23, 494-500 (1967).
13. O. Yonemitsu, H. Nakai, Y. Kanaoka, I.L. Karle and B. Witkop, J. Amer. Chem. Soc. 92, 5691-5700 (1970); I.L. Karle and J. Karle, Acta Cryst. B26, 1276-1282 (1970).
14. T. Iwakuma, H. Nakai, O. Yonemitsu, D. Jones, I.L. Karle and B. Witkop, J. Amer. Chem. Soc. 94, 5136-5139 (1972).
15. D.S. Jones and I.L. Karle, Acta Cryst. B30, 617-623 (1974).
16. Jean M. Karle, Ph.D. Dissertation, Duke University, Durham (1976). Also see P.W. Jeffs, H.F. Campbell, D.S. Farrier, G. Ganguli, N.H. Martin and G. Molina, Phytochem. 13, 933-945 (1974).
17. Jean M. Karle, Acta Cryst., in press
18. E.X. Albuquerque, J.W. Daly and B. Witkop, Science 172, 995-1002 (1971).
19. J. Daly and B. Witkop, Aldrichimica Acta 3, No.4 3-6 (1970).
20. T. Tokuyama, J. Daly, B. Witkop, I.L. Karle and J. Karle, J. Amer. Chem. Soc. 90, 1917-1918 (1968).
21. I.L. Karle and J. Karle, Acta Cryst. B25, 428-434 (1969).
22. R.D. Gilardi, Acta Cryst. B26, 440-441 (1970).
23. E. Gossinger, W. Graf, R. Imhof and H. Wehrli, Helv. Chim. Acta 54, 2785-2793 (1971) and references therein.
24. T. Tokuyama, J. Daly and B. Witkop, J. Amer. Chem. Soc. 91, 3931-3938 (1969).
25. J.W. Daly, I. Karle, C.W. Myers, T. Tokuyama, J.A. Waters and B. Witkop, Proc. Nat. Acad. Sci. USA, 68, 1870-1875 (1971).
26. I.L. Karle, J. Amer. Chem. Soc. 95, 4036-4040 (1973).
27. I.L. Karle, J.W. Daly, T. Tokuyama and B. Witkop, Manuscript in preparation.
28. J.W. Daly, T. Tokuyama, G. Habermehl, I.L. Karle and B. Witkop, Liebigs Ann. Chem. 729, 198-204 (1969).
29. T. Ueki, T. Ashida, M. Kakudo, Y. Sasada and Y. Katsube, Acta Cryst. B25, 1840-1849 (1969).
30. T. Ueki, S. Bando, T. Ashida and M. Kakudo, Acta Cryst. B27, 2219-2231 (1971).
31. A. Zalkin, J.D. Forrester and D.H. Templeton, J. Amer. Chem. Soc. 88, 1810-1814 (1966).
32. See e.g. I.L. Karle, "The State of the Art of X-Ray Crystallography of Peptides" in Peptides: Chemistry Structure and Biology, eds. R. Walter and J. Meienhofer, Ann Arbor Science, Ann Arbor, pp.61-84 (1975).

33. C.M. Venkatachalam, Biopolymers 6, 1425-1436 (1968).
34. Y. Iitaka, H. Nakamura, K. Takada and T. Takita, Acta Cryst. B30, 2817-2825 (1974).
35. J.L. Flippen and I.L. Karle, Biopolymers 15, 1081-1092 (1976).
36. A. Closse and R. Huguenin, Helv. Chim. Acta 57, 533-545 (1974).
37. N. Go and H.A. Scheraga, Macromolecules 3, 178-187 (1970).
38. I.L. Karle, L. Pease and C. Watson. Manuscript in preparation.
39. W.L. Duax, H. Hauptman, C.M. Weeks and D.A. Norton, Science 176, 911-914 (1972).
40. I.L. Karle, J. Amer. Chem. Soc. 97, 4379-4386 (1975).
41. I.L. Karle, Macromolecules 9, 61-66 (1976).
42. V.I. Tsetlin, E.N. Shepel, V.T. Ivanov and Yu. A. Ovchinnikov, Biorg. Khim. 1, 407-415 (1975).
43. See e.g. Th. Wieland in Chemistry and Biology of Peptides, Ed. J. Meienhofer, Ann Arbor Science, Ann Arbor, pp.377-396 (1972).
44. I.L. Karle, Biochem. 13, 2155-2162 (1974).
45. I.L. Karle, J. Karle, Th. Wieland, W. Burgermeister, H. Faulstich and B. Witkop, Proc. Nat. Acad. Sci. USA 70, 1836-1840 (1973).
46. I.L. Karle, J. Amer. Chem. Soc. 96, 4000-4006 (1974).
47. See e.g. Yu. A. Ovchinnikov and V.T. Ivanov, Tetrahedron 31, 2177-2209 (1975).
48. I.L. Karle, J. Karle, Th. Wieland, W. Burgermeister and B. Witkop, Proc. Nat. Acad. Sci. USA, 73, 1782-1785 (1976).