16. Molecules and Crystals, 1926–1970[1])

by **J. M. Robertson**
Chemistry Department, University of Glasgow, Scotland

(12. XI. 71)

Summary. Some features in the history and development of the *X*-ray crystallography of organic molecules are described, from the simple but fundamental structures of the 1920's and 1930's to the complex biological structures of today. The most important developments were probably the introduction of phase determining methods based on the use of heavy atoms, and the invention of the electronic digital computer.

It has been suggested that I should try and describe the history and development of chemical crystallography. However, in the brief time available this subject is much too wide and varied. So instead I have decided to make it much more personal and deal mainly with those aspects in which I have been involved. These relate almost entirely to the study of organic molecules by *X*-ray diffraction methods, and this is indeed quite a big enough subject.

During the early 1920's I was working for my Ph. D. in organic chemistry. I was studying certain members of the bicyclic sesquiterpene series, including caryophyllene, by the classical methods of chemical degradation, oxidation by various reagents, and so on. This was not successful. The reagents were unsuitable, and nothing was known about the structure of caryophyllene at that time. I therefore decided to try a more physical approach.

The study of organic crystals by *X*-rays was then just beginning, and *W. H. Bragg* had measured the unit cells of naphthalene and anthracene. Most of my terpenoid compounds were non-crystalline, but I had prepared a new caryophyllene derivative, isoclovene hydrochloride, which was nicely crystalline. So I took these crystals to the Royal Institution in London with a two-year Fellowship intending to work out the structure by *X*-ray analysis. This approach was successful, but the task took thirty years instead of two. I remember long afterwards describing this structure at a Chemical Society meeting in London, and afterwards *Derek Barton* congratulated me on having finally completed my Ph. D. after thirty years of effort! Of course, I did solve some other structures as well during that time, but he did not mention this.

It is largely the history of these thirty years that I wish to discuss today, and the advances that have made the analysis of these and even more complex molecules possible.

During the 1920's a lot of time was occupied in building our own *X*-ray tubes and spectrometers, making our own diffusion pumps for high vacuum, and so on. Quite a lot of work on organic crystal structures was published, but it was mainly guess work and nearly all wrong. But there were some notable exceptions. The structure of hexamethylenetetramine, which is cubic, was completely solved by

[1]) Lecture presented at the meeting of the Swiss Chemical Society, Fribourg, 8. 10. 1971; published by special decision of the Editorial Committee.

Dickinson & Raymond in 1923. Some other high symmetry crystals like the benzene hexahalides, urea, and thiourea were also rather completely determined a little later. Such high symmetry in organic crystal structures is, however, very rare.

The achievement which had the greatest impact on organic chemistry during this decade was *Kathleen Lonsdale*'s work on hexamethylbenzene. She was able to show that the benzene ring is a regular planar hexagon, and she measured its dimensions accurately for the first time. The crystal is triclinic with only a centre of symmetry, but the problem was simplified by the fact that all the carbon atoms lie in one crystal plane. Even so, there are 24 parameters to determine, and this was a formidable undertaking in those days. Commenting on this result at that time, Professor *C. K. Ingold* wrote 'One paper like this brings more certainty into organic chemistry than generations of activity by us professionals'.

By 1930 I had decided that my sesquiterpenoid structures like caryophyllene of unknown chemical constitution were too difficult, and in fact impossible unless there was some method of solving the fundamental phase problem in *X*-ray crystallography. So I concentrated on the structures of naphthalene and anthracene. Although it was now likely that these molecules were planar there was no simplifying feature beyond a centre of symmetry, and the molecules could be anywhere in the unit cell. Tedious trial and error calculations based on possible models had to be made. When rough agreements had been obtained it was possible to carry out refinements by the *Fourier* series method. With no computers at that time this also entailed a great deal of work. This was the first organic crystal without any simplifying features which had been solved, and the results were exciting. Fig. 1 shows the principal projection obtained by the *Fourier* synthesis, and the molecule is very clearly outlined. The atoms which are close together in this projection are not separately resolved, but their positions were accurately measured in other projections. The molecule turns out to be accurately planar with an average C–C bond length of 1.41 Å. There is an indication of hydrogen atoms, but they are very weak.

At the time when I published this work in 1933, *Linus Pauling* was working on a theory of the chemical bond, and he came to the conclusion that in a molecule like anthracene the C–C bond length should vary a little in different parts of the molecule. He wrote to me to say that he could detect these variations in my published maps. However, the accuracy (about $\pm$ 0.03–0.04 Å) was not really sufficient to confirm this.

It was not until many years later, in 1950, that I was able to complete a more accurate analysis in three dimensions, which is shown in Fig. 2. Here the electron density is evaluated at every point in the mean plane of the molecule and all the carbon atoms are very clearly resolved. The bond distances can be estimated to within a hundredth of an Angstrom, and the variations predicted by *Pauling* are fully confirmed. The distances vary from about 1.37 to 1.44 Å, the first established case of such variations in an aromatic hydrocarbon structure.

In all these pictures of molecules obtained by *X*-ray diffraction, the hydrogen atoms are very weak and ill defined because there is only one electron available to scatter the *X*-rays. However, similar diffraction effects can be obtained with a beam of neutrons and in this case it is the atomic nuclei and not the electrons that do the scattering. Now, the scattering power of the nucleus for neutrons varies abruptly

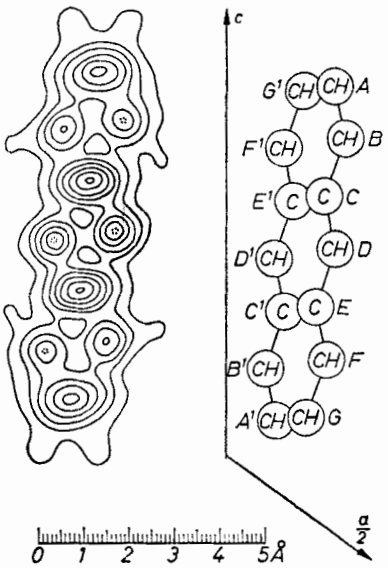

Fig. 1. *Electron density projection of the anthracene molecule*

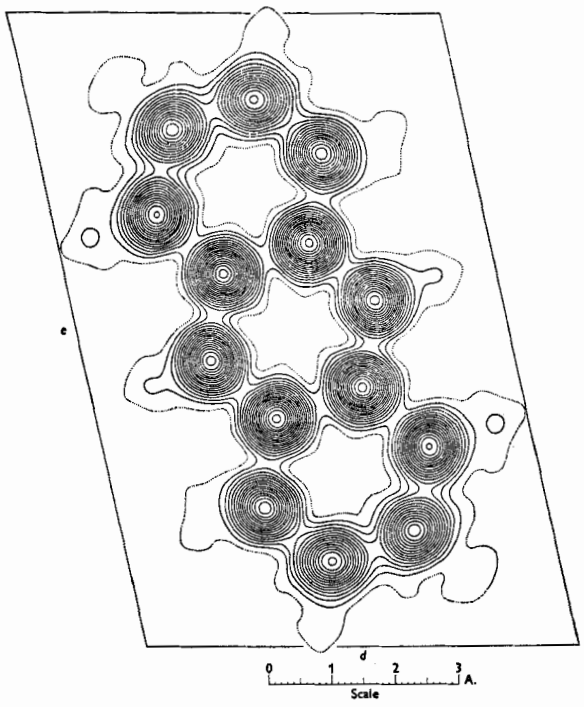

Fig. 2. *Electron density distribution in the mean plane of the anthracene molecule*

from element to element, and in the case of hydrogen, and especially deuterium, it is as large or larger than that for many heavy elements. So if we substitute deuterium for hydrogen and use neutrons instead of *X*-rays we can get a picture of the molecule that shows the hydrogen positions very clearly indeed. I am indebted to Dr. *J. C. Speakman* for Fig.3 which shows the naphthalene molecule with eight deuterium atoms attached ($C_{10}D_8$) calculated from the neutron diffraction pattern.

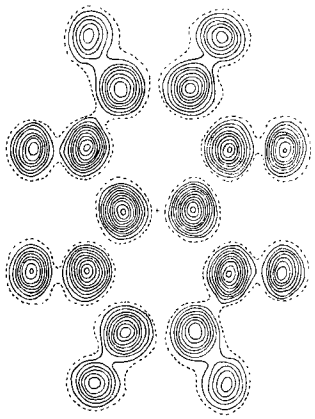

Fig. 3. *Neutron diffraction* Fourier *map of the naphthalene molecule,* $C_{10}D_8$ (*J. C. Speakman*)

During the 1930's many other chemically well known molecules were accurately studied by these methods, including some non-centrosymmetrical ones like resorcinol. The results were of great importance in providing exact tests for chemical theory, in the evaluation of hydrogen bonding, etc.

But at this time there was no way of solving the structures of molecules of unknown chemical constitution. Given 20 or 30 atoms that might be anywhere we cannot possibly set up a trial model as a starting point. There are far too many variables.

In general, measurement of cell dimensions was of little help. But there was at least one very notable exception. In 1932 *J. D. Bernal* made a few simple measurements of cell dimensions and optical properties for a number of steroids, and he was able to show that the then accepted chemical formula must be wrong. This was an essential step in the chemical elucidation of the correct structural formula, and led to immense advances in the organic chemistry of the steroids.

However, to find the precise position of every atom in a completely unknown structure by *X*-ray methods is a different matter. To do this we have somehow to solve the fundamental phase problem. We have to find some way of determining the phase of the diffracted wave relative to some chosen origin. Is it a peak, or a trough, or something in between? Intensities give the amplitudes, but to apply the *Fourier* series method we require to know the phases as well.

As the reflections have to be measured separately in time and space no *physical* method is possible. We can try *mathematical* methods for solving the structure from a knowledge of the amplitudes only without the phases, and in recent years this approach has been very successful indeed. The foundations were laid in the 1930's

with the discovery of the *Patterson* vector method in 1934, and the *Harker* synthesis in 1936. But for large molecules containing many carbon atoms of the same weight, these methods were too difficult at that time.

There is, however, a *chemical* method which has effectively solved the problem. This was developed in the course of my phthalocyanine work from 1935 onwards. The diffracted wave which we measure is built up from the contributions from all the atoms in the structure, and it may be either a peak or a trough (or something in between) at our chosen origin. We now do a rather difficult chemical experiment and insert another fairly heavy atom at or near this chosen origin, doing this with as little disturbance as possible to the remainder of the structure. The wave scattered from this additional atom must correspond to a peak at the origin, because that is where we have put the atom. If the original wave had a peak at this point, the resultant wave in the new heavy atom compound will have a much higher peak (intensity increased). On the other hand, if the original wave had a trough at this point, the algebraic resultant in the heavy atom compound will be a diminished peak (intensity reduced). In this way we have converted unknown phase differences which cannot be measured into amplitude differences which can be measured.

The case where we have neither a peak or a trough at the chosen origin but something in between is more difficult. The principle is the same, but in general we require two isomorphous heavy atom derivatives in addition to the parent compound for a complete and unambiguous solution. Very often, however, a solution may be obtained by successive approximations from a single heavy atom derivative, basing the phases entirely on the known heavy atom contribution.

This is a much over-simplified statement of the heavy atom and isomorphous substitution methods, which have been the principal means employed during the last 30 years for solving unknown chemical structures.

As I have mentioned, these methods were first developed and applied to organic molecular structures in my phthalocyanine work. The results are shown in Fig. 4. The large molecule has four benzene rings attached to a 16-membered porphyrin-like

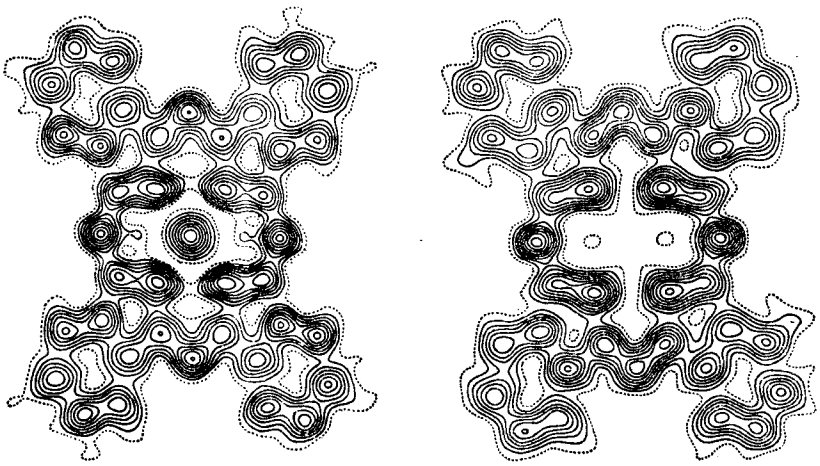

Fig. 4. *Electron density projections of metal free and nickel phthalocyanine*

ring of carbon and nitrogen atoms. A metal atom can be attached in the centre, and this can be removed, or another metal substituted, without appreciably altering the overall structure. The compounds are nicely crystalline and the conditions are ideal for the application of the method I have just outlined. By comparing the intensities of corresponding *X*-ray reflections from the two compounds, all the phases can be determined without ambiguity.

The contour maps, which show projections of the electron density, were prepared by summing the appropriate double *Fourier* series. In doing this calculation, nothing has been assumed about the structure of the molecule. We have not even assumed that it consists of atoms. As well as showing the structure of the molecule in complete detail, this result may be regarded as a direct physical proof of *Dalton*'s atomic theory!

I think the most exciting moment in my life was when I first calculated this result in 1936, because it obviously provided the key for solving any chemical structure of unknown constitution, no matter how complex. Indeed, about two years later I wrote an optimistic paper pointing out that even a protein structure like insulin containing thousands of atoms might be solved by these methods if suitable heavy atom derivatives could be prepared.

This has now been done, but it could not be done at that time. Although we knew the way, we could not solve even much smaller natural product molecules at that time, and I would now like to explain why. It was all a matter of arithmetic.

For complex molecules it is essential to work in three dimensions, in order to resolve all the atoms by *Fourier* series methods. The calculations are simple but very tedious. They generally involve evaluating triple series of the type summing

$$\varrho(xyz) = \sum_{-\infty}^{+\infty}\sum_{-\infty}^{+\infty}\sum_{-\infty}^{+\infty} \frac{F(hkl)}{V} \cos 2\pi \left(\frac{hx}{a} + \frac{ky}{b} + \frac{lz}{c}\right)$$

over all values of h, k, l and x, y, z. In a complex structure there may be 5000 or more terms in this series covering all values of h, k and l. And the series has to be summed at a sufficient number of points (values of x, y and z) throughout the unit cell to give a complete picture of the structure. This may involve 50 or more points along each of the axes, a, b and c; say 100,000 points in all. So we may have a total of 500,000,000 terms of this type to evaluate and add together for a complete structure. This and similar calculations will usually have to be repeated many times as we approach the true structure by successive approximations.

In the early days this was quite impossible, and we had to confine ourselves to small molecules and two-dimensional projections. It was a very frustrating situation, because although we knew how to do the work, it was impossible to complete the calculations within a life-time, even with a team of people.

The next big advance in crystallography began with the invention of the electronic digital computer. Nowadays calculations of this type can be completed within an hour, on the fastest machines. But the computer did not become available till after 1950, and even then the first machines were very crude. In the meantime a number of important and heroic determinations of complex organic structures had been made.

Some of the earliest were in the steroid series, when *Carlisle & Crowfoot* (*Dorothy Hodgkin*) solved the structure of cholesterol iodide in 1945, and *Crowfoot & Dunitz* obtained all the atomic positions in a complex calciferol derivative in 1948. The important penicillin structure was also solved by *Dorothy Hodgkin* and her collaborators about this time. The complex alkaloid structure, strychnine, was solved by *Beevers* and independently by *Bijvoet* in 1950 just after the chemical structure had been fully elucidated by the work of *Robinson & Woodward*. We also solved the first caryophyllene structure about this time, with results in agreement with the chemical work of *Šorm & Barton*, but going much further on some stereochemical points. All this and much more was accomplished without computers, using the heavy atom and isomorphous substitution methods. Crystallographers led laborious lives in those days!

When computers finally came on the scene in the later 1950's and early 1960's there was a rapid and enormous expansion. This has continued at an ever increasing pace until now hardly a week passes without some new structure being determined, usually by the heavy atom method. Recently, however, and also largely due to the availability of fast computers, there has been a great development of the so-called 'direct methods' of analysis, which only use the measured intensities and mathematical relations between the structure factors. This important development was initiated by the work of *Harker & Kasper* in the 1940's and greatly extended in scope by the use of probability methods developed by *Karle & Hauptman, Sayre, Zachariasen*, and many others. Such methods may often eliminate the need for a heavy atom derivative; but the simple heavy atom method is sometimes easier to apply and more direct. The output of structures by all these methods is now so great that it is impossible to give any kind of summary, and I can only mention a few examples.

Even when a structure is fully known by chemical methods there is still enormous scope for *X*-ray work, as Professors *Prelog & Dunitz* have shown so clearly, in the beautiful stereochemical work that has been done in Zurich on large and medium ring compounds.

One early and important structure is the *citrus* bitter principle limonin, intensively studied by *Arigoni, Barton, Corey, Jeger*, and many others. Our *X*-ray work defined the stereochemistry completely and cleared up some doubtful points. This opened quite a new chapter in organic chemistry, and the structure of cedrelone, gedunin and other molecules soon became clear.

Alkaloid structures have always been a favourite with *X*-ray crystallographers because nicely crystalline hydrochlorides and bromides are often easy to make. Quite a number have been solved in Glasgow, and in 1960 the calycanthine and echitamine structures cleared up many long standing chemical problems. Professor *Schmid*'s isocalebassine structure caused us a great deal of trouble, but this was also finally solved. Many of my students have made outstanding contributions to alkaloid structures, notably *George Sim* in Glasgow, *Maria Przybylska* in Ottawa, *A. McL. Mathieson* in Melbourne, and *J. Trotter* in Vancouver.

The last and most complex structure determination that I wish to mention during this period is vitamin B_{12} by *Dorothy Hodgkin, John White*, and their many collaborators. This is a beautiful example of what the *X*-ray method, and perhaps only the

X-ray method, can do. The result of this structure determination has led to profound and continuing chemical advances.

However, all this was only the beginning of a new chapter, and molecular structures of a much higher order of complexity were soon to be solved. Within the last 10 years complete three-dimensional analyses of about 30 protein structures have been carried out, and this surely represents a supreme triumph for *X*-ray crystallography. Owing to their enormous complexity it is impossible to describe these structures briefly. We can write the formulas, *e.g.* $C_{1561}H_{2352}N_{406}O_{465}S_5Zn$, but it would take days to describe each structure at all adequately. The methods used to solve these structures have invariably been those based on the use of heavy atoms and isomorphous replacement.

The first successful protein analyses were those of *Perutz* & *Kendrew* on haemoglobin and myoglobin, and these were followed by the unravelling of the first enzyme structure, lysozyme, by *Phillips*. Many others have followed rapidly, including ribonuclease, carboxypipidase, lactate dehydrogenase, papain, and the peptide hormone insulin. The very beautiful electron density distributions that have been published for these structures resemble those obtained from the smaller organic molecules, but there is a radical difference. The protein maps do not in general reveal the individual atoms but only small groups of atoms that are not separately resolved. A certain amount of disorder is always present which limits the resolution to about 2 Å even in the most favourable cases.

The success of protein structure determinations is due to a combination of chemical and *X*-ray analyses. *Sanger* and others have shown how it is possible to determine the *sequence* of the amino acid residues that are present in the main chain. With this guidance, and if the resolution can be increased to about 2 Å in the *Fourier* maps, it is possible to see enough shape in the atomic distribution to identify the various amino acids even although the individual atoms are merged together. It is then possible to build up a complete structural model in three dimensions. When this is done for lysozyme, for example, it is found that the polypeptide chain containing 129 amino acid sub-units is cross-linked in four places by disulphide bonds (–S–S–) at the sites of the cysteine residues. As expected, many of the amino acid residues are found to be in α-helical conformation. The complete three-dimensional structure is, of course, exceedingly complex, but one important result of this study is the discovery of a cleft, or pocket, in the structure into which the substrate (a complex sugar) very probably fits. Many important structural studies involving different substrates are now in progress.

One of the most striking and complete *X*-ray studies of protein structure is that carried out on the zinc containing enzyme carboxypeptidase A by *Lipscomb* and his co-workers. It is notable as being the first enzyme structure to be determined by *X*-ray analysis before more than very small portions of the chemical sequence were known. It is also one of the more complex structures, containing 307 amino acid residues, and with a molecular weight of 34,472. This intensive analysis was based on the measurement of over 20,000 *X*-ray reflections from the native enzyme, and in addition four heavy atom derivatives had to be very completely measured for phase determination. The results of this very beautiful and detailed study have indicated

the probable binding sites of polypeptide substrates and have enabled the mechanism of the enzymatic activity to be studied in great detail.

A recent exciting development in protein structure analysis has been the application of neutron diffraction. This has been done by *Schoenborn* at Brookhaven who has obtained some astonishingly beautiful and accurate *Fourier* maps of myoglobin by this method. The crystal seems to withstand thermal neutrons even better than *X*-rays, although the large size of crystal required is certainly a difficulty. This may be overcome if more powerful neutron sources become available. The results are exciting because some of the groups in myoglobin are outlined more clearly than in the corresponding *X*-ray *Fourier*'s, a result which is probably due to the negative contribution of the hydrogen atoms. Finally, there is the exciting possibility of making an ideal isomorphous substitution for phase determination by employing isotopes of different scattering powers instead of a succession of heavy elements. The future possibilities of neutron diffraction in this field would seem to be very great.

It will be clear, I think, that we have come a very long way since the 1920's when work on organic crystal structures first began. But there are still endless new fields to explore, in chemistry, and as we move on to still more complex biological molecules.

---

17. Zur Diazotierung aromatischer Amine in Schwefelsäure

von **J. Konecny** und **H. Wenger**[1])

Chemie-Ingenieur Technik und Analytik, CIBA-GEIGY Aktiengesellschaft, Basel

(26. XI. 71)

Summary. The diazotization of 2-chloro-4,6-dinitroaniline, 2,6-dichloro-4-nitroaniline, and 4-nitroaniline in concentrated sulfuric acid is very strongly catalysed by water. At a given water concentration the reaction rates of these amines are in the ratio 50/20/1. The relation between the bimolecular rate constants k and the acidity function H_0 is very simple, the plots of $\log k$ versus H_0 being linear with a slope of 2.

Sehr schwach basische aromatische Amine werden üblicherweise in konzentrierter Schwefelsäure oder in Gemischen von Schwefelsäure und Essigsäure diazotiert. Trotz der technischen Bedeutung dieser Verfahren fehlen Angaben über die Kinetik der Reaktionen.

Besser bekannt ist der Verlauf der Diazotierung stärkerer Basen – Anilin, *p*-Toluidin, *p*-Chloranilin und *p*-Nitro-anilin – im hochaciden Bereich (66–74-proz. Schwefelsäure, 57–61-proz.[2]) Perchlorsäure) [1] [2]: die Reaktionsgeschwindigkeit nimmt mit steigendem Wassergehalt, d.h. abnehmender Acidität, rapid zu, viel stärker, als nach der Verschiebung des Gleichgewichtes (1)

$$RNH_3^+ \rightleftharpoons RNH_2 + H^+ \qquad (1)$$

zugunsten der freien Base zu erwarten wäre.

---

[1]) Der experimentelle Teil dieser Arbeit wurde von den Herren *R. Ferrat, M. Sieber & A. Schneider* ausgeführt.

[2]) Angaben über die prozentuale Zusammensetzung sind stets als Gewichtsprozente zu verstehen.