L. Zechmeister
Herausgeber · Editor

Fortschritte der Chemie organischer Naturstoffe

Progress in the Chemistry of Organic Natural Products

XXVI

FORTSCHRITTE DER CHEMIE ORGANISCHER NATURSTOFFE

PROGRESS IN THE CHEMISTRY OF ORGANIC NATURAL PRODUCTS

HERAUSGEGEBEN VON · EDITED BY

L. ZECHMEISTER

CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA

SECHSUNDZWANZIGSTER BAND
TWENTY-SIXTH VOLUME

VERFASSER · AUTHORS

K. BERNAUER · J. W. BUCHLER · R. B. COREY · D. L. DREYER
D. DÜTTING · H. GERLACH · W. HOFHEINZ · H. H. INHOFFEN
P. JÄGER · W. KELLER-SCHIERLEIN · K. LÜBKE · R. E. MARSH
E. SCHRÖDER · G. P. SCHWARTZ · A. C. TRAKATELLIS

MIT 97 ABBILDUNGEN · WITH 97 FIGURES

1968

WIEN · SPRINGER-VERLAG · NEW YORK

LIBRARY OF CONGRESS CATALOG CARD NUMBER AC 39-1015

PRINTED IN AUSTRIA

Titel Nr. 8233

Inhaltsverzeichnis
Contents

X-Ray Diffraction Studies of Crystalline Amino Acids, Peptides and Proteins

By **R. B. Corey** and **R. E. Marsh**, Pasadena, California

With 14 Figures

Contents Page

Acknowledgement. The preparation of this paper was supported in part by a Public Health Service Research Grant No. HE-02143 from the National Heart Institute of the National Institutes of Health, Public Health Service. We should like to thank Miss LILLIAN CASLER for making most of the drawings, and Miss ALLISON KIMBALL for help in preparing the manuscript.

Introduction

Three previous reviews have appeared in this Series describing the use of X-ray diffraction methods for the investigation of molecular structure: one by KRATKY and MARK (*4*) on proteins and other natural products, one by COREY (*1*) on amino acids and peptides, and one by PAULING and COREY (*11*) on the configuration of polypeptide chains. During the 14 years that have passed since the last of these reviews, great progress has been made in X-ray diffraction techniques. As a result of these improved techniques, much more accurate and reliable data are now available concerning the dimensions of small biological molecules; in addition, detailed information concerning the structure of vastly more complex molecules is now being obtained. Whereas in 1954 structural information concerning proteins had to be deduced from knowledge of the structures of, at best, simple dipeptides, today such information is being obtained directly from crystalline proteins themselves.

It is the purpose of this review to discuss the results and significance of recent X-ray diffraction investigations of, first, amino acids and peptides and, later, of crystalline proteins.

I. Outline of the X-Ray Method

The determination of the structure of a crystal is based on two fundamental equations. The first is Bragg's Law,

$$\lambda = 2\, \mathrm{d}(h\,k\,l) \sin \Theta\,(h\,k\,l), \qquad (1)$$

which defines the conditions for which a diffraction maximum may be observed when a crystal is irradiated. Here, λ is the wave length of the radiation, Θ is the angle of incidence (and reflection) between the X-rays and the diffracting set of crystallographic planes, h, k, and l are the Miller indices defining that set of planes, and d is the spacing between successive, parallel planes. By measuring values of Θ for a number of sets of planes, using a radiation of known wave length, it is immediately possible to determine the size and shape of the crystallographic unit cell—the unit of structure which, when repeated translationally in three dimensions, generates the crystal. Typically, unit cells of crystals of moderately complex compounds have dimensions in the range 10 to 20 Å; for crystalline proteins these dimensions may be as large as 100 Å. The wave lengths of the X-rays used to examine the crystals are about 1 to 2 Å.

References, pp. 40—47

The determination of the structure of a crystalline compound involves knowing not only the size and shape of the unit cell but also the locations of the various atoms within the cell. This knowledge can be obtained from measurements of the intensities I ($h\,k\,l$) of the X-ray beam after it has been diffracted by a set of crystal planes $h\,k\,l$. The intensity of the diffracted beam is proportional to the square of the magnitude of the quantity $F_{h\,k\,l}$—the so-called "structure factor". The structure factor is the basis of the second fundamental equation:

$$F_{h\,k\,l} = \sum_j f_j\, T_j \exp.\ 2\,\pi\, i\,(h\,x_j + k\,y_j + l\,z_j). \tag{2}$$

Here, $x_j\ y_j\ z_j$ are the fractional coordinates, relative to the unit cell axes a, b and c, defining the position of the jth atom of the structure, f_j is the scattering power of that atom (approximately proportional to its atomic number), and T_j is a term describing its vibrational motion within the crystal lattice; the summation is over all atoms in the unit cell.

Thus, if one knows the positions, identities and patterns of motion of all atoms within the unit cell, one can calculate the structure factors and, from them, the intensities of all diffraction maxima. The inverse is not true; for in order to determine the positions of the atoms from a knowledge of the intensities, one must know not only the magnitudes but also the phase angles of the structure factors. Determination of these phase angles—the "phase problem"—is therefore the crucial factor in a successful X-ray diffraction investigation.

Various methods are used to solve the phase problem. One of the most powerful tools is the Patterson function

$$P_{u\,v\,w} = \sum_{h,\,k,\,l\,=\,-\infty}^{\infty} \sum^{\infty} \sum^{\infty} F_{h\,k\,l}^2 \cos 2\,\pi\,(h\,u + k\,v + l\,w). \tag{3}$$

This function involves a triple Fourier summation of the observed values of F^2 (which are readily derived from intensity measurements) for all crystal planes. The result of this summation is a three-dimensional map in which maximum values of $P_{u\,v\,w}$ occur at positions—defined by the coordinates u, v, w—corresponding to vectors between pairs of atoms in the structure. Since a structure with n atoms will generate n^2 such vectors, Patterson maps are difficult to interpret except for small compounds or for molecules containing one or two very heavy atoms, in which case vectors involving these atoms will be prominent. Nevertheless, a majority of successful crystal-structure determinations have been based on the interpretation of Patterson maps.

Other methods used for solving the phase problem include: (i) attempting to guess the structure from the size and symmetry of the unit cell, making use of known features of molecular geometry, packing and

hydrogen bonding; (ii) preparing isomorphous, heavy-atom derivatives, for which the heavy atom can be readily located from Patterson maps and then used, in conjunction with the differences in intensities between the two derivatives, to assign phase angles; (iii) using statistical methods of phasing, in which the trigonometric nature of the structure factor is made the basis of probability relationships between the phase angles of related reflections [a recent application of this method is given by KARLE

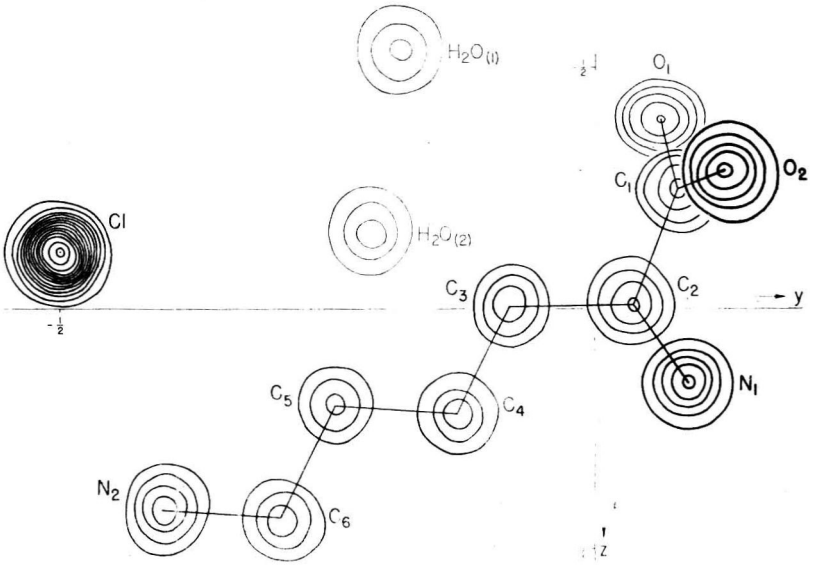


Fig. 1. A representation of the electron density in crystals of L-lysine hydrochloride dihydrate (*38*), as viewed down the *a* axis of the unit cell. Contours are drawn at equal intervals of electron density. The electron density associated with the hydrogen atoms is too small to be apparent in maps of this type. The complete structure of the crystal is shown in Figure 3 [Acta Crystallogr. 15, 54 (1962)].

and KARLE (*22*, *45*)]. In general, method (i) is applicable only to very simple molecules; method (iii) has been used on relatively complex structures—up to about 50 atoms; while method (ii) can be used on very complex structures and is the method usually employed in protein crystallography.

Once an approximate structure, or set of phase angles, has been derived, an improved picture of the structure can be obtained by calculating the three-dimensional Fourier series

$$\varrho\,(x\,y\,z) = \sum_{h,k,l=-\infty}^{\infty}\sum\sum F\,(h\,k\,l)\,\exp.\,-\,2\,\pi\,i\,(h\,x + k\,y + l\,z). \qquad (4)$$

A plot of this function depicts the value of the electron density ϱ at each point $(x\,y\,z)$ in the unit cell. Regions of high electron density cor-

respond to the positions of atoms, the electron density being approximately proportional to the atomic number of the atom. Typical electron density maps are shown in *Figures 1 and 2*.

The final step in a crystal-structure investigation is the refinement of the positional coordinates x, y, z and vibrational parameters T of all the atoms in the unit cell, resulting in a set of calculated structure factors

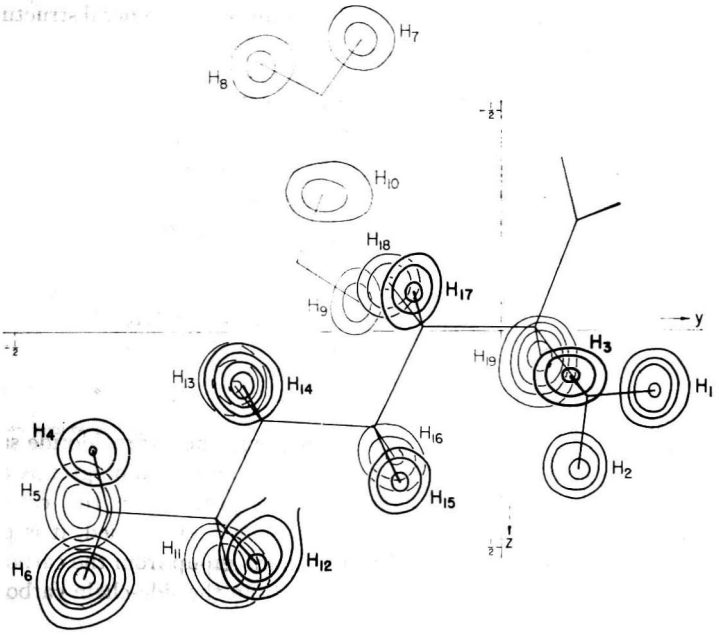


Fig. 2. An electron density map of L-lysine hydrochloride, showing the locations of the hydrogen atoms. The terms entering into this Fourier summation (Equation 4, p. 4) were values of $F(h\,k\,l)$ from which the calculated contributions of the C, N, O and Cl atoms had been subtracted [Acta Crystallogr. 15, 54 (1962)]

$F(h\,k\,l)$ in optimum agreement in magnitude with the observed values. This refinement is now carried out almost exclusively by the method of least-squares, and makes full use of the speed and storage capabilities of modern digital computers. The intensities $I(h\,k\,l)$ calculated from these coordinates should be in satisfactory agreement with the observed intensities. The accuracy of the resulting set of atomic parameters is often such that interatomic distances are accurate to less than 0.01 Å and angles to about 0.1°.

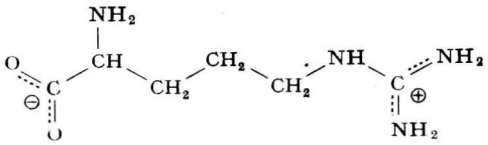
II. Crystal Structures of the Amino Acids

The crystal structures of nearly all of the amino acids—or, in some cases, of one or more of their hydrohalide salts—have by now been deter-

mined (*12–39*). The accuracy of these determinations varies widely, the standard deviations in the interatomic distances ranging from perhaps 0.1 Å for early investigations based on partial, two-dimensional intensity data to less than 0.005 Å in a few recent instances where exceptionally careful three-dimensional refinements were carried out. In all cases, however, the accuracy has been sufficient to permit the fundamental structural and configurational features to be seen. We shall not discuss each of these structures in detail, but rather point out the general structural features which seem to be common to all.

1. Zwitterion Structures

All amino acids that have been studied (as well as all simple peptides) crystallize as zwitterions. In all cases except arginine, the proton from the carboxyl group is transferred to the amino group on the alpha carbon atom; in arginine, it is transferred to the guanidinium group. In the cases



of the basic amino acids lysine and histidine, only the hydrochloride salts have been investigated; in these cases both the alpha amino groups and the side-chain nitrogen atoms have an extra proton. In the case of glutamic acid, the only free dibasic acid to have been studied, it is probable that the proton is transferred to the amino group from the carboxyl group on the alpha carbon atom rather than from the side-chain carboxyl group.

2. Hydrogen Bonding

Intermolecular hydrogen bonding is an important feature of the crystal structures of all the amino acids. The ammonium group ($-NH_3^+$) of the zwitterion is an excellent hydrogen-bond donor and the carboxylate ion ($-CO_2^-$) is an excellent acceptor; as a result, strong hydrogen bonds are invariably formed involving these two groups. The length of the hydrogen bond, the N ... O distance, is usually in the range 2.8 to 2.9 Å, but occasionally is as long as 3.1 Å.

Typical hydrogen-bond arrangements, as found in crystals of L-lysine hydrochloride, glycine, and of L- and DL-alanine, are shown in *Figures 3–6*. The relationship between the structures of the two forms of alanine is particularly interesting, and points up the importance of hydrogen bonding in determining the arrangement of molecules in crystals. Crystals of DL-alanine contain planes of symmetry which relate molecules of

References, pp. 40—47

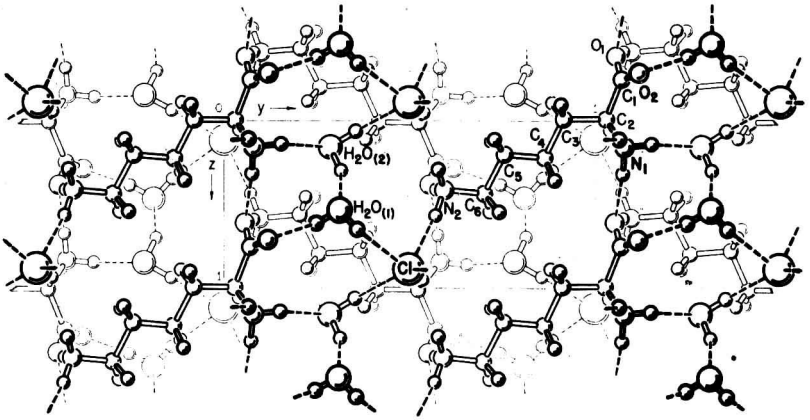


Fig. 3. The crystal structure of L-lysine hydrochloride dihydrate, as determined by X-ray diffraction procedures (*38*). The view is along the *a* axis. All ten acidic protons—three from each of the ammonium groups and two from each of the water molecules—form hydrogen bonds, indicated by dashed lines; the chloride ion and the oxygen atoms of the carboxylate groups and water molecules serve as acceptors. The molecule shown in Figures 1 and 2 is at the extreme left [Acta Crystallogr. 15, 54 (1962)]

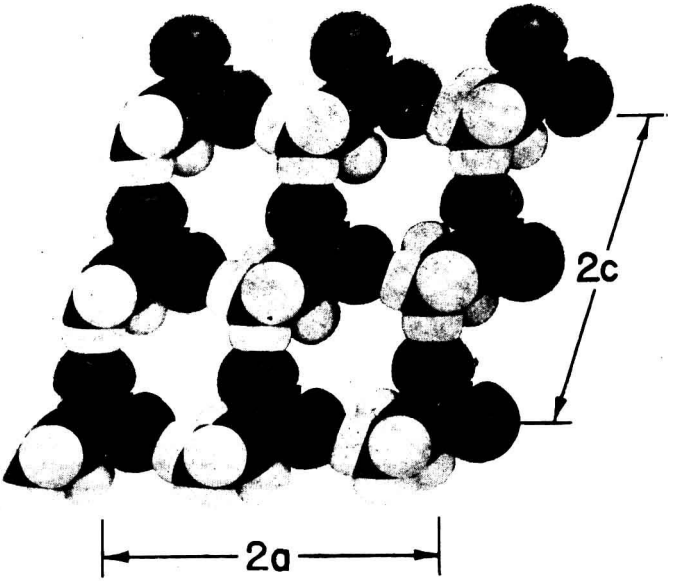


Fig. 4. A photograph of space-filling atomic models representing a portion of the crystal structure of glycine (*12*, *24*). Only two of the three hydrogen bonds are shown; the remaining hydrogen atoms (white) of the ammonium groups are bonded to oxygen atoms in a layer above