# INTERATOMIC DISTANCES IN PROTEINS AND RELATED SUBSTANCES[1,2]

ROBERT B. COREY

*Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California*

*Received October 3, 1939*

In the study of the structure of proteins the ultimate goal is a complete chemical and physical picture of the molecule,—of the nature and number of the atoms which compose it and the details of the manner of their combination. Its attainment, even in part, might be expected to bring these compounds, so fundamentally associated with life processes, under a control similar to that which the chemist now exercises over less complicated organic molecules. The great strides which have been made in this direction during recent years have resulted largely from the use of both chemical and physical techniques in the attack upon the problems of protein structure.

Chemical studies have led to the general acceptance of the idea, advanced many years ago (25), that protein molecules are made up of amino acids held together in long peptide chains by linkages between their carbon and nitrogen atoms. More than twenty amino acids have been identified as integral parts of proteins, and the development of special analytical techniques has permitted their quantitative estimation with ever-increasing accuracy. Recent results have suggested the possibility of definite periodicities in the arrangement of amino acid residues along the peptide chain (11, 12), a hypothesis which should be most fruitful in stimulating critical evaluation of existing data and in emphasizing the need of more precise analytical methods. Certainly much work remains to be done before the chemical composition of any single protein can be considered as established (34).

Preparation of increasing numbers of both plant and animal proteins in the crystalline form has encouraged the attempt to seek information concerning their structure through x-ray diffraction studies, a method which has proved so successful in determining atomic positions in crystals

[1] Contribution No. 729 from the Gates and Crellin Laboratories.

[2] Presented at the Symposium on X-ray Studies of Substances of High Molecular Weight, held at the Ninety-eighth Meeting of the American Chemical Society in Boston, Massachusetts, September, 1939, under the auspices of the Divisions of Physical and Inorganic Chemistry, Colloid Chemistry, Organic Chemistry, Paint and Varnish Chemistry, and Rubber Chemistry of the American Chemical Society.

of simpler substances. In the case of some crystalline proteins (5, 14, 20, 21) it has been possible to establish the nature and dimensions of the crystal lattice and to draw some conclusions regarding the size and shape of the molecule. However, in spite of the advances which have been made in x-ray technique, the great size and complexity of protein molecules seem to preclude the possibility of arriving at positions and relationships of individual atoms directly by these means.

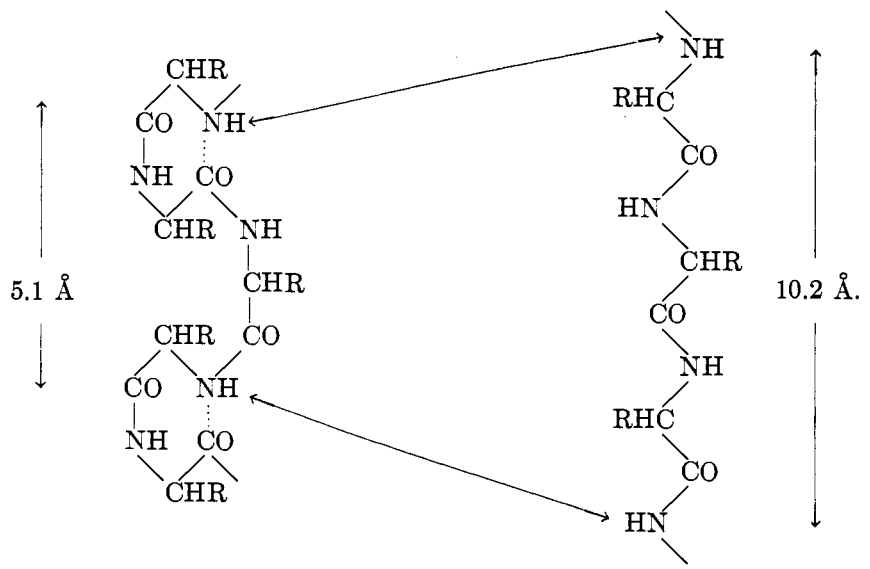
When applied to fibrous proteins, on the other hand, x-ray methods have met with considerable success in throwing light upon the intramolecular arrangements of their atoms. Silk fibroin, for example, yields excellent fiber photographs having well-defined spots and bearing marked resemblance to those obtained from comparatively simple crystalline substances. Many attempts have been made to subject these patterns to systematic analysis leading to the establishment of a fundamental unit of structure (15, 29, 30, 31, 32, 40), but it is not possible to reconcile any of the units proposed with the known facts concerning the chemical constitution of the fibroin molecule (13). All of these studies, however, indicate that the protein of silk exists as fully extended polypeptide chains arrayed nearly parallel to the axis of the fiber. Thus the photographs all reveal a definite periodicity of 7.0 $\pm$ 0.2 Å. along the fiber axis. From early crystal structure data it could be safely assumed that the distance between adjacent, chemically bonded carbon atoms is 1.54 Å. and that the corresponding separation between carbon and nitrogen is 1.3 Å. to 1.4 Å. with approximately a tetrahedral angle (109° 28′) between all bonds. In a fully extended polypeptide chain having these dimensions one amino acid residue will occupy a distance of 3.38 Å. to 3.54 Å. along the axis of the chain. These data suggest that the repetition distance, 7.0 Å. (= 2 $\times$ 3.5 Å.), found along the fiber axis of silk fibroin corresponds to two amino acid residues. Equatorial reflections, the most prominent of which represent spacings of 4.3 Å. and 4.6 A., are to be associated in some way with the packing together of adjacent chains in the fiber.

X-ray studies of other fibrous proteins lend additional significance to this picture of the fully extended polypeptide chain. In their extensive investigations of keratin, the fundamental protein constituent of animal hairs, horns, quills, etc., Astbury and his coworkers have found a characteristic spacing of 3.38 A. along the axis of the fiber when the latter is stretched to its maximum elongation (2). A corresponding periodicity of 3.3 Å. has been found in samples of stretched feather keratin (7), and a spacing along the fiber axis of 6.7 Å. (= 2 $\times$ 3.35 Å.) is reported (28) in spun filaments of blood fibroin.

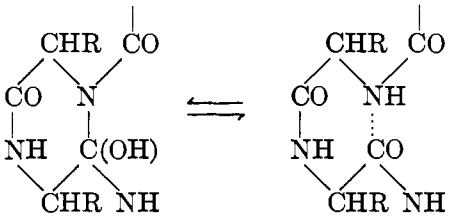
Striking uniformity is also observed in the spacings perpendicular to the fiber axis which are to be ascribed to regularity in the packing of the protein chains with their immediate neighbors. In the case of $\beta$-keratin (stretched

hair) Astbury (2) observed maxima corresponding to 4.65 Å. and 9.8 Å., the former of which he attributes to the distance between polypeptide chains held in close contact by the interaction of their =NH and =CO groups. The latter distance has been found to be practically perpendicular to the former (8) and probably lies in the direction in which the main chains are unable to approach each other more closely because of their respective side chains. Approximately these same equatorial spacings have been found in fibrin (28), and their significance is further increased by the fact that similar distances are characteristic of all denatured proteins that have been studied by x-rays (6).

Brief mention should also be made of the changes in the x-ray photographs of keratin which are observed when the specimen is stretched. Photographs of unstretched hair indicate a spacing along the fiber axis of 5.06 Å. (10). Under proper conditions hair may be stretched 100 per cent without rupture, and in this fully extended state the 5.06 Å. spacing has vanished, to be replaced by the characteristic 3.38 Å. distance associated with nearly completely extended polypeptide chains. The 9.8 Å. "side-chain" spacing perpendicular to the fiber axis is found unchanged in photographs of hair before and after stretching. The shorter (4.65 Å.) equatorial spacing, however, is found only in stretched specimens. These facts suggest that in unstretched keratin the protein is longitudinally folded in some fashion which allows of reversible extension and contraction. A mechanism for this folding was first suggested by Astbury (9) in 1930, and has been discussed by him in many subsequent publications, in which the transformation from $\alpha$- to $\beta$-keratin was represented by the following scheme:



In a review published in 1935 Astbury (3) stated that "when this scheme was first put forward the precise nature of the linkage (shown dotted above) between CO and NH groups in the hexagonal folds of the $\alpha$-form was left open; but we know now from accumulated x-ray crystallographic data that hexagons of the dimensions required cannot be built unless this linkage is of a covalent type that would bring the carbon and nitrogen atoms to within some $1\frac{1}{2}$ Å. of each other. A way out of the difficulty has been suggested by F. C. Frank, by postulating a sort of lactam–lactim interchange between the CO and NH groups, thus:"




More recently experimental evidence from many sources (23, 24, 26, 27, 33, 39) has shown that the existence in protein molecules of closed rings of the sort postulated is exceedingly unlikely, so that the nature of the

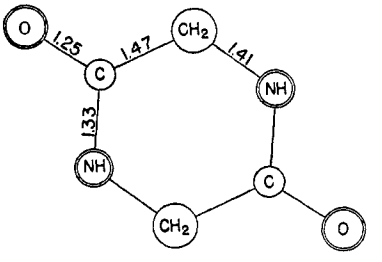


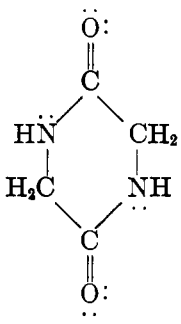

FIG. 1. A diagram of the molecule of diketopiperazine. Figures give interatomic distances in Ångström units. Angles between all bonds are close to 120°.

folding of polypeptide chains in fibrous and globular proteins is still completely undetermined. This situation has been thoroughly appreciated by Astbury who, in a paper (4) published only last year, called attention to the fact that "whether the kind of hexagonal fold postulated for the fibrous proteins provides the theme of the globular proteins too, or whether, even, it is really correct for either, it is also not yet possible to decide."

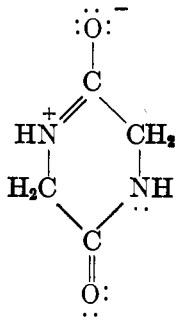
X-ray studies of fibrous proteins have been of great assistance in constructing a picture of the probable arrangement of the molecules in these substances in accord with their physical and chemical properties, and in some cases the periodicities observed are to be regarded as repetition dis-

tances within the molecules themselves. Although they thus constitute a body of experimental facts which a complete picture of the protein molecule must fully explain, they fail entirely to give any direct information regarding the distances between discrete atoms in the polypeptide chain or, indeed, in any part of the protein molecule. At present the only means of obtaining such information appears to be through x-ray investigations of crystallized products of protein hydrolysis. For this reason a series of investigations upon the crystal structures of the amino acids and polypeptides is being prosecuted in these Laboratories as a portion of a program of research upon the structure of proteins in general.

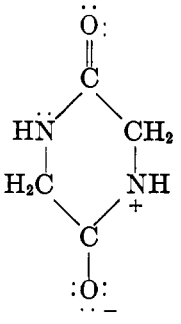
The first substance to be thus analyzed was the cyclic dipeptide diketopiperazine, or "glycine anhydride" (19). The general shape of the molecule and the interatomic distances found within it are shown diagrammatically in figure 1. The molecule is a nearly plane hexagon with the angles between all bonds 120° ± 3°. It may be expected to resonate among the structures



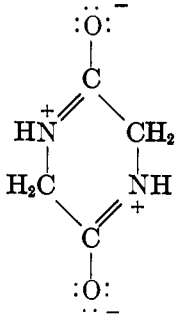
I



II



III



IV

so that the bond distances C—O and OC—N should have the values characteristic of resonance of this type. The distances found, C—O = 1.25 Å. and C—N = 1.33 Å., are in good agreement with those to be anticipated (37). Thus in urea (42) the interatomic distances within the molecule are C—O = 1.25 Å. and C—$NH_2$ = 1.37 Å. In thiourea (41) C—$NH_2$ = 1.35 Å. On the other hand, the distance N—$CH_2$, 1.41 Å., is surprisingly short, for there appears to be no reason for so great a departure from the normal single-bond distance, 1.47 Å., found in the compounds $CH_3NO_2$ (17), $CH_3N_3$ (36), $CH_3NC$ (16), and $N(CH_3)_3$ (18). The C—C distance, 1.47 Å., is likewise much shorter than the normal value, 1.54 Å.

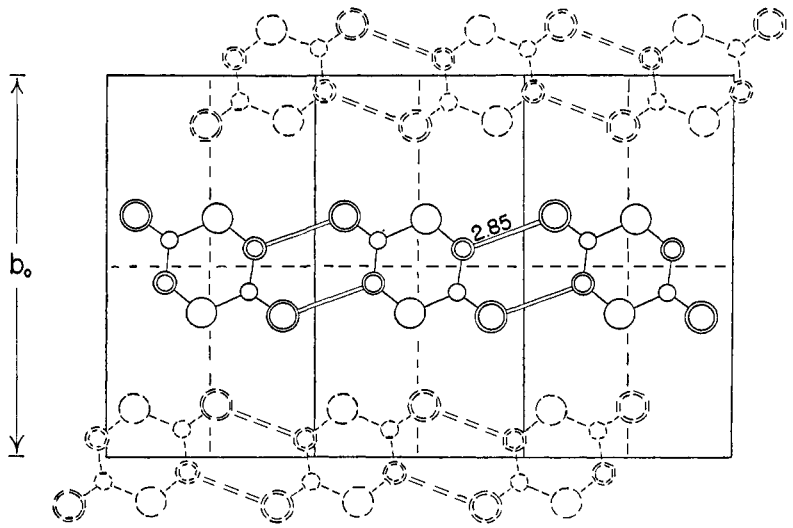



FIG. 2. A view perpendicular to the (101) plane of the crystal of diketopiperazine, showing chains of molecules held together by hydrogen bonds (double lines) 2.85 Å. in length between oxygen atoms and =NH groups.

It is hoped that work upon some substituted diketopiperazines which is now in progress in these Laboratories may throw some light upon the causes of these anomalies.

In crystals of diketopiperazine the molecules are linked together by hydrogen bonds between their respective oxygen atoms and =NH groups to form flat continuous chains throughout the structure. The positions of these chains are indicated in figure 2, which is a view of the crystal of diketopiperazine perpendicular to the plane (101). The distance between oxygen and nitrogen atoms connected by hydrogen bonds is 2.85 Å., which is in satisfactory agreement with corresponding separations in $(NH_4)H_2PO_2$, 2.81 Å. (43), and in $(NH_4)_2C_2O_4 \cdot H_2O$, 2.76 to 2.88 Å. (22).

The determination of the crystal structure of glycine (1), the simplest of the amino acids, affords additional data concerning interatomic distances within these least complicated products of protein hydrolysis. Distances and bond angles found in the glycine molecule are indicated

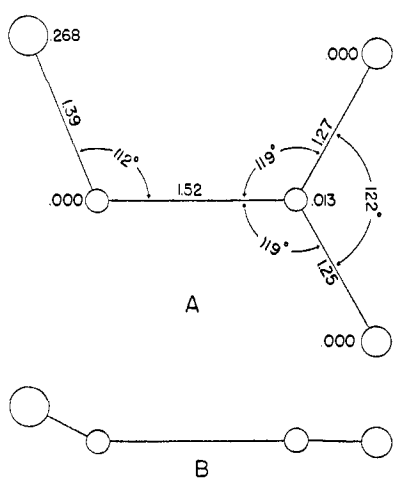



FIG. 3

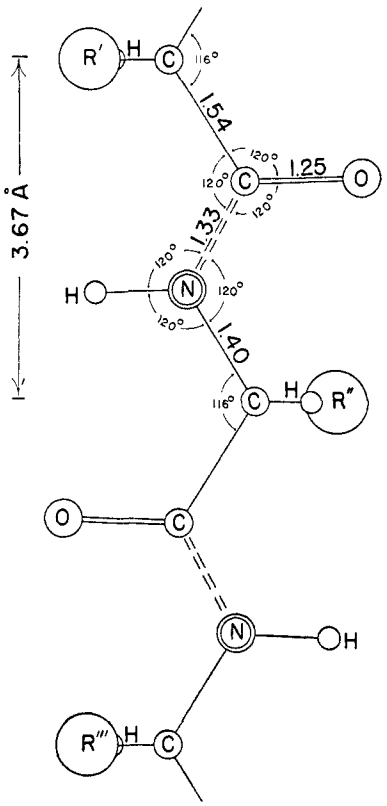



FIG. 4

FIG. 3. The molecule of glycine viewed (A) perpendicular to a plane containing the two oxygen and the $\alpha$-carbon atoms, and (B) parallel to this plane.

FIG. 4. A diagrammatic representation of a fully extended polypeptide chain, based upon the interatomic distances and bond angles found in crystals of diketopiperazine and glycine. The double dashed lines represent resonating bonds (1.33 Å.) between the keto-carbon and nitrogen atoms.

in figure 3. The only value which departs from normal is the C—N separation, 1.39 Å., which, however, is in agreement with the analogous distance found in diketopiperazine. Interatomic distances found within the molecules of these two compounds are summarized in table 1.

Even from these limited data it should be possible to make some definite

statements regarding the linkages present in the polypeptide chain in proteins. The distance between the nitrogen atom and the $\alpha$-carbon atom is 1.40 Å., with a probable error of 0.02 Å., in both diketopiperazine and glycine. Although this distance is about 0.07 Å. shorter than that to be expected from the normal covalent radii of the atoms (35, 38), there seems to be little doubt that this new distance found in these hydrolytic products of proteins exists also in the protein molecule itself. Whether the abnormally short C—C distance found in the cyclic dipeptide prevails also in the long peptide chains of proteins is a question which can be settled only by complete crystal structure studies of simple straight-chain polypeptides and substituted diketopiperazines. There seems to be every reason for believing that the type of resonance evidenced by diketopiperazine, urea, etc., will be present in the polypeptide chain so that the distance between the nitrogen atom and its adjacent keto-carbon atom must be very near

TABLE 1

*Corresponding interatomic distances found in molecules of diketopiperazine and glycine*

| | DIKETOPIPERAZINE | GLYCINE |
|---|---|---|
| | Å. | Å. |
| C—O | 1.25 | 1.26 |
| $\alpha$-C—N | 1.41 | 1.39 |
| C—C | 1.47 | 1.52 |
| OC—N | 1.33 | |

to 1.33 Å. For the same reason the oxygen atom must be about 1.25 Å. from the keto-carbon atom.

A diagram of a fully extended polypeptide chain in which these dimensions have been incorporated is shown in figure 4. The bond angles about the carbonyl carbon and the nitrogen atoms are assumed to be 120°, as found in diketopiperazine. Although the true angles between these bonds doubtless differ somewhat from this value, it is unlikely that any of them departs from it by more than 5°. Around the $\alpha$-carbon atom the bonds are assumed to be arranged in nearly tetrahedral fashion, except that the C—C—N bond angle is taken as 116°, a mean of the values found in glycine (112°) and diketopiperazine (120°). The fully extended chain is coplanar, with the carbonyl oxygen atoms and amide hydrogen atoms included in the plane. The distance along the chain corresponding to one amino acid residue is 3.67 Å., which is somewhat greater than that estimated from x-ray photographs of any fibrous protein. This fact suggests that in these substances the chain is never fully extended in a truly coplanar configuration but that interactions, steric or otherwise, with its immediate neighbors cause slight distortions, probably involving rotations about the C—C

bond. This view finds some confirmation if the attempt is made to place these fully extended chains side by side at about 4.65 Å. from each other and in such a fashion as to form hydrogen bonds between the =CO and =NH groups of adjacent chains. It is then found that all such arrangements result in steric interferences between side chains, or between side chains and oxygen atoms, as long as complete coplanarity of the polypeptide chain is rigidly maintained. However, for certain relative positions it is possible to avoid these conflicts by slight rotations back and forth about the carbon–carbon bond. Unfortunately, existing x-ray data are inadequate as reliable guides in making a selection among the many possible configurations which present themselves.

Thus, although our present information is insufficient to supply a definite solution of even the simpler problems of the atomic arrangements in proteins, there seems to be ample justification for the belief that more precise knowledge of the interatomic distances within the polypeptide chain, together with the results of more refined x-ray and chemical techniques, will provide experimental foundation for the development of a critical theory of protein structure.

It is a pleasure to acknowledge the kindness of Professor Linus Pauling, whose helpful discussions have been a great aid in the preparation of this paper.

## REFERENCES

(1) ALBRECHT, G., AND COREY, R. B.: J. Am. Chem. Soc. **61**, 1087 (1939).
(2) ASTBURY, W. T.: Trans. Faraday Soc. **29**, 193 (1933).
(3) ASTBURY, W. T.: J. Textile Inst. **27**, 282 (1936).
(4) ASTBURY, W. T.: Trans. Faraday Soc. **34**, 378 (1938).
(5) ASTBURY, W. T., AND LOMAX, R.: Nature **133**, 795 (1934).
(6) ASTBURY, W. T., AND LOMAX, R.: J. Chem. Soc. **1935**, 846.
(7) ASTBURY, W. T., AND MARWICK, T.: Nature **130**, 309 (1932).
(8) ASTBURY, W. T., AND SISSON, W. A.: Proc. Roy. Soc. (London) **A150**, 533 (1935).
(9) ASTBURY, W. T., AND WOODS, H. J.: Nature **126**, 913 (1930).
(10) ASTBURY, W. T., AND WOODS, H. J.: Phil. Trans. **A232**, 333 (1933).
(11) BERGMANN, M.: Harvey Lectures **31**, 37 (1936); Chem. Rev. **22**, 423 (1938).
(12) BERGMANN, M., AND NIEMANN, C.: J. Biol. Chem. **115**, 77 (1936); **118**, 301 (1937); Science **86**, 187 (1937).
(13) BERGMANN, M., AND NIEMANN, C.: J. Biol. Chem. **122**, 577 (1938).
(14) BERNAL, J. D., AND CROWFOOT, D.: Nature **133**, 794 (1934).
(15) BRILL, R.: Ann. **434**, 204 (1923).
(16) BROCKWAY, L. O.: J. Am. Chem. Soc. **58**, 2516 (1936).
(17) BROCKWAY, L. O., BEACH, J. Y., AND PAULING, L.: J. Am. Chem. Soc. **57**, 2693 (1935).
(18) BROCKWAY, L. O., AND JENKINS, H. O.: J. Am. Chem. Soc. **58**, 2036 (1936).
(19) COREY, R. B.: J. Am. Chem. Soc. **60**, 1598 (1938).
(20) CROWFOOT, D.: Nature **136**, 591 (1935).

(21) CROWFOOT, D.: Proc. Roy. Soc. (London) **A164**, 580 (1938).
(22) HENDRICKS, S. B., AND JEFFERSON, M. E.: J. Chem. Phys. **4**, 102 (1936).
(23) HAUROWITZ, T.: Z. physiol. Chem. **256**, 28 (1938).
(24) HAUROWITZ, T., AND ASTRUP, T.: Nature **143**, 118 (1939).
(25) HOFMEISTER, F.: Ergeb. Physiol. biol. Chem. exptl. Pharmakol. **1**, 759 (1902).
(26) HUGGINS, M. L.: J. Am. Chem. Soc. **61**, 755 (1939).
(27) JENKINS, G. I., AND TAYLOR, T. W. J.: J. Chem. Soc. **1937**, 495.
(28) KATZ, J. R., AND DE ROOY, A.: Rec. trav. chim. **52**, 742 (1933).
(29) KRATKY, O.: Z. physik. Chem. **B5**, 297 (1929).
(30) KRATKY, O., AND KURIYAMA, S.: Z. physik. Chem. **B11**, 363 (1931).
(31) MEYER, K. H., AND MARK, H.: *Der Aufbau der hochpolymeren organischen Naturstoffe*, p. 222. Akademische Verlagsgesellschaft, Leipzig (1930).
(32) MEYER, K. H., AND MARK, H.: Ber. **61**, 1932 (1928).
(33) NEURATH, H., AND BULL, H. D.: Chem. Rev. **23**, 427 (1938).
(34) NIEMANN, C.: Cold Spring Harbor Symposia Quant. Biol. **6**, 58 (1938).
(35) PAULING, L.: *The Nature of the Chemical Bond*, p. 154. Cornell University Press, Ithaca, New York (1939).
(36) PAULING, L., AND BROCKWAY, L. O.: J. Am. Chem. Soc. **59**, 13 (1937).
(37) PAULING, L., AND BROCKWAY, L. O.: J. Am. Chem. Soc. **59**, 1223 (1937).
(38) PAULING, L., AND HUGGINS, M. L.: Z. Krist. **87**, 205 (1934).
(39) PAULING, L., AND NIEMANN, C.: J. Am. Chem. Soc. **61**, 1860 (1939).
(40) TROGUS, C., AND HESS, K.: Biochem. Z. **260**, 376 (1933).
(41) WYCKOFF, R. W. G., AND COREY, R. B.: Z. Krist. **81**, 386 (1932).
(42) WYCKOFF, R. W. G., AND COREY, R. B.: Z. Krist. **89**, 462 (1934).
(43) ZACHARIASEN, W. H., AND MOONEY, R. C. L.: J. Chem. Phys. **2**, 34 (1934).