Mr. President, Ladies and Gentlemen:

It is a great honor for me to have been invited to speak at the opening session of the Thirteenth International Congress of Pure and Applied Chemistry, and I express my thanks to the officers of the Congress and of the International Union of Pure and Applied Chemistry for having extended the invitation to me and to you for your courteous welcome.

My subject today is the stochastic method and the structure of proteins. Many scientists have been interested in the question of the way in which scientific discoveries are made. A popular idea is that scientists apply their powerful intellects in the straightforward, logical induction of new general principles from known facts, and deduction of previously unrecognized conclusions from known principles. This method is, of course, sometimes used; but the advances in knowledge that are made by it are less significant than those that
result from mental processes of another sort - in large part subconscious processes. Henri Poincaré, in his essay on mathematical creation, said that knowledge of mathematics and of the rules of logic is not enough to make a man a creative mathematician; he must also be gifted with an intuition that permits him to select from among the infinite number of combinations with mathematical entities already known, most of them absolutely without interest, those combinations which will lead to useful and interesting results. He illustrated the role of the subconscious by describing his investigations of the Fuchsian functions, which he had discovered while working at Caen. He left Caen on a geologic excursion, and for sometime, while traveling, made no conscious effort to attack the problem; then one day, as he put his foot on the step of an omnibus, the idea suddenly came to him that the transformations that he had used to define the Fuchsian functions were identical with those of non-Euclidian geometry. He verified this conclusion after on his return to Caen. Then, after a period of non-productive effort, he spent a few days at the seaside, and one morning, while walking on the
The field of the determination of the structure of crystals by the x-ray diffraction method provides interesting illustrations of the ways that scientific progress is achieved. Work in this field consists in the solution of individual, largely unrelated problems - the determination of the atomic arrangement of individual crystals. If the atomic arrangement (the structure) is sufficiently simple from the observed x-ray diffraction pattern, it can be determined by straightforward, completely logical arguments.

The procedure developed by Fldshikawa, Wyckoff, and Dickinson before 1920 consisted in the tabulation, with use of the theory of space groups, of all atomic arrangements compatible with the symmetry and size of unit required to account for the x-ray pattern, and the rigorous elimination with use of the observed x-ray intensities (especially of qualitative inequalities in intensity of pairs of diffraction maxima) of all of these atomic arrangements except one, which was then accepted as the structure of the crystal. During the dozen years after the determination of the first crystal structures by W. L. Bragg and W. H. Bragg most structure
It is well known that the electron distribution in a crystal can be expressed as a three-dimensional Fourier series in which the coefficients of the various Fourier terms are proportional to the square roots of the intensities of the corresponding diffraction maxima. The electron-distribution function depends, however, on the phases of the Fourier terms, and there is no method of determining these phases. For a crystal of even moderate complexity, such as an amino acid, the number of possible atomic arrangements provided by the theory of space groups is so great that the exhaustive consideration of them and elimination of all but one by use of the observed x-ray intensities cannot be carried out even with the aid of electronic calculating machines, and the attack on the problems presented by these crystals must be made in other ways.

During the first few years of my activity as a research man I carried through a number of structure determinations by the rigorous method, in collaboration with my teacher Roscoe G. Dickinson, and independently. I was deeply interested, however, in more complex substances
than those that could be attacked in this way, and from 1928 on I applied a different method of attack to many substances.

One evening in 1933, after I had described the new method of crystal structure determination to him at New York, and contrasted it with the rigorous method, Dr. Karl Darrow suggested that I call it the stochastic method, and referred me to the introduction of this expression by the chemist Alexander Smith, who had written, in his "Inorganic Chemistry", 1909, page 142, the following: "... When Mitcherlich discovered that Glauber's salt gave a definite pressure of water vapor, he at once formed the hypothesis, that is, supposition, that other hydrates would be found to do likewise. Experiments showed this supposition to be correct. The hypothesis was at once displaced by the fact. This sort of hypothesis predicts the probable existence of certain facts or connections of facts, hence, reviving a disused word, we call it a stochastic hypothesis (Greek stochastikos, apt to divine the truth by conjecture). It differs from the other kind in that it professes to be composed entirely of verifiable facts and is subjected to verification as quickly as possible. In the stochastic method of treating very complex guessed crystals a plausible structure is with the aid of hints provided
The stochastic method is, of course, used all the time by the organic chemist. He synthesizes a new substance and assigns it a structure — a stochastic structure — on the basis of his knowledge of chemical principles. Then he checks a few points — ultimate analysis, molecular weight, presence of a methoxy group or carboxyl group — and is satisfied that his guess is correct if the results of the tests are those that he has predicted.

In the
by the observed size of unit and space-group symmetry, and knowledge of
genral principles of molecular structure, and the stochastic hypothesis
that this is the actual structure of the crystal thereupon is either veri-
fied or disproved by the comparison of calculated and observed x-ray
intensities.

As a rule, if the agreement between observed calculated intensities is excellent
the proposed structure may be accepted as the correct one. There is, however,
always some possibility that the agreement is fortuitous. This was em-
phasized by a discovery that Dr. M. D. Shappell and I made in 1930, 1
during our study of the structure of bixbyite (Mn, Fe)₂O₃, and the C-
modification of the rare-earth sesquioxides. W. H. Zachariasen had in-
vestigated these crystals, and had assigned positions to the 32 metal atoms
in the unit cube. 2 During our reinvestigation of the structure we

2 W. H. Zachariasen, Z. Krist., 67, 455 (1928)

1 L. Pauling and M. D. Shappell, Z. Krist., 75, 128 (1930).
discovered that the space group $T_b^7$ provides certain pairs of physically distinct arrangement which give exactly the same intensities of reflection of x-rays from all crystallographic planes, so that no unambiguous structure determination can be made by the consideration of x-ray intensities alone. This ambiguity, which has not turned out to be an important one in practice, has been further investigated by Patterson.

I shall content myself with a few examples of the application of the stochastic method.

The mineral enargite, Cu$_3$AsS$_4$, was found on x-ray investigation to have an orthorhombic unit of structure with $a = 6.46 \AA$, $b = 7.43 \AA$, and $c = 6.18 \AA$. This unit contains six copper atoms, two arsenic atoms, and eight sulfur atoms. The dimensions are closely similar to those of the hexagonal mineral wurtzite, ZnS: the dimensions of a double orthohexagonal unit of wurtzite, containing eight zinc atoms and eight sulfur atoms,
are $\sqrt{3}a = 6.65\ \text{Å}, \ 2a = 7.68\ \text{Å}, \ \text{and} \ c = 6.28\ \text{Å}$. This suggests that the atomic arrangement is that shown in Figure 1, which results from replacing one fourth of the zinc atoms in wurtzite by arsenic and the remaining three fourths by copper atoms, in such a way as to give discrete $\text{AsS}_4$ x-ray groups. It was found that the calculated intensities for this ideal structure well agree with the observed ones, and somewhat improved agreement is obtained by moving the sulfur atoms slightly closer to the arsenic atom and away from the copper atoms, giving the distances arsenic-sulfur and 2.32 Å for copper-sulfur. Although the structure depends upon thirteen parameters, there is little doubt that it is correct.

A similar investigation, with, however, different results, was carried out for the mineral sulvanite. Sulvanite, $\text{Cu}_3\text{VS}_4$, has been found as a massive mineral in Burra Burra, Australia, and in cleavable masses and a few small crystals near Mercur, Utah. It was found to be

---

L. Pauling and R. Hultgren, Z. Krist., 84, 204 (1933).

---


cubic, with unit of structure with \( a = 5.370 \) \( \text{Å} \), containing three copper atoms, one vanadium atom, and four sulfur atoms. The fact that the unit of structure of the cubic form of zinc sulfide has \( a = 5.42 \text{Å} \) and contains strongly four zinc atoms and four sulfur atoms / suggests that sulvanite has a similar structure, with copper atoms replacing three of the zinc atoms and vanadium the fourth. It immediately noted, however, that the observed intensities differ greatly from those calculated for this structure; the stochastic method in this case has failed. There is, moreover, no other plausible structure that suggests itself. Fortunately, the crystal is simple enough to permit the rigorous method of structure determination to be applied. Its application was found to lead to the structure shown in Figure 2.

This structure is a surprising one - it puzzles me nearly as much today as it did twenty years ago, when it was discovered. The four sulfur atoms are in the same positions as for sphalerite, and the three copper atoms are in the positions occupied by three of the four zinc atoms in sphalerite. The position of the fourth zinc atom was noted, however, unoccupied; instead, the vanadium atom is intrinsically squeezed-
in a position on the same side of the sulfur atoms as that toward which the sulfur-copper bonds extend, rather than the opposite side, which would complete the tetrahedral configuration of metal atoms about the sulfur atom. This structure is satisfactory in that each metal atom is surrounded tetrahedrally by sulfur atoms. It is surprising, however, that the bonds formed by the sulfur atoms are not tetrahedrally directed. The explanation of the stability of this peculiar atomic arrangement may lie in the formation of covalent bonds between the vanadium atom and the six surrounding copper atoms, and also in the Coulomb stabilization of this structure, relative to the structure similar to that of sphalerite, through the achievement of a small negative electric charge by the copper atoms and a small positive charge by the vanadium atoms, which would result from the formation of single bonds with the four sulfur atoms, and by the copper atoms of bonds with approximately covalent character with the four sulfur atoms.

One of the most complex crystal structures to have been discovered is that of zenyite.\(^5\) Zenyite is an aluminosilicate that occurs as

colorless, transparent tetrahedra, about as hard as quartz. The mineral has been found in San Juan County, Colorado, intimately mixed with guiterminate, a lead arsenate, and in Ouray County, Colorado, in an altered porphyrite; sharp crystals of zunyite have also been found with hematite as powder in pots from graves in Uaxactun, Guatemala. Its composition could not be determined from the chemical analyses until the x-ray studies had been made, which provided a determination of the formula weight. The cubic unit of structure was found to have $a = 13.82 \text{Å}$, and to contain four molecules of composition $\text{Al}_3\text{Si}_5\text{O}_{20}(\text{OH}, \text{F})_8\text{Cl}$.

It was known in 1933 that in all silicates investigated up to that time each silicon atom was surrounded tetrahedrally by four oxygen atoms, with the silicon-oxygen distance about 1.60 Å, and that aluminum atoms were usually surrounded octahedrally by six oxygen atoms, with aluminum-oxygen distance 1.90 Å, although a tetrahedral environment of aluminum atoms had occasionally been found in a few minerals. It was not easy to find a way of combining tetrahedra and octahedra by the sharing of corners and edges to give a cubic structure with the right dimensions, but finally a single structure
was discovered, and the calculated/intensities for the structure were found to be in good agreement with experiment. In this structure in the crystal are complexes of twelve aluminum octahedra, which are held together by sharing corners with one another, and also with a single aluminum tetrahedron and a complex of five silicon tetrahedra. This tetrahedron-octahedron framework defines cavities which are occupied by chloride ions.

The structures of many other minerals (topaz, mica, musolinite, chlorite, swedenborgite, etc.) were determined by the stochastic method.

During recent years a number of complex intermetallic compounds have been successfully investigated in the same way. A fundamental structural principle for metals and alloys is that in general each atom surrounds itself by as many neighbors as possible. Some of the most interesting metal structures that are known have been discovered as a result of the recognition of the significance of polyhedra of the icosahedral group. These polyhedra, which have five-fold axes of symmetry as well as three-fold axes and two-fold axes, cannot, of course, retain their symmetry elements completely in crystals, which cannot have true icosahedral polyhedra five-fold axes. However, it is found that they can be present in crystals.
with only slight deformation.

It is interesting that the icosahedral arrangement of twelve atoms around a central atom provides what may be called a closer packing than closest packing — that is, than closest packing of spheres of the same size. In the icosahedral configuration the central atom can be 10 percent smaller than the twelve surrounding atoms, which are still able to make contact with it. The most complex metal structure to have been determined is that of Mg$_3$Zn, an icosahedron of twelve slightly larger atoms. This group of thirteen atoms is then surrounded by twenty atoms, at the corners of a pentagonal dodecahedron, each of these twenty atoms thus lying directly out from the center of one of the twenty faces of the icosahedron. The next twelve atoms lie out from the centers of the pentagonal faces of the dodecahedron; this gives a complex of forty-five atoms, the outer thirty-two of which lie at the corners of a rhombic triacontahedron.

The next shell consists of sixty atoms, each directly out from the center of a triangle which forms one half of one of the
thirty rhombs bounding the rhombic triacontahedron. These sixty atoms lie at the corners of a truncated icosahedron, which has twenty hexagonal faces and twelve pentagonal faces. Twelve additional atoms are then located out from the centers of twelve of the twenty hexagonal faces. The complexes are then condensed together, with their centers at the point of a body-centered cubic lattice, in such a way that each of the seventy-two outer atoms is shared between two complexes; each then contributes thirty-six atoms per lattice point, which with the central complex of forty-five atoms gives eighty-one atoms per lattice point, or 162 in the unit cube.

The complexity of this structure may be attributed to the difficulty of fitting complexes with icosahedral symmetry into a crystal with cubic symmetry.

It seems not unlikely that the complexity of the intermetallic compound with the simple formula NaCd₂ is to be attributed to the same cause. I have now been working on the problem of the structure of this crystal for thirty years, and the structure still remains

undiscovered. The crystal is cubic, with the edge of the unit cube slightly over 30 Å. This unit contains about 384 sodium atoms and 768 cadmium atoms. The attempt to determine its structure by rigorous methods would, of course, be hopeless; but I think that the stochastic method will ultimately be successful.

The power of the stochastic method is illustrated by its recent application to the problem of the configuration of polypeptide chains in proteins. The history of this application is illuminating also in showing that an investigator who strives to apply the method must have confidence in himself. In 1937, after I had become interested in proteins, and had carried out a number of experimental studies of their properties (especially the magnetic properties of hemoglobin), and, with Dr. Alfred Mirsky, had formulated a general theory of the structure and process of denaturation of proteins, I spent several months in an unsuccessful effort to apply the stochastic method in the discovery of an acceptable structure for a keratin.

It was possible at that time to predict that the amide group in a peptide would be planar, because of the resonance of the double bond between the two positions C-O and C-N, and it was possible to predict the interatomic distances and bond angles, essentially as given in Figure 5, with reasonable confidence. In addition it was recognized that a stable configuration would be one that permitted the formation of hydrogen bonds between the N-H group and the oxygen atom of the carbonyl group, with N-H...O distance approximately 2.80 Å. No reasonable configuration was found in this early investigation, however, and in consequence the possibility was considered that the structural parameters of the polypeptide chain might be significantly different from those predicted from information obtained through the determination of the structures of somewhat distantly related substances. At that time no structure determination had been made of any amino acid, simple peptide, or other simple substance closely related to proteins. My colleague Professor Robert B. Corey, who had some years earlier made x-ray photographs of several proteins with R. W. G. Wyckoff, had attempted to analyze and I concluded that we should embark upon a program of precise structure determination of these simple
steady substances. This program has been/under way since 1937, and has led
precise to the/determination of the structures of crystals of half a dozen amino
acids, several simple peptides, and several other simple substances (such
as acetylglycine) closely related to proteins. Through these investigations it was found that the planarity of the amide
group and the interatomic distances and bond angles are preserved in all
of the simple substances and can confidently be expected to apply also
to the polypeptide chains in proteins. In addition, hydrogen bonds with
N-H...O distance $2.79 \pm 0.12$ Å have been found to be universally
present.

Even after this program was well under way, and it was recognized that the structural parameters of the polypeptide chain
would be reasonably well predicted, there was delay in the application
of the stochastic method to the problem of the structure of proteins.
This delay probably resulted from the failure of the earlier effort and
from the feeling that the chemical complexity of the proteins - their
construction from about twenty different kinds of amino-acid residues -
might well indicate a corresponding structural complexity. Then one day
in March 1948, while I was again recovering from a cold, I decided to
attack the problem of the configuration of polypeptide chains, for the
first time in eleven years. It occurred to me to make a search for the
simplest configurations - those in which all of the amino-acid residues
are structurally equivalent. The most general operation that converts an
asymmetric element (not its mirror image) is a rotation about an axis combined with a translation along the axis. The
repetition of this general operation automatically leads to a helix.

I attempted accordingly to find configurations of polypeptide chains involving planar amide groups with known dimensions, such that suitable hydrogen bonds were formed. Within an hour, with the aid of a pencil and a piece of paper, I had discovered a satisfactory helical structure. It

did not, however, explain satisfactorily the details of the x-ray diagram of hair and other a-keratin proteins, and nothing more was done along those lines for some months.
After my return to Pasadena, Professor Corey and I suggested to Dr. H. R. Branson, a young man interested in the application of mathematics to chemical problems, that he make a search for other satisfactory helical configurations. He found one additional satisfactory configuration, and in 1951 a description of the two helixes, the α helix and the γ helix, was published.


Although the α helix, which has about 3.6 amino-acid residues per turn and a pitch of about 5.4 Å, did not explain in an obvious way the principal feature of the x-ray diagram of the α-keratin proteins, a strong meridional reflection with spacing 5.15 Å, its predicted x-ray pattern was found to be in excellent agreement with the observed pattern for synthetic polypeptides. Moreover, the general similarity in appearance of the x-ray diagram for the α-keratin proteins and that for the synthetic polypeptides gave reasonably strong support to the assignment of the α helix to these fibrous proteins also. The problem of the
origin of the 5.15 Å meridional reflection seems to have been solved by the simultaneous suggestion by F. H. C. Crick, and Professor Corey and me that in the α-keratin proteins the α helixes are twisted about one another. A detailed structure is that shown in Figures 7 and 8; it involves the seven-strand cable of α helixes, with two additional α helixes in the interstitial positions.

\[\text{References}\]
\[\text{F. H. C. Crick, Nature, 170, 1882 (1952)}\]

Reasonably convincing evidence has also been obtained, largely through the work of Perutz and Kendrew, who have studied hemoglobin and myoglobin, and of Riley and Arndt, who have prepared radial-distribution curves of a number of proteins, that globular proteins contain segments with the configuration of the α helix. In hemoglobin and myoglobin these α-helix segments lie in approximately parallel orientation to one another.
No detailed information has so far been obtained about the way in which the polypeptide chains make the transition from one α-helix segment to another.

Silk fibroin is shown by its X-ray diagram to have a structure which repeats in the distance 7.00 Å along the fiber axis. The antiparallel chain pleated-sheet structure, shown in Figure 1, is predicted to have the identity distance 7.00 Å along the fiber axis, and in other respects it accounts satisfactorily for the X-ray diagram of silk fibroin; it can be confidently accepted as representing this protein. It is probable that the β-keratin proteins (such as stretched hair) have a closely similar structure, with, however, the polypeptide chains in parallel, rather than antiparallel, orientation.

The X-ray diagram of collagen and gelatin is characteristic of these proteins, showing that they have a structure different from that of α-proteins of the keratin class, and also different from that of silk and the β-keratin proteins. The stochastic method was used in the formulation of a structure for collagen and gelatin, which, however, has since been found to be in disagreement with some of the features of the X-ray diagram. Despite much effort that has been expended on the problem, L. Pauling and R. B. Corey, Proc. Nat. Acad. Sci., 37, 272 (1951)
no satisfactory structure for collagen and gelatin has yet been found. The problem of the structure of collagen and gelatin may be used to illustrate an important aspect of the stochastic method. The first step in the application of this method is to make a hypothesis - a guess. The second step is to test the hypothesis, by some comparison with experiment. In general the test cannot be sufficiently thorough to provide proof that the hypothesis is correct - it may easily be shown that the hypothesis is incorrect, through the discovery of a significant disagreement with experiment. Agreement on a limited number of points cannot be accepted as verification of the hypothesis. In order for the stochastic method to be significant, the principles used in formulating the hypothesis must be restrictive enough to make the hypothesis itself essentially unique; in other words, investigator who makes use of this method should be allowed one guess. If he were allowed many guesses he would sooner or later make one that was not in disagreement with the limited number of test points, but there would then be little justification for accepting that hypothesis as correct. I may, however, contend that Professor Corey and I should be allowed two guesses on
and we are determined that our second one shall be right.

We may ask whether or not the quality of simplicity in postulated explanations of natural structures and phenomena is to be recognized. It is clear that living organisms are not simple, and yet there is a tendency for simplicity—the use of the same structural element or mechanism over and over again: the hemoglobin apparatus, cytochrome, and other conjugated proteins; L amino acid residues instead of D; the use of structural gene duplication, enzyme activity, and other phenomena.

Perhaps the time has come to penetrate into the principles of the extensive method, in order to get a better understanding of it, to make it more powerful, the problems of biology and medicine are of great importance, and they are so difficult that we cannot afford to allow a possible solution.
Legends for Figures

Fig. 1. The structure of enargite, $\text{Cu}_3\text{AsS}_4$. The large circles represent sulfur atoms, the small circles copper atoms, and the small shaded circles arsenic atoms.

Fig. 2. The structure of sulvanite, $\text{Cu}_3\text{VS}_4$.

Fig. 3. A portion of the structure of zunyite, $\text{Al}_{13}\text{Si}_{5}\text{O}_{20}(\text{OH}, \text{F})_{18}\text{Cl}$. $\text{SiO}_4$ and $\text{AlO}_4$ groups by tetrahedra, the last being marked Al. Smaller spheres represent oxygen atoms, larger spheres chloride ions. Groups of five tetrahedra and twelve octahedra preserve their identity in the structure.

Fig. 4. Diagrammatic representation of the configuration of the polypeptide chain in the $\alpha$ helix.

Fig. 5. The $\alpha$ helix.

Fig. 6. At the left, a compound $\alpha$ helix — an $\alpha$ helix whose axis describes a helical configuration. The diameter, shown as about 10 Å.
includes the volume occupied by side chains as well as the main chains of the protein. Center, a 7-strand α-cable. In the proposed structures of proteins of the α-keratin type these καβλε cables are packed together, with compound helixes as shown at the left in the interstices. At the right, a 3-strand rope of α helixes.

Fig. 6. A cross section of the α-keratin structure, showing the 7-strand α-cables AB₆ and the interstitial compound helixes C.

The orientation of the cross-section of the cable changes with coordinate along the fiber axis. The central cable is shown in the most unfavorable orientation for the interstitial α helixes. The protein chains are not so nearly circular in cross section as indicated in the drawing, and space is filled more effectively than is indicated.

Fig. 8. Drawing representing the antiparallel-chain pleated-sheet structure.
Fig. 4. The structure of the intermetallic compound Mg$_{32}$(Zn,Al)$_{49}$.
The six drawings, from left to right in the top row and then left to right in the bottom row, have the following significance: a central atom surrounded by a nearly regular icosahedron; the icosahedral group of thirteen atoms surrounded by a pentagonal dodecahedron; the complex of thirty-three atoms surrounded by twelve atoms at the corners of an icosahedron; the outermost shell of sixty atoms at the corners of a truncated icosahedron, plus twelve atoms out from the centers of twelve of the hexagons of this polyhedron; packing drawing showing the complex of forty-five atoms plus an outer shell of seventy-two atoms; the structure of the crystal, in which these complexes located about the points of a body-centered cubic lattice share all of the seventy-two atoms of the outermost shell with neighboring complexes.