MACROMOLECULAR STRUCTURE FROM ANOMALOUS DISPERSION

A simple algebraic analysis of diffraction data from the unusual scattering near an absorption edge reveals the structures of substances that play a primary role in the processes of life.

Jerome Karle

Knowledge of the structures of macromolecules is of major importance both to basic science and to new, burgeoning industries. The determination of these structures is improving our understanding of fundamental biological processes, and so it sets the stage for the broad explanation of diseases on the molecular level and for the rational design of drugs. It promises to give insight into such topics as the mechanisms of enzyme action, the implications of protein structure for evolutionary development, the relation of amino acid sequences to three-dimensional structure and the relation between structure and function in complicated systems such as viruses, muscle, ribosomes and chromatin.

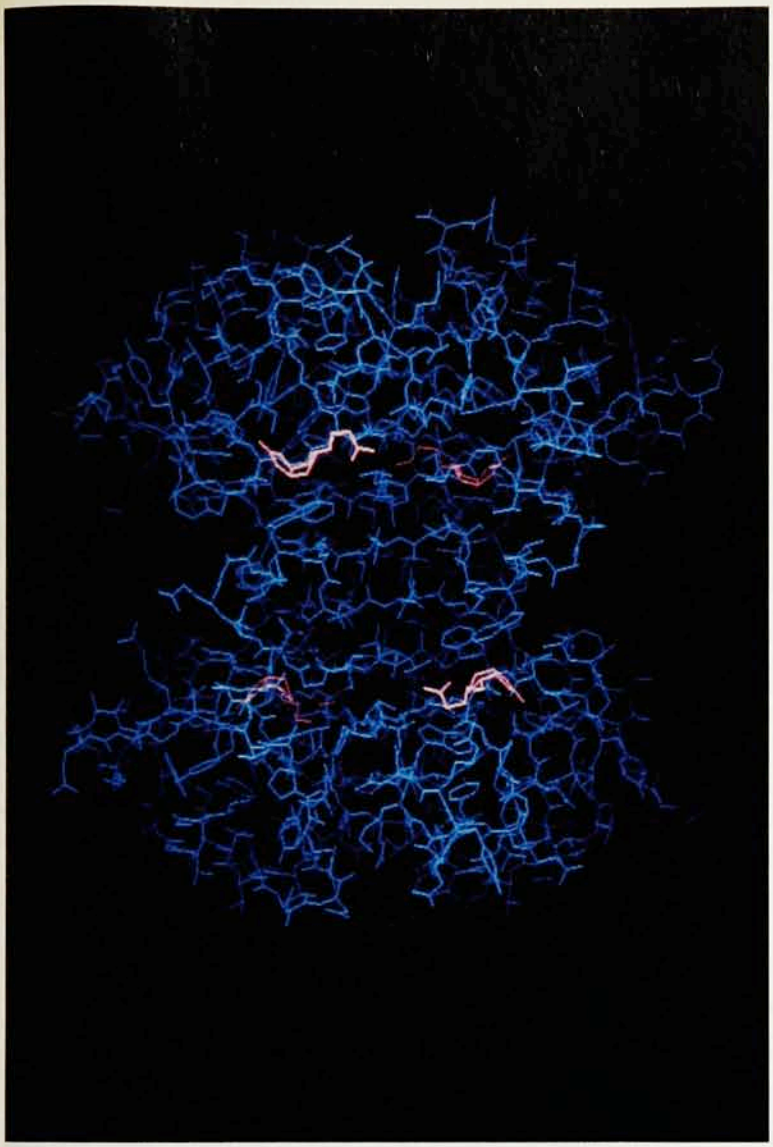
It is a remarkable experience to witness the elucidation of a complex structure from the information in a diffraction pattern. The beauty of such revelations is not diminished by the fact that successful determinations of structure are now commonplace. Figure 1 shows a protein structure recently solved[1] by the technique I will discuss in this article. It is the macromolecule streptavidin, from the bacterium *Streptomyces avidinii*. Avidins have the property of binding very strongly to a growth factor, biotin, found in all cells in minute amounts. Biotin, also known as vitamin H, plays an indispensable role in many carboxylation reactions in living systems. Figure 2 shows the structural form of streptavidin and the electron density distribution (obtained from diffraction data) from which the structure was derived.

The use of diffraction methods to investigate the structures of substances in the crystalline state originated in physical inquiry and has benefited greatly from developments in physics. Advances in structure analysis have accompanied advances in diffraction physics, automated equipment, high-intensity x-ray and neutron sources, computing facilities and analytical techniques. This article describes an analytical technique that takes advantage of the phenomenon of anomalous dispersion to determine the structures of large molecules. Anomalous dispersion refers to a modification of scattering intensities due to absorption processes involving interactions between an incident beam of photons or particles and the atoms in a structure. The technique is facilitated by the use of tunable synchrotron radiation and by several developments in instrumentation, including the efficient recording of diffraction patterns.

Structure factors

The immediate objective of structure analysis of crystalline materials by diffraction methods is to determine the average positions of the atoms that form the crystal. There are additional phenomena of interest that can be determined, such as internal motion and precise electron density distributions. This article, however, is confined to the subject of average atomic positions. I will begin by reviewing some of the basic mathematical and physical concepts used to represent and analyze crystal structures. I will then explain how these concepts can be used as a foundation for describing the special features of anomalous dispersion and the consequences of its presence in

Jerome Karle occupies a Chair of Science at the Naval Research Laboratory, in Washington, DC, and is chief scientist at NRL's Laboratory for the Structure of Matter.



Final structure of the macromolecule streptavidin, a protein, in the form of a tetramer. Attached to each of the protomers (blue) is a molecule of selenobiotin (pink). The "beta sheets" of each protomer are arranged in such a way as to give the protomer a barrel-like shape. The beta sheets, represented by wide arrows in figure 5, are made up of adjacent polypeptide chains held together weakly by hydrogen bonding. (Courtesy of Wayne A. Hendrickson and Arno Pähler, Howard Hughes Medical Institute, Columbia University.) **Figure 1**

diffraction experiments.

A crystal is characterized by a basic structural unit called the unit cell, which, through a myriad of repetitions in three dimensions, generates the crystal. The problem of determining the structure of a crystal is solved when the structure of a unit cell is determined. Furthermore, when a crystal possesses symmetries such as axes and planes of symmetry (which is usually the case), the structure of only part of a unit cell called the asymmetric unit needs to be determined.

Consistent with the three-dimensional periodicity of a crystal, the mathematical description of a crystal structure is given by a Fourier series representing the electron density distribution $\rho(\mathbf{r})$ in the crystal:

$$\rho(\mathbf{r}) = V^{-1} \sum_{\mathbf{h}=-\infty}^{\infty} F_{\mathbf{h}} \exp(-2\pi i\, \mathbf{h}\cdot\mathbf{r}) \qquad (1)$$

In this expression V is the volume of the unit cell, $\mathbf{r}$ is a vector whose components locate a point in the unit cell, and $\mathbf{h}$ is a vector whose integral components identify imaginary planes cutting through the crystal. The components of $\mathbf{h}$ are the integers h, k and l, which are inversely proportional to the intercepts of the corresponding planes on the axes chosen to define the unit cell. In view of the fact that the maxima of the electron density distribution $\rho(\mathbf{r})$ locate the centers of atoms, it is apparent that equation 1 represents the arrangements of atoms in crystals.

The coefficients $F_{\mathbf{h}}$, called structure factors, may be real or complex numbers that represent characteristics of the x-ray scattering associated with the planes $\mathbf{h}$. The structure factors $F_{\mathbf{h}}$ are defined as the ratios of the amplitudes of coherent scattering from the plane segments in a unit cell to the amplitude of scattering from a free electron at the origin of the unit cell under the same conditions.

The structure factors can be written

$$F_{\mathbf{h}} = |F_{\mathbf{h}}| \exp(i\phi_{\mathbf{h}})$$

where the angle $\phi_{\mathbf{h}}$ is the phase associated with the factor $F_{\mathbf{h}}$. The incident x-ray wave is scattered by the atoms in such a way that it appears to be reflected by segments of crystal planes in a unit cell labeled by $\mathbf{h}$ with a maximum scattered amplitude of $|F_{\mathbf{h}}|$ and at a point in the scattered wave cycle given by the phase $\phi_{\mathbf{h}}$. This is not the usual concept of reflection because the reflected beam occurs only for a specific value of the angle of incidence. It is therefore necessary to change the orientation of a crystal

continually when using monochromatic radiation, so as to collect a complete set of diffraction data. The phases $\phi_{\mathbf{h}}$ may assume all values between $-\pi$ and π except under special circumstances. A crystal with a center of symmetry is one such special case. Here the phases can assume only two possible values, 0 or π.

Need for specialized techniques

If the coefficients $F_{\mathbf{h}}$ of the Fourier series were directly obtainable from an x-ray diffraction experiment, the structure of any crystal could be readily computed from equation 1. The intensities of the diffracted beams, however, are proportional to the quantities $|F_{\mathbf{h}}|^2$. Hence, in an ordinary diffraction experiment only the magnitudes $|F_{\mathbf{h}}|$ are obtained; the associated phases $\phi_{\mathbf{h}}$ are not.

At one time the absence of phase information was perceived as unsolvable in principle, and it was generally thought that crystal structures could be solved only by use of special techniques such as trial and error reinforced with experience or by introducing heavy atoms into the structure. An important feature of the electron density distribution $\rho(\mathbf{r})$, however, makes this view too pessimistic—namely that $\rho(\mathbf{r})$ can be represented as a sum of N discrete, known atomic electron distributions per unit cell. This is demonstrated by performing the Fourier inversion of equation 1 and reducing the integral expression for the Fourier coefficient to the sum

$$|F_{\mathbf{h}}| \exp(i\phi_{\mathbf{h}}) = \sum_{j=1}^{N} f_{j\mathbf{h}} \exp(2\pi i\, \mathbf{h}\cdot\mathbf{r}_j) \qquad (2)$$

In this sum, $f_{j\mathbf{h}}$ is the atomic scattering factor of the jth atom in a unit cell containing N atoms, and $\mathbf{r}_j$ is its position. The atomic scattering factor is defined as the ratio of the amplitude of coherent scattering for an atom to that for a free electron at the atomic center under the same conditions.

Equation 2 can be considered a system of simultaneous equations with two equations for each $\mathbf{h}$—one for the real and one for the imaginary part. An analysis of the equation indicates that for the number of independent data obtained from the use of Kα radiation from a copper-source x-ray tube, the problem of locating the unknown atomic positions is highly overdetermined. (Kα radiation arises from specific electron transitions from the L shell to the K shell.) This is also true for the unknown phases. The quantities in equation 2 that are known to good accuracy are the $|F_{\mathbf{h}}|$ that come from the experimental intensities and the $f_{j\mathbf{h}}$ that have been calculated for free atoms and tabulated.

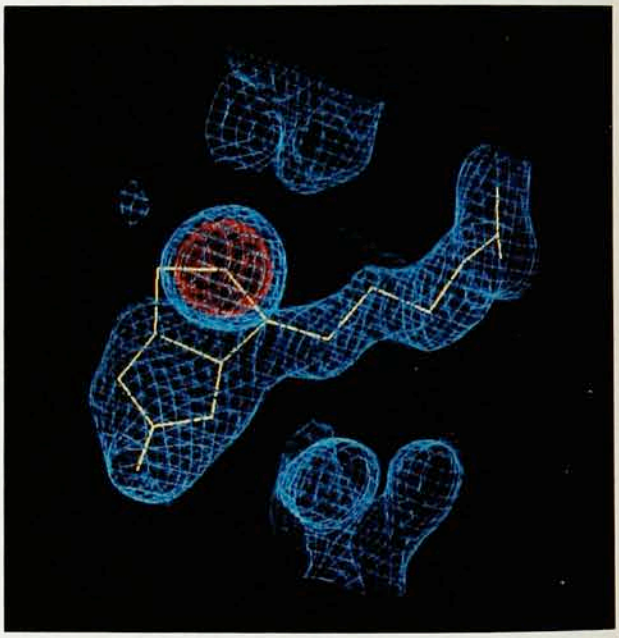
Useful relationships between phases and the magnitudes of structure factors arise from the simple fact that the electron density distribution in a crystal is a nonnegative function. It is these relationships, and their corresponding probabilistic implications, that provide the mathematical foundation for the solution to the phase problem. After some years of theoretical work and further years of practical development and experimental experience, a broadly applicable procedure for "direct" phase determination was developed.[2] This procedure, with various modifications, has been in use for the past 25 years. Increasing structural complexity leads to a decrease in the reliability of the probability measures and to a decrease in the relative amount of data. This accounts for the difficulty in applying the procedures for direct phase determination to complex structures.

A Fourier series called a Patterson map, denoted as $P(\mathbf{r})$, represents the interatomic vectors in a crystal and offers an opportunity to determine atomic positions directly from the measured intensities.[3] The function is

$$P(\mathbf{r}) = \sum_{\mathbf{h}=-\infty}^{\infty} |F_{\mathbf{h}}|^2 \exp(-2\pi i\, \mathbf{h}\cdot\mathbf{r}) \qquad (3)$$

The coefficients are known from intensity measurements over the experimental range and contrast with the coefficients in equation 1, whose magnitudes are known from intensity measurements but whose phases must somehow be otherwise evaluated. General application of equation 3 has been limited by a lack of resolution of the $N(N-1)$ interatomic vectors in the map as calculated from the available data. Nevertheless a Patterson map can become quite useful, even for complex structures, when particular structural features are known or when some heavy atoms are present. Interatomic vectors associated with a few heavy atoms can usually be readily identified. The positions of the heavy atoms can be deduced from their interatomic vectors, permitting one to apply a variety of so-called heavy-atom techniques to the recovery of an entire structure.

The discussion so far answers the often-asked question



Structural form of the selenobiotin molecule (yellow) in the complex with streptavidin, and the electron density distribution from which this structure was derived (blue). The electron density distribution was obtained from diffraction data at 2-Å resolution. The blue contours are drawn at a separation of approximately 0.35 electrons/Å^3, and the red contours, which depict the selenium position, are drawn at approximately 3.5 electrons/Å^3. (Courtesy of Hendrickson and Pähler.) **Figure 2**

of why direct methods of phase determination, which are so broadly applicable to small structures, are not also applied to macromolecules. The simplest proteins are made up of several hundred atoms, and because of positional disorders such as large-amplitude motions and conformational variation, the range of measurable data is much less than for smaller, more rigid structures. As noted above, such circumstances make current direct methods for phase determination inapplicable, and so one must turn to some specialized techniques for determining the phase angles. One such technique is the main subject of this article.

Heavy atoms

Heavy-atom derivatives of macromolecules have provided the means for evaluating the phases required for structure determination. The techniques used in this evaluation are isomorphous replacement and anomalous dispersion. Crystals having the same unit-cell geometry but different chemical composition are called isomorphous. Isomorphous crystals that play a useful role in protein structure analysis usually consist of the native protein and the same protein in which some heavy atoms have been added or have replaced lighter atoms.

Isomorphous replacement has played a major role in the vast progress made in the elucidation of protein structures since the first structures were solved about 30 years ago. Anomalous dispersion has been applied to a lesser extent, usually in combination with isomorphous replacement. Although the value of each technique is enhanced by its use in combination with the other, anomalous dispersion is now beginning to be used as the sole source of phase information in the investigation of macromolecular structure. It is particularly useful when the formation of isomorphous crystals does not appear to be possible.

Isomorphous replacement. Protein crystals generally have a good deal of water in them. It is therefore possible to add heavy-atom moieties merely by soaking the crystals in solutions of the heavy-atom compounds. The usefulness of isomorphous replacement depends on the rather strict positional invariance of essentially all the atoms except, of course, for the substitutions or additions.

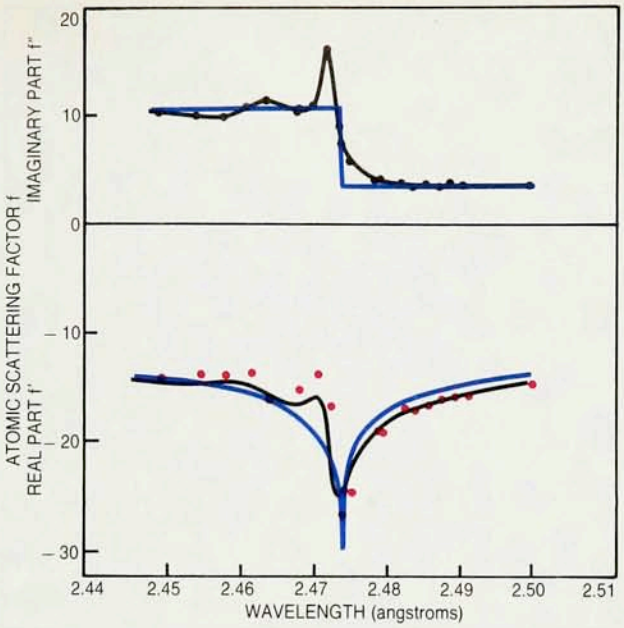
An appropriate equation for an isomorphous pair of crystals is

$$F_{\mathbf{h},R+X} + F_{\mathbf{h},Y-X} = F_{\mathbf{h},R+Y} \tag{4}$$

Here $F_{\mathbf{h},R+X}$ is the structure factor for the structure made up of invariant atoms R and replaceable atoms X; $F_{\mathbf{h},R+Y}$ is the same except with replaceable atoms Y; and $F_{\mathbf{h},Y-X}$ is the structure factor for the difference of the configurations for atoms Y and atoms X, that is, the difference between the structure factor for the Y atoms and the structure factor for the X atoms. If it is assumed that heavy atoms are added to an unsubstituted crystal, equation 4 becomes

$$F_{\mathbf{h},R} + F_{\mathbf{h},Y} = F_{\mathbf{h},R+Y} \tag{5}$$

The analysis of equations 4 and 5 benefits from the information that Patterson maps determine, for example, about the atomic positions of the replaceable atoms X and Y. The structure factors $F_{\mathbf{h},Y-X}$ and $F_{\mathbf{h},Y}$ can then be computed from the known atomic positions via equation 2. From knowledge of $F_{\mathbf{h},Y-X}$ or $F_{\mathbf{h},Y}$ and the magnitudes of the other structure factors in equations 4 or 5, one can derive phase information. If the two structures of an isomorphous pair of crystals are centrosymmetric, the phase values determined through the use of equations 4 or 5 will be unique; if the two structures are not centrosymmetric, the phases will have a twofold ambiguity. A number of methods have been developed to resolve the ambiguity. One is to make additional isomorphous substitutions. Ideally, a second substitution would suffice.



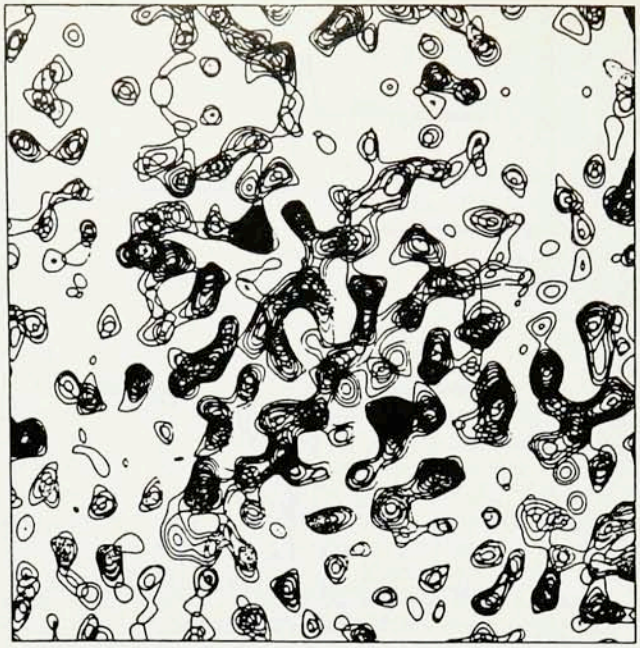
Corrections to the normal, or nonanomalous, atomic scattering factor due to anomalous dispersion of x rays. The upper and lower curves are plots of f'' and f', the imaginary and real parts of the correction, respectively. The circles in the upper plot represent data taken near the L_3 edge of cesium. The circles in the lower plot were calculated from those in the upper plot from the Kramers–Kronig dispersion relation. The colored lines are theoretical plots based on calculations by Don T. Cromer of Los Alamos National Laboratory. (Adapted from reference 4.) **Figure 3**

It is quite customary in the investigation of protein structure to make several isomorphous replacements to overcome problems that arise from experimental uncertainty, lack of exact isomorphy and the occasional lack of distinction between ambiguous phase indications from two different isomorphous replacements. The method of phase determination based on several isomorphous structures is called the multiple isomorphous replacement method. As noted, it has played a most important role in the progress of protein crystallography.

Anomalous dispersion. This phenomenon affects diffraction intensities through its modification of the atomic scattering factors f. The atomic scattering factors from which structure factors can be computed through equation 2 may be defined by

$$f = f^{n} + f' + if'' \tag{6}$$

In this expression f^{n} is the normal, or nonanomalous, scattering factor computed at a radiation frequency that is much larger than the absorption frequencies of the constituent atoms, since anomalous scattering is a conse-



Electron density distribution for the protein selenobiotinyl streptavidin. This section of the distribution was obtained from multiple-wavelength anomalous dispersion phase information at 3.3-Å resolution. Fitting a structural model to such a distribution is facilitated by knowledge of the sequence of peptide residues in the protein. The location in the structure of the peptide residues is determined from structural models, which give approximate values for the atomic positions. The approximate coordinates are usually refined by least-squares methods and by incorporating data of higher resolution. (From reference 1.) **Figure 4**

quence of the absorption process. In the vicinity of an atom's absorption edge there can be significant corrections for f^n. The quantities f' and f'' (often written $\Delta f'$ and $\Delta f''$) are the real and imaginary parts of the correction to f^n arising from the anomalous dispersion. An important consequence of having a significant imaginary part in equation 6 is that the intensity measured for a so-called acentric reflection $\mathbf{h}$ is, in general, different from that for $-\mathbf{h}$, which is also written $\bar{\mathbf{h}}$. The term "acentric" indicates reflections that act like general reflections from noncentrosymmetric crystals. These contrast with centric reflections, which act like reflections from centrosymmetric crystals. The latter have the property that $|F_{\mathbf{h}}| = |F_{\bar{\mathbf{h}}}|$ even when there are strong effects from anomalous dispersion. When the scattering factor f is real, $|F_{\mathbf{h}}| = |F_{\bar{\mathbf{h}}}|$ for all crystals. For acentric reflections affected by anomalous dispersion, $|F_{\mathbf{h}}| \neq |F_{\bar{\mathbf{h}}}|$.

David H. Templeton and Lieselotte K. Templeton of the University of California, Berkeley, and their colleagues have made a number of investigations of absorption edges with the aim of facilitating crystallographic applications of anomalous dispersion. In one study they emphasized the value of using the L absorption edge, the second lowest shell of electrons in terms of energy.[4] Energy is absorbed from the incident beam in removing an electron from the L shell. The Templetons found, for example, that near the L_3 edge of cesium, f' varies between -26.7 and -13.9 electrons, and f'' varies between 4.0 and 16.1 electrons in a wavelength interval of 0.008 Å. They also pointed out that in the actual curves representing f' and f'' there is detail that does not appear in the theoretical calculations, which are smoother than the actual curves because of certain approximations. This is illustrated in figure 3, which shows the variation of f' and f'' near the L_3 edge of cesium for both experimental and theoretical values. It appears that taking full advantage of the major features in the vicinity of absorption edges is made easier by the experimental determination of these features. Theoretical values should be sufficiently accurate for applications as the wavelength departs from the absorption edge. It should be noted, however, that the tabulated values for f' and f'' are for Kα radiation only (from various sources). The use of L edges therefore requires their determination by experimental or theoretical means.

Johannes M. Bijvoet and his colleagues at the University of Utrecht in the Netherlands introduced the anomalous dispersion technique into crystal structure analysis in 1951. Bijvoet and colleagues pioneered the technique in their determination of the absolute configuration of *d*-tartaric acid in the form of its sodium rubidium double salt.[5] Adrianus J. van Bommel at Utrecht later made the same determination using the acid rubidium salt.[6] This work confirmed Emil Fischer's arbitrary convention, which assigned a particular optical rotation (*d* or *l*) to one of the two mirror images of the spatial configuration of tartaric acid.

Bijvoet also recognized in the 1950s that anomalous dispersion could be used for phase determination. He found that a single-wavelength experiment could give phase values for acentric reflections, with a twofold ambiguity similar to the experimental result of single isomorphous replacement. The combination of single-wavelength anomalous dispersion and single isomorphous replacement could resolve the ambiguity because the false answers obtained from the techniques are the ones that do not agree with each other. In the algebraic analysis to be presented here, the two techniques may be combined in such a way that the false answers never appear. Over the years, a variety of investigations have been devoted to methods for enhancing the quality of the data, for analyzing the data and for resolving the phase ambiguities.[7]

A number of developments since the mid-1970s have enhanced interest in the anomalous dispersion technique. They are theoretical, instrumental and experimental in character. For example, by modifying the algebraic analysis of anomalous dispersion experiments, it is now possible to derive a readily applied system of exact, linear simultaneous equations whose unknown quantities are of special utility in determining crystal structures. Improvements in instrumentation include the high intensity and tunability of synchrotron radiation sources, which can greatly facilitate the collection of anomalous dispersion data. The rate of data collection has also been enhanced by the development of instrumentation such as area detectors for determining the positions and intensities of the diffracted x rays. To broaden the applicability of the techniques, some general procedures in genetic engineering are being applied to build proteins with suitable anomalous scatterers.

Algebraic analysis

In the late 1970s Wayne A. Hendrickson and Martha M. Teeter at my laboratory used anomalous scattering data to solve the structure of a small (approximately 5000 dalton)

protein called crambin.[8] The protein molecule contains six cysteine residues in disulfide linkages, and the structure was solved by anomalous scattering from the sulfur atoms at a single wavelength—that of the CuKα line, which at 1.54 Å is quite far from the absorption edge of sulfur at 5.36 Å. A positive result from the use of relatively weak anomalous scatterers in a one-wavelength experiment so far from the absorption edge seemed to imply that further development of the anomalous dispersion technique held considerable potential. This rekindled in me an earlier interest in the analysis of multiple-wavelength anomalous dispersion data.

The key feature of the algebraic analysis whose results are discussed here is a simple alteration in the way the analysis proceeds. It involves treating the atomic scattering factor in equation 2 in terms of two separate entities: the normal scattering f^n and the corrections f' and f''. Prior to the development of this analysis it had been customary to combine f^n with the real part of the correction, f', and to treat the imaginary part f'' separately, thereby forming a real part and an imaginary part of the scattering factor, both of which vary with wavelength. In contrast, keeping f^n separate from the corrections permits an important portion of the mathematical description to be invariant with respect to wavelength. The resulting equations are composed of unknown wavelength-invariant quantities that describe the structure of interest and wavelength-dependent quantities that act only as coefficients of the unknowns and are defined in terms of tabulated normal atomic scattering factors and their corrections.

The structure factor $F_{\lambda\mathbf{h}}$ in the presence of atoms that scatter anomalously to a significant extent may be written as

$$F_{\lambda\mathbf{h}} = F^{\,n}_{\mathbf{h}} + F^{\,a}_{\lambda\mathbf{h}} \tag{7}$$

Here λ indicates variation with wavelength, $|F_{\lambda\mathbf{h}}|$ is obtained from the measured x-ray intensities, $F^{\,n}_{\mathbf{h}}$ is the structure factor that would be obtained if all atoms scattered nonanomalously, and $F^{\,a}_{\lambda\mathbf{h}}$ is the structure factor obtained solely from the anomalous corrections to the normal atomic scattering. This analysis assumes that vibrational effects can be ignored, for it is possible to remove them from $|F_{\lambda\mathbf{h}}|$ to good approximation. The structure factors $F^{\,n}_{\mathbf{h}}$ and $F^{\,a}_{\lambda\mathbf{h}}$ are defined as follows:

$$F^{\,n}_{\mathbf{h}} = \sum_{j=1}^{N} f^{\,n}_{j\mathbf{h}} \exp(2\pi i\, \mathbf{h}\cdot\mathbf{r}_j)$$

$$F^{\,a}_{\lambda\mathbf{h}} = \sum_{j=1}^{N} (f'_{\lambda j} + i f''_{\lambda j}) \exp(2\pi i\, \mathbf{h}\cdot\mathbf{r}_j) \tag{8}$$

The corrections $f'_{\lambda j}$ and $f''_{\lambda j}$ are treated as constants with respect to the $\mathbf{h}$. As mentioned, the factors $f^{\,n}_{j\mathbf{h}}$ and some values of $f'_{\lambda j}$ and $f''_{\lambda j}$ have been tabulated. For many atoms the anomalous corrections are very small, in which case the corresponding f' and f'' in equation 8 can be set to zero. An alternative way of writing equation 8 is

$$F^{\,a}_{\lambda\mathbf{h}} = \sum_{j=1}^{N} f^{\,a}_{\lambda j} \exp(i\delta_{\lambda j}) \exp(2\pi i\, \mathbf{h}\cdot\mathbf{r}_j)$$

where

$$f^{\,a}_{\lambda j} = [(f'_{\lambda j})^2 + (f''_{\lambda j})^2]^{1/2}$$

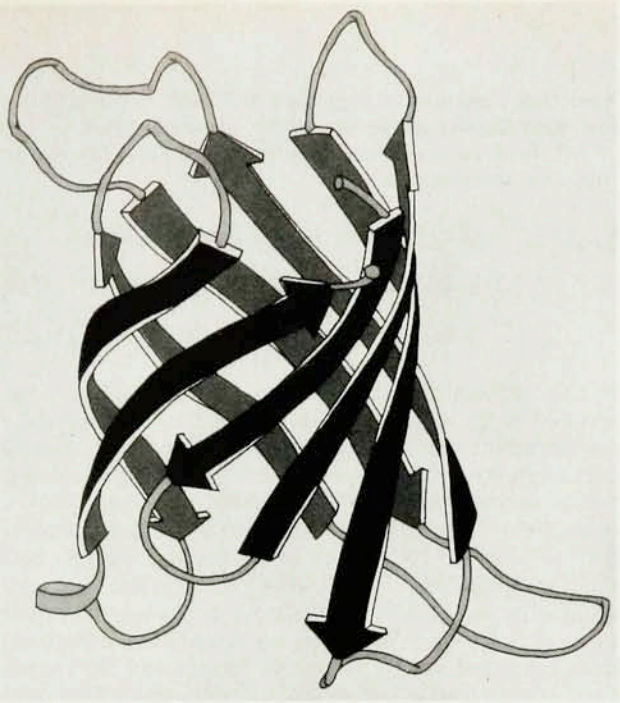
and

$$\delta_{\lambda j} = \tan^{-1}(f''_{\lambda j}/f'_{\lambda j})$$

It follows from equation 7 that

$$|F_{\lambda\mathbf{h}}|^2 = |F^{\,n}_{\mathbf{h}}|^2 + |F^{\,a}_{\lambda\mathbf{h}}|^2 + 2|F^{\,n}_{\mathbf{h}}||F^{\,a}_{\lambda\mathbf{h}}|\cos(\phi^{\,n}_{\mathbf{h}} - \phi^{\,a}_{\lambda\mathbf{h}}) \tag{9}$$

The three quantities on the right side of equation 9 can be expressed in terms of structure factor magnitudes and phases that describe individual atomic arrangements of anomalous scatterers of different types. These magnitudes and phases also have the virtue of being invariant with respect to wavelength and thus form suitable unknown quantities for simultaneous equations that vary with wavelength. A completely general expression has been obtained that is valid for any number of atoms and any number of types of atoms.[9]



Symbolic representation of the structure of the protomer of selenobiotinyl streptavidin. The molecule is approximately barrel-shaped—the curved "staves" represent beta sheets. Such a structure is known as a beta barrel. (Courtesy of Hendrickson and Pähler.) **Figure 5**

A specialization

Crystallographers have begun to apply a simple special case of the general mathematical system to the solution of protein structures. In this case there are a large number of atoms whose anomalous dispersion corrections are so

weak that they may be neglected, and there are also only a few anomalously scattering atoms, all of one type.

If the system of simultaneous equations is applied to this case, the result is[9]

$$|F_{\lambda\mathbf{h}}|^2 = |F^n_{1,\mathbf{h}}|^2 + [1 + Q(Q + 2\cos\delta_{\lambda 2})]\,|F^n_{2,\mathbf{h}}|^2 + 2(1 + Q\cos\delta_{\lambda 2})|F^n_{1,\mathbf{h}}||F^n_{2,\mathbf{h}}|\cos(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}}) + 2Q\sin\delta_{\lambda 2}\,|F^n_{1,\mathbf{h}}||F^n_{2,\mathbf{h}}|\sin(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}}) \quad (10)$$

In this expression Q is the ratio $f''_{\lambda 2}/f^n_{2,\mathbf{h}}$, $|F^n_{1,\mathbf{h}}|$ is the magnitude of the structure factor for the normally scattering atoms, $|F^n_{2,\mathbf{h}}|$ is the magnitude of the normal part of the structure factor for the anomalously scattering atoms, and $\phi^n_{1,\mathbf{h}}$ and $\phi^n_{2,\mathbf{h}}$ are the associated phase angles, respectively. The variables in equation 10 can be taken to be $|F^n_{1,\mathbf{h}}|^2$, $|F^n_{2,\mathbf{h}}|^2$, $|F^n_{1,\mathbf{h}}||F^n_{2,\mathbf{h}}|\cos(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}})$ and $|F^n_{1,\mathbf{h}}||F^n_{2,\mathbf{h}}|\sin(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}})$. One can see that the variation with wavelength is restricted to the coefficients of some of these variables. The coefficients are calculated from tabulated atomic scattering factors and the corrections arising from anomalous dispersion, which may need to be evaluated experimentally.

There is a second equation formed from $|F_{\lambda\bar{\mathbf{h}}}|^2$. It is identical with equation 10 except for a minus sign in front of the last term, which contains $\sin(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}})$. This is so because $|F^n_{\bar{\mathbf{h}}}| = |F^n_{\mathbf{h}}|$ and $\phi^n_{\bar{\mathbf{h}}} = -\phi^n_{\mathbf{h}}$. There will therefore be two independent equations at each wavelength for which data are collected. A third equation, a quadratic, may be added to the two independent equations because of the trigonometric identity $\sin^2 x + \cos^2 x = 1$:

$$[|F^n_{1,\mathbf{h}}||F^n_{2,\mathbf{h}}|\cos(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}})]^2 + [|F^n_{1,\mathbf{h}}||F^n_{2,\mathbf{h}}|\sin(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}})]^2 = |F^n_{1,\mathbf{h}}|^2\,|F^n_{2,\mathbf{h}}|^2 \quad (11)$$

There are four unknown quantities in equation 10. If anomalous dispersion data are collected at two wavelengths, there will be five equations defining the variables: two of the form of equation 10 for each wavelength, generated by experimental values for $|F_{\lambda\mathbf{h}}|^2$ and $|F_{\lambda\bar{\mathbf{h}}}|^2$, and the quadratic equation 11. In practice, data may be collected at more than two wavelengths to increase accuracy and to make accessible those reflections **h** that show little difference between $|F_{\lambda\mathbf{h}}|^2$ and $|F_{\lambda\bar{\mathbf{h}}}|^2$ at some wavelengths.

Solution of the equations gives values for the quantities $|F^n_{1,\mathbf{h}}|$, $|F^n_{2,\mathbf{h}}|$ and $(\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}})$. If a value for $\phi^n_{1,\mathbf{h}}$ could be extracted and combined with $|F^n_{1,\mathbf{h}}|$ to form $F^n_{1,\mathbf{h}}$, it would be readily possible to compute equation 1 and obtain the structure of the macromolecule except for the anomalous scatterers. The structure of the anomalous scatterers is obtained from knowledge of $F^n_{2,\mathbf{h}}$. Actually, to proceed with the analysis it is necessary to determine the structure of the anomalous scatterers or, at least, the value of the phases $\phi^n_{2,\mathbf{h}}$. It is often possible to determine the structures of the anomalous scatterers from Patterson functions (equation 3) because such structures usually involve few atoms. Alternatively, direct phase determination could be performed based on the known values of $|F^n_{2,\mathbf{h}}|$. Once the structure of the anomalous scatterers is known, phases can be computed with equation 2. From the values of $\phi^n_{2,\mathbf{h}}$ thus determined, and the known values of $\phi^n_{1,\mathbf{h}} - \phi^n_{2,\mathbf{h}}$, the values of $\phi^n_{1,\mathbf{h}}$ can be obtained. The macromolecular structure can then be computed.

This technique for structure determination is called multiple-wavelength anomalous dispersion, or MAD.

Applications

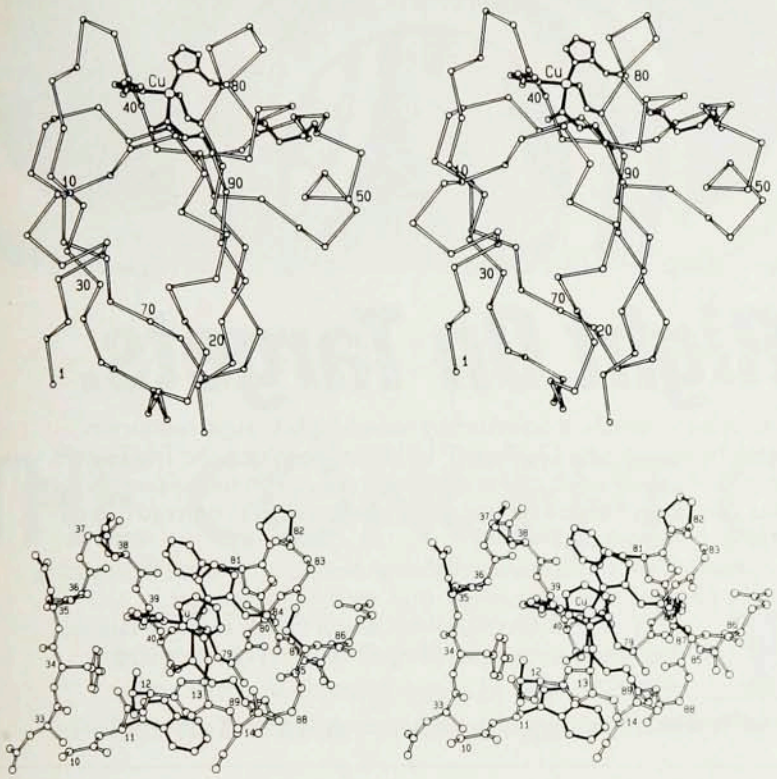
Hendrickson and his colleagues have used exact, linear equations to carry out several multiple-wavelength anomalous dispersion investigations of macromolecular structures. For example, they analyzed ferredoxin from *Clostridium acidi-urici* and streptavidin[1] from *Streptomyces avidinii*. Their analysis of ferredoxin was limited by the 5-Å resolution of their measurements. However, they found the phase values to be comparable to previously determined values for a related ferredoxin. For streptavidin, they took data at three wavelengths from a single crystal of a complex with selenobiotin. Selenobiotin is biotin in which the normally occurring sulfur atom has been replaced by selenium. Biotin adheres very strongly to avidin, and the selenium is the source of the anomalous scattering used in determining the structure. As a source it is superior to sulfur. The multiple-wavelength data led to a very good map at 3.3-Å resolution, which has been further refined with data at 2-Å resolution.

Figure 4 is a section of the electron density map for selenobiotinyl streptavidin at 3.3-Å resolution. In the normal course of events, one fits atomic models of the structure to the electron density distribution using knowledge of the peptide sequence in the protein. Then least-squares methods are applied to the initial model, a process in which agreement with the intensity data is improved by altering the atomic coordinates without sacrificing chemically sensible geometry. Figure 1 shows the final structure of streptavidin in the form of a tetramer.

Figure 5 shows the final structure of a protomer schematically. The structural features of the protomer, namely beta sheets, are pervasive and are symbolized by the wide arrows. This type of arrangement is called a beta barrel because the molecule is approximately barrel-shaped. The beta sheets are agglomerations of polypeptide chains held together by hydrogen bonds and forming a sheet-like structure that actually contains numerous puckers. Curiously, there are no alpha helices in the structure.

Another investigation applied the multiple-wavelength anomalous dispersion technique to determine the structure of a basic copper protein.[10] This experiment used as a source of the anomalous scattering an atom that is an inherent part of the protein structure, namely the copper atom. The opportunity to apply the multiple-wavelength technique to metalloproteins of small to medium size is especially helpful because attempts to make isomorphous or nonisomorphous heavier-atom derivatives of such structures are often unsuccessful.

Since the anomalous scattering source in metalloproteins may not be strong, it is worthwhile to optimize the experimental procedures. Thus the anomalous dispersion experiments on the basic copper protein were carried out at and near the absorption edge, where the corrections f' and f'' have extreme values and may change rapidly. Also, the experiments were done with intense, tunable synchrotron x-ray sources and multiwire electronic area detectors. These techniques are also quite appropriate for selenium-substituted proteins and were used for the analysis of the selenobiotin–streptavidin complex. In the determination of the structure of the basic copper protein, measurements were made with wavelengths at and in the vicinity of the copper Kα absorption edge. The structure has now been refined to 1.8-Å resolution. Figure 6 shows



Stereoviews of the copper basic protein. Such images allow one to see the special features of a protein structure in three dimensions. When viewed properly, the right eye's image of the figures on the right merges with the left eye's image of the corresponding figure on the left; some practice is required. The top stereoview shows one atom (the Cα atom) from each peptide unit of the polypeptide backbone, the side chains of the copper binding residues and the cystine disulfide bridge (formed from the two cysteine residues associated with the Cα atoms at the 52nd and 85th positions in the chain). The lower stereoview depicts the same protein, but shows the copper site and its environment in detail. (From reference 10; copyright 1988 AAAS.) **Figure 6**

aspects of the structure in stereoview, which is a good way to get clear views of complex structures and to facilitate understanding of the relationship of structure to function in macromolecules.

There have been additional investigations with the multiple-wavelength anomalous dispersion technique. In addition to other studies on macromolecules noted in reference 10, Gervais Chapuis and the Templetons[11] investigated the use of L edges in the multiple-wavelength method with crystals of NaHo(edta)·8H_2O and NaSm(edta)·8H_2O. Both studies were successful. Synchrotron radiation on the samarium compound gave especially high accuracy for the phases—the average uncertainty was 5°.

Future potential

Advances in theory and technique promise to extend the anomalous dispersion method in the future. The theoretical work suggests additional types of experiments and analytical procedures, and the technical advances make it possible to modify a variety of proteins to make them suitable for anomalous dispersion experiments.

As indicated above, in many instances selenium is a suitable atom for the application of the multiple-wavelength anomalous dispersion technique. Specifically, selenium-substituted biotin, when attached to streptavidin, afforded a selenium atom that was a sufficiently strong anomalous scatterer to permit determination of the structure. This experience led Hendrickson to suggest a procedure for producing a variety of selenium-substituted proteins.

A statistical survey indicates that about 1 in 59 of the amino acids that make up a protein is methionine, a sulfur-containing amino acid. The replacement of sulfur by selenium has several advantages, and Hendrickson's suggestion is aimed at creating proteins in which methionine is replaced by selenomethionine. His procedure takes advantage of the fact that the synthetic mechanism that incorporates methionine into proteins also works with selenomethionine. In brief, the procedure involves taking the DNA that codes for the protein of interest, along with attached promoter, and inserting it into a host cell deficient in methionine. The host cell is then cloned. Selenomethionine is added, the promoter system is stimulated, the repressor is removed, and selenomethionine is incorporated into proteins that need methionine.

This example gives an additional perspective on the potential of protein engineering. Not only can it provide suitable materials for structural characterization, but it can also make possible planned modifications that afford the opportunity for further clarification of the relation of structure to function.

References

1. W. A. Hendrickson, A. Pähler, J. L. Smith, Y. Satow, E. A. Merritt, R. P. Phizackerley, Proc. Natl. Acad. Sci. USA **86**, 2190 (1989).
2. J. Karle, Angew. Chem. (English ed.) **25**, 614 (1986).
3. A. L. Patterson, Z. Kristallogr. **90**, 517 (1935).
4. D. H. Templeton, L. K. Templeton, J. C. Phillips, K. O. Hodgson, Acta Crystallogr. A **36**, 436 (1980).
5. A. F. Peerdeman, A. J. van Bommel, J. M. Bijvoet, Proc. R. Soc. Amsterdam B **54**, 16 (1951).
6. A. J. van Bommel, Proc. R. Soc. Amsterdam B **56**, 268 (1953).
7. S. Ramaseshan, S. C. Abrahams, eds., *Anomalous Scattering*, Munksgaard, Copenhagen (1975).
8. W. A. Hendrickson, M. M. Teeter, Nature **290**, 107 (1981).
9. J. Karle, Int. J. Quantum Chem. **7**, 357 (1980).
10. J. M. Guss, E. A. Merritt, R. P. Phizackerley, B. Hedman, M. Murata, K. O. Hodgson, H. C. Freeman, Science **241**, 806 (1988).
11. G. Chapuis, D. H. Templeton, L. K. Templeton, Acta Crystallogr. A **41**, 274 (1985). ■