The Crystal-Solution Problem
of Sperm Whale Myoglobin

PETER URNES
From the Department of Chemistry, Harvard University, Cambridge, Massachusetts

ABSTRACT The central question to be discussed in this paper is whether the structure established for sperm whale myoglobin in the crystalline state is the same as that of the protein in solution. As judged by its ultraviolet optical rotatory dispersion, the helical content of metmyoglobin in solution does not differ from that in the crystal, 77 per cent. Although an uncertainty of about ±5 per cent must attach to this result, it excludes many alternative arrangements of the polypeptide chain. The folding of the chain may be further restricted to the basic form seen in the crystal if the dimensions of the molecule in solution and the interactions of specific chemical groups are taken into account. Since the rotatory dispersion of metmyoglobin is constant with respect to ionic strength, and since the dispersions of reduced and oxymyoglobin reveal no change in helical content upon their formation from metmyoglobin, one may infer that the structure of the protein is largely maintained both as it dissolves and during its reversible combination with oxygen. The crystallographic model of myoglobin thus offers a structural basis for attempting to explain its physiological function in solution. The relevance of this conclusion to the crystal-solution problems presented by other species of protein is then best seen in the light of common factors that govern the equilibrium of all proteins between crystal and solution.

INTRODUCTION

In the absence of any complete crystallographic model, the problem of whether the structure of a crystalline protein persists as it dissolves would not be pressing. Yet the importance of this issue as a matter of principle was recognized by Kendrew in 1949, well before the detailed structure of sperm whale myoglobin, the first protein to yield to analysis by x-ray diffraction, was at hand. At that time, in giving cautious support to a flattened disc model for crystalline horse myoglobin by appeal to its hydrodynamic properties in solution, he stated, "We cannot even exclude the possibility of a reversible change of configuration during the transition from crystal to solution. The x-ray crystallographer would fervently hope that this does not turn out to be the truth, since it would mean that his studies of protein crystals are of less interest to the biologist, who is mainly concerned with the state of the protein in solution" (Kendrew, 1949). Now that the structures of crystalline sperm...
whale myoglobin (Kendrew et al., 1960, 1961) and horse hemoglobin (Perutz et al., 1960; Cullis et al., 1962) have been very largely established, we would wish a clear answer to this uncertainty. Only then shall we know whether the use of these intricate structures is justified in attempting to explain the function of myoglobin and hemoglobin under physiological conditions. "The question most frequently asked," says Perutz, "is not whether the results of our crystal structure analyses are reliable, but whether they really represent the structure of the native molecules in solution" (Perutz, 1962).

Although there are several types of argument favoring the structural identity of any native protein in crystal and solution, perhaps the most convincing approach to myoglobin and hemoglobin is to compare specific features known to exist in their crystals with those that can be detected in solution. As a basis for this comparison, the crystallographic models themselves set structural hypotheses about the dissolved molecules, for example, their size and shape, helical content, and the availability to solvent of particular side chains. Of the two proteins, myoglobin is the more suitable for an initial test, for unlike hemoglobin, its structure is known almost to atomic resolution, while in many essentials it is one of the four subunits of the larger protein. The central topic in this discussion will therefore be the question of whether the structure of sperm whale myoglobin is the same in crystal and solution.

Before we examine the manner in which the structure of myoglobin in solution can be assessed, two related questions that will require subsequent attention may be asked. The first of these arises from the fact that nearly all of the crystallographic effort has focused upon metmyoglobin, that is, the stable form of the protein in which the iron is in the ferric state and is hence incapable of combining with oxygen. For this reason, it is natural that the principal studies in solution should also have been carried out upon metmyoglobin. However, given the biological significance of the problem, one must inquire into the structural similarity of this form to its physiologically active derivatives, reduced and oxymyoglobin, that participate in the reversible binding of oxygen and define the function of the protein in muscle as an oxygen carrier.

The second question is more comprehensive in scope. Once we have reviewed the specific case of myoglobin, we must ask whether there are any general lines of reasoning that apply as well to the crystal-solution problems presented by other species of protein, the functions of which take place chiefly in solution but the structures of which can be determined only by recourse to x-ray analysis of the crystalline state. In particular, is there any evidence that will let us approach the correspondence of crystal and solution in advance of having complete models for each of the many biologically active proteins that can be crystallized? For reasons that will become clear, a strictly struc-
ural comparison will in most instances prove less satisfactory than is the case for myoglobin. Some other manner of assuring structural continuity with solution may thus prove advantageous when new structures in fact come to be established.

THE STRUCTURAL CORRESPONDENCE OF MYOGLOBIN IN CRYSTAL AND SOLUTION

In the crystallographic model of sperm whale myoglobin, its polypeptide chain is packed into a space roughly $45 \times 35 \times 25$ Å in size (Kendrew, 1969). Of the 153 residues in the chain, 118 or 77 per cent are folded into eight right-handed $\alpha$-helical segments that vary in length from seven to twenty-four residues, while the remaining portions of the chain, although rigidly fixed, appear to be devoid of obvious long-range order. Nearly all side chains within this framework can now be identified, both by x-ray analysis and with the aid of the known chemical sequence along the chain (Edmondson and Hirs, 1962). In this way, the immediate environment of the heme group can be described with a high degree of precision. Unlike earlier attempts to delineate the molecule, this structure is founded solely on the independent physical basis of x-ray diffraction. Even as late as 1959, Bodo et al. (1959) cited two properties of myoglobin in solution, its optical rotation and hydrogen-deuterium exchange, as support for their working hypothesis that the chain in the crystal was largely folded into $\alpha$-helical rods. Now, however, it is possible to avoid the circularity that this procedure entails and to ask without equivocation whether the known crystalline structure persists in solution.

Wide-angle x-ray diffraction itself cannot establish continuity with the crystal, for diffraction patterns of randomly oriented protein molecules, both in solution (Echols and Anderegg, 1960) and in dry powdered specimens (Arndt and Riley, 1955), yield little information that can be used to discriminate one native protein from another. One must therefore fall back upon methods more appropriate for structural appraisal in solution, even though no one of these can provide a picture of the molecule in its entirety such as results from x-ray diffraction.

Given the primary covalent sequence of the chain, which almost certainly undergoes no chemical change in the transition from crystal to solution, the structural task for solution may be stated as one of ruling out alternative spatial arrangements of the chain that could conceivably exist. The size and shape of a macromolecule in solution can be estimated by hydrodynamic and light-scattering techniques (Edsall, 1953). Yet these methods alone establish nothing except the compactness with which the chain is folded. More discrete structural information can be obtained through various chemical techniques that measure the reactivity of specific side chains in the molecule and thereby permit one to estimate the number and site along the main chain of particular
groups available to solvent (Hermans, 1962; Breslow and Gurd, 1962; Banaszak et al., 1963). Although this count can then be matched with the position of these groups in the crystallographic model, reactivity in solution can locate a group only with respect to the surface of the protein and hence does little to define the course of the polypeptide chain. In contrast, if it can be demonstrated that the high helical content present in crystalline myoglobin persists in solution, then many possible orderings of the main chain are excluded by the preponderance of this rigid manner of folding. Conversely, if the amount of helix in solution proves to be significantly different, it follows that many interactions seen in the crystalline state must have been disrupted. As an initial single test for structural correspondence, the measurement of helical content thus offers a direct way to eliminate possible structures and at the same time provides a means for detecting any distinct change that may occur in the transition from crystal to solution.

Of the three phenomena in current use for detecting secondary structure in polypeptide chains—hydrogen-deuterium exchange, far ultraviolet hypochromism, and optical rotatory dispersion—it is the last that is the most sensitive reflection of helical content and is the best calibrated by model polymers in known conformation. Although the other two methods have been applied to myoglobin (Benson and Linderstrøm-Lang, 1958; Benson, 1959; Beychok et al., 1962; Rosenheck and Doty, 1961), several difficulties stand in the way of a clear interpretation. For example, deuterium slightly alters the coordinates of the α-helix (Tomita et al., 1962), the cumulative effect of which may decrease the stability of a protein. In the case of far ultraviolet hypochromism, correction for side chain absorbance and especially for that of the heme group is uncertain. One may therefore turn to optical rotatory dispersion for the most reliable estimate of helical content in myoglobin.

**Optical Rotation and Helical Content in Proteins**

The optical rotatory properties of polypeptides and proteins have recently been reviewed at length (Urnes and Doty, 1961). However, as background for the modified treatment to be described, it will be helpful to repeat here the essential argument by which helical content and optical rotation may be correlated.

We must first be able to compare the optical rotatory power of different polypeptide chains at a given wavelength. For this purpose, rotatory power is usually expressed as the reduced mean residue rotation, \([m']_\lambda\), a quantity which places measured rotations on a molar basis with respect to the number of amino acid residues and incorporates as well the Lorentz correction for refractive index of the solvent. It is defined by the formula

\[
[m']_\lambda = \left(\frac{3}{\pi^2 + 2}\right) \left(\frac{M_x}{d \cdot c}\right) \alpha^{obs}_\lambda
\]
in which $n$ is the refractive index of the solvent at wavelength $\lambda$, $M_r$ the mean residue weight of the polypeptide, $d$ the path length in decimeters, $c$ the concentration of solute in grams per 100 ml, and $\alpha_0$ the observed rotation. In this way differences among various specimens and solvents that are thought to be irrelevant to the inherent rotatory power of the chain are eliminated.

Optical rotatory dispersion, then, is simply the variation in optical rotatory power with wavelength. As a basis for measuring helical content in solution, it is founded on the fact that the helical and disordered forms of a polypeptide chain can generally be distinguished by their quite different rotatory attributes. For example, if the helix of poly-1-glutamic acid is formed by lowering the pH, its rotatory power at wavelengths above 300 m$\mu$ becomes more positive while the curvature of the dispersion becomes more pronounced (Fig. 1). A similar change in optical rotation, both in direction and curvature of the dispersion, takes place as the helix of poly-$\gamma$-benzyl-1-glutamate is formed in organic solvents.

That this rotatory change reflects the transition from a disordered chain to a helix rather than some other process is established, first, by alterations in hydrodynamic and light-scattering behavior which indicate the appearance of rod-like particles (Doty, 1957; Doty et al., 1954, 1956, 1957). Second, this conclusion is given support by the observation that this pattern of change is, for a majority of synthetic polypeptides, largely independent of side chain, solvent, and the manner in which the transition is brought about, whether this be through change in ionization of side chains, solvent polarity, or temperature (Urnes and Doty, 1961). Third, the analysis of recent measurements at low wavelength of circular dichroism (Holzwarth et al., 1962; Holzwarth, 1964; Holzwarth and Doty, 1965), polarized absorption spectra (Gratzer et al., 1961), and optical rotatory dispersion (Blout et al., 1962) demonstrates that these rotatory differences arise specifically from alteration in electronic transitions of the amide chromophore, the chemical group which all polypeptides have in common and the mutual orientation of which in a polymer defines the conformation of the backbone. There is thus good reason to believe that the difference in rotatory dispersion observed between a helix and its disordered form represents directly the change in conformational state of the main chain.

In order to exploit this phenomenon for the measurement of helical content, it must be described in some way that will permit the difference between a helix and a disordered chain to be given a numerical value. Although it is possible to use the change in optical rotatory power at individual wavelengths for this purpose, an expression for rotatory power as a function of wavelength embodies additional information that can be related to conformation. The rotatory dispersion of a typical polypeptide helix from 300 to 600 m$\mu$ can be represented by the Moffitt equation (Moffitt and Yang, 1956),
The constant $\lambda_0$ is chosen as that wavelength which will permit the data to be described in linear fashion when $[m']_\lambda (\lambda^2 - \lambda^2_0) / \lambda^2_0$ is plotted against $\lambda^2 /$ 

\[
[m']_\lambda = \frac{a_0 \lambda^2_0}{\lambda^2 - \lambda^2_0} + \frac{b_0 \lambda^2}{(\lambda^2 - \lambda^2_0)^3}
\]  

(Fig. 2). For helical polypeptides lacking strongly chromophoric side chains, $\lambda_0$ determined in this manner is found uniformly to be near 212 m\(\mu\) for data above 300 m\(\mu\). The values of $a_0$ and $b_0$ show variation with species of side 

**Figure 1.** Optical rotatory dispersions for two synthetic polypeptides and a globular protein. PBG, poly-\(\gamma\)-benzyl-\(\ell\)-glutamate. Helical structure in \(m\)-cresol, •; disordered chain in dichloroacetic acid, O (Yang and Doty, 1957). PGA, poly-\(\ell\)-glutamic acid. Helical structure at pH 4.72 in 1:2 water:dioxane mixture, 0.2 \(\mu\) NaCl, ▲; disordered chain at pH 6.56 in the same solvent, Δ (Doty et al., 1957). BSA, bovine serum albumin. Native protein in water at pH 5.47, ■; denatured protein in 8 \(\mu\) urea at pH 5.5, ○. (Dispersion calculated from values of [\(\alpha\)]$_{489}$ and $\lambda_c$ reported by Schellman, 1958 $\delta$.) (From Urnes and Doty, 1961.)
chain and solvent, but \( b_0 \) is generally in the vicinity of \(-600\) to \(-700\), whereas \( a_0 \) lies between \(-100\) and 0 in aqueous solution and between \(+100\) and \(+400\) in organic solvents.

In contrast to helices, disordered chains do not require as complex an expression as the Moffitt equation since the two constants of the simple Drude

\[
[M] = \left(\frac{\lambda_0^2 - \lambda^2}{\lambda_0^2 - \lambda^2 - \lambda_0^2}ight)
\]

\[
[\eta] = \left(\frac{\lambda_0^2 - \lambda^2}{\lambda_0^2 - \lambda^2 - \lambda_0^2}ight)
\]

Figure 2. Graphical treatment of dispersion data in Fig. 1 according to the Moffitt equation (2). The dispersions of the helical polypeptides, poly-\( \gamma \)-benzyl-t-glutamate (PBG, •) and poly-t-glutamic acid (PGA, △) are made linear, thereby permitting quantitative comparison with their disordered forms (PBG, ○; PGA, △) in terms of a slope and intercept. The plots for bovine serum albumin (BSA) are likewise linear both for the native protein (■) and the denatured molecule (□), so that its dispersion parameters can be placed on scales furnished by synthetic polypeptides and interpreted in terms of helical content (from Urnes, 1963).
suffice for the linear description of their dispersions. However, it is important to note that \( \lambda_e \) is generally near 200 m\(\mu\), while the value of \( \lambda_o \) is 212 m\(\mu\); it is hence necessary to recast the dispersion of disordered chains in Moffitt form in order to obtain parameters by which the two conformations may be directly compared. The value of \( b_0 \) is then typically found to be slightly positive, +50 to +80, while \( a_0 \) has shifted to a more negative range, centering near -200 for organic solvents and varying between -600 and -800 for aqueous media.

In the Moffitt parameters, then, we have a set of numbers for expressing quantitatively the features of rotatory dispersion that distinguish a polypeptide helix from its disordered chain. The positive change in intercept, \( a_0 \), may be thought of as representing the relatively positive rotation of the helix at high wavelength. The change in \( b_0 \) from a number near zero to a large negative value reflects the pronounced negative curvature in the dispersion of the helix and hence the need for the second term in the Moffitt equation. This term characteristically has a sharper dependence upon wavelength than the first, which is identical in form with the simple Drude equation. Both on formal grounds (Moffitt, 1956; Moffitt and Yang, 1956; Kauzmann, 1957, 1959) and now as a matter of observation at low wavelength (Holzwarth et al., 1962; Holzwarth and Doty, 1965), the curvature peculiar to the helix as well as its relatively positive rotation at high wavelength can be explained as the effect of opposing contributions from rotatory bands that are close together but different in sign (see Fig. 6). In contrast, the simple dispersion of the disordered form can be seen to arise from a single dominant rotatory band.

Before these differences in rotatory dispersion can be applied in measuring the helical content of proteins, we would wish some assurance that the synthetic polymers chosen as models are representative of proteins both in amino acid composition and in solvent conditions. We should therefore confine the species of amino acid residue to those that occur naturally and then study chains containing typical proportions of the principal classes of side chain, characterized by ionic, aliphatic, and aromatic groups. As solvents, aqueous media near physiological ionic strengths are more appropriate than many organic substances, for example, chloroform and trifluoroacetic acid, that are frequently used for synthetic polymers. Although the internal environment of a globular protein is probably provided largely by non-polar side chains, as seen in crystalline myoglobin (Kendrew, 1962), much of the course of the chain in myoglobin, including the major helical segments, is near the surface of the molecule and hence exposed to an aqueous environment.
Average values of $a_0$ and $b_0$ obtained for three synthetic polypeptides that are soluble in water and differ in content of L-glutamic acid, L-lysine, L-alanine, and L-tyrosine are listed in Table I, together with the change in each parameter that occurs upon forming a helix from a disordered chain. The superscripts $H$ and $D$ denote the helical and disordered forms, respectively. The three polymers upon which these values are based are poly-L-glutamic acid (PGA) in 0.25 M NaCl at pH values of 4.25 and 7.8 (Klemperer and Doty, unpublished results); a copolymer of 5 per cent L-tyrosine with L-glutamic acid (PTGA) in 0.1 M phosphate buffer at pH values of 4.0 and 7.0 (Urnes, 1963); and a copolymer of 30 per cent L-alanine, 28 per cent L-lysine, and 42 per cent L-glutamic acid (PALGA) in 0.15 M NaCl at pH 3.5 (Friedman and Doty, unpublished results). Rotatory measurements were made with a Rudolph 200S photoelectric polarimeter and data corrected for the dispersion of refractive index of water by the equation of Duclaux and Jeantet (Dorsey, 1940). All concentrations were measured by microKjeldahl determination for total nitrogen. Moffitt parameters were obtained by least squares treatment of data from 300 to 600 m$\mu$ with $\lambda_0$ set to 212 m$\mu$. While the helical parameters in Table I represent the averaging of the three polypeptides, those for disordered chains are based only on data for PGA and PTGA since PALGA does not appear to be completely unfolded in aqueous solution, even in 8 M urea (Friedman et al., 1962). With the reservation that these chains are only a first approximation to proteins in composition and solvent, the values of $[m']_{380}^{(H-D)}$, $a_0^{(H-D)}$, and $b_0^{(H-D)}$ may be taken as representing typical changes in optical rotatory dispersion that accompany the formation of a helix from its disordered chain. As can be seen from an examination of the literature (Urnes and Doty, 1961), these values are in substantial agreement with the results of many other studies.

Given these reference parameters, it remains to be shown that the rotatory properties observed for a polypeptide or protein of unknown conformation, for example, those illustrated for bovine serum albumin in Figs. 1 and 2, can be interpreted as reflecting a mixture of helical segments and disordered regions. If we assume that the observed rotation at any wavelength, $[m']_{380}$, is the sum of the rotations characteristic of the pure forms, $[m']_{380}^{(H)}$ and $[m']_{380}^{(D)}$, Table I

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\lambda_0 = 212$ m$\mu$</th>
<th>$[m']_{380}^{(H)}$</th>
<th>$[m']_{380}^{(D)}$</th>
<th>$[m']_{380}^{(H-D)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>-21°</td>
<td>-111°</td>
<td>+90°</td>
<td></td>
</tr>
<tr>
<td>$b_0$</td>
<td>-778</td>
<td>+734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_0^{(H)}$</td>
<td>-44</td>
<td>-778</td>
<td>+734</td>
<td></td>
</tr>
<tr>
<td>$b_0^{(H)}$</td>
<td>-630</td>
<td>+72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_0^{(D)}$</td>
<td>-778</td>
<td>+734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b_0^{(D)}$</td>
<td>-630</td>
<td>+72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_0^{(H-D)}$</td>
<td>+90°</td>
<td>+734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b_0^{(H-D)}$</td>
<td>-702</td>
<td>-702</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
each weighted by the fraction of residues in that conformation, \( f_D \) or \( f_H \), the following equation for a mixture may be written.

\[
[m']_\lambda^{obs} = f_D [m']_\lambda^D + f_H [m']_\lambda^H \tag{4}
\]

Substituting for \([m']_\lambda^D\) and \([m']_\lambda^H\) the Moffitt expressions to which each is equivalent and employing the relationship that \( f_D \) is equal to \((1 - f_H)\), one obtains an equation for the dispersion of the mixture.

\[
[m']_\lambda^{obs} = \frac{(a_0^D + f_H a_0^{H-D})\lambda_0^2}{\lambda^2 - \lambda_0^2} + \frac{(b_0^D + f_H b_0^{H-D})\lambda_0^4}{(\lambda^2 - \lambda_0^2)^2} \tag{5}
\]

This is an equation of Moffitt form in which the coefficients of each term may be equated with the intercept, \( a_0^{obs} \), and slope, \( b_0^{obs} \), that one expects to observe experimentally,

\[
a_0^{obs} = a_0^D + f_H a_0^{H-D} \tag{6}
\]
\[
b_0^{obs} = b_0^D + f_H b_0^{H-D} \tag{7}
\]

This formal derivation predicts that \( f_H \) will be a linear function of both \( a_0^{obs} \) and \( b_0^{obs} \) and further, given a set of reference parameters such as those listed in Table I, it stipulates how the helical content, \( f_H \), may be calculated. An analogous expression based upon \([m']_\lambda^{obs}\) may be derived from equation (4) and also used to compute \( f_H \).

\[
[m']_\lambda^{obs} = [m']_\lambda^D + f_H [m']_\lambda^{H-D} \tag{8}
\]

What evidence is there that this interpretation of the observed rotatory parameters is correct? First, the self-consistency of the method may be tested by following the parameters over the course of a helix-coil transition since equations (6) and (7) require that \( a_0^{obs} \) and \( b_0^{obs} \) vary together as the helical content changes. To a very good approximation, this is the result for the helix-coil transitions of poly-L-lysine (Applequist and Doty, 1962; Leonard and Foster, 1963) and poly-L-glutamic acid (Leonard and Foster, 1963). Second, the calculation of \( f_H \) by this means can be checked against an independent method for estimating helical content. In the case of copolymers of L-glutamic acid and L-lysine, there is a good correspondence between optical rotation and hydrogen-deuterium exchange (Blout and Idelson, 1958; Blout et al., 1961). In interpreting the rotatory properties of proteins, this same dual criterion of self-consistency and independent check is met, though with less precision than is found for synthetic polypeptides (Urnes and Doty, 1961). In general, \( b_0 \) appears to be a more reliable index of helical content than either \( a_0 \) or \([m']_\lambda\).

This simple pattern of analysis is appropriate for the conformational mixture.
seen in the crystalline structure of sperm whale myoglobin, which consists of helical segments connected by disordered stretches of chain. Unfortunately, the heme group itself is optically active, thus precluding any direct application of the Moffitt treatment above 300 mμ. Although an early estimate of 74 per cent helix was obtained from data at single wavelengths greater than the heme absorption bands (Urnes et al., 1961), this result cannot be given the same weight as an analysis of rotatory dispersion. It was therefore necessary to learn whether the rotatory dispersion of myoglobin at lower wavelengths would permit quantitative interpretation by comparison with synthetic polypeptides, thus yielding a more systematic calculation of its helical content in solution.

Changes in the Moffitt Parameters at Low Wavelength

At wavelengths less than 300 mμ, two problems became apparent. The first was the finding that rotatory artifacts can be easily produced at absorption bands by stray light and that they become more prominent in regions of low arc emission, as is the case with mercury and xenon arcs below 300 mμ (Urnes and Doty, 1961). In order to avoid artifact with the instrument then in use, it was necessary to work at absorbances of 2 or less and to ensure that the observed rotations of myoglobin were independent of the amount of absorbing material in the optical path.

The second problem is one of interpretation rather than experimental caution. If data for a typical polypeptide helix between 240 and 315 mμ are treated according to the Moffitt equation with λ0 set to 212 mμ, the resulting plot is found to be curved, so that no intercept or slope can be obtained for describing the dispersion. However, as shown in Fig. 3, this curvature in the case of the helical copolymer of glutamic acid, lysine, and alanine can be eliminated by increasing the value of λ0 to 220 mμ. This choice of λ0 is confirmed by computer analysis according to the statistical treatment of Sogami et al. (1963). For the helix of poly-L-glutamic acid, this same treatment again selects 220 mμ as the best value for λ0, while for the helical copolymer of 5 per cent tyrosine with glutamic acid, the best value for λ0 lies between 219 and 220 mμ. In every case, obvious curvature is now introduced above 315 mμ, thus indicating that no one value of λ0 will give a precisely linear plot for the entire range from 240 to 600 mμ. Both high and low wavelength data can be reasonably well accommodated by a compromise value of 216 mμ for λ0, as shown in Fig. 3. It is, however, only the more restricted part of the spectrum at low wavelength that is of concern here in setting up a pattern of analysis for myoglobin.

Since b0 becomes less negative as λ0 increases, a new set of Moffitt parameters must be computed for the low wavelength region. Disordered chains show little dependence upon λ0 for their linear representation and hence can be easily assimilated into a Moffitt equation with λ0 set to 220 mμ. In a manner
entirely analogous to the way in which the constants in Table I were obtained, averaging of the rotatory parameters for the same three synthetic models between 240 and 315 mµ leads to the values listed in Table II. Since the rotatory properties of these substances at higher wavelength are typical of a

![Graphical treatment according to the Moffitt equation (2) of the rotatory dispersion of a helical copolymer of L-alanine (30 per cent), L-lysine (28 per cent), and L-glutamic acid (42 per cent) (PALGA) in 0.15 M NaCl at pH 3.5. The wavelength range is 240 to 598 mµ. The lines represent plots at three values of λ₀, 212, 216, and 220 mµ, while points at 217, 218, and 219 mµ are also indicated. The line at a λ₀ of 216 mµ has been extrapolated from that fitting data in the wavelength range above 300 mµ (from Urnes, 1963; Urnes and Doty, 1965).](image-url)

**Figure 3.**

<table>
<thead>
<tr>
<th>MOFFITT REFERENCE PARAMETERS, 240-315 mµ</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ₀ = 220 mµ</td>
</tr>
<tr>
<td>( a_0^H )</td>
</tr>
<tr>
<td>( b_0^H )</td>
</tr>
<tr>
<td>( a_0^D )</td>
</tr>
<tr>
<td>( b_0^D )</td>
</tr>
<tr>
<td>( a_{H-D} )</td>
</tr>
<tr>
<td>( b_{H-D} )</td>
</tr>
<tr>
<td>-153</td>
</tr>
<tr>
<td>-363</td>
</tr>
<tr>
<td>-704</td>
</tr>
<tr>
<td>+64</td>
</tr>
<tr>
<td>+551</td>
</tr>
<tr>
<td>-427</td>
</tr>
</tbody>
</table>

wide range of synthetic polypeptides in known conformation, we may conclude that the rotatory differences observed at lower wavelength are likewise representative of the difference between a helix and its disordered chain. Given these new reference values, rotatory data for any protein between 240 and 315 mµ can be interpreted in exactly the same way as information obtained at more accessible wavelengths.
The Helical Content of Metmyoglobin in Solution

In setting out to compare the helical content of myoglobin in crystal and solution by means of optical rotatory dispersion, we must first ascertain whether the helical regions delineated in the crystallographic model are likely to have optical activity in proportion to their extent. Since the distinctive rotatory attributes of a helix arise from the mutual interaction of amide bonds, it is these groups rather than the amino acid residues themselves that define the physically relevant portions of the chain. From the standpoint of optical rotation, it is therefore significant that the conventions followed by Kendrew et al. (1961) give each helical segment the same number of amide bonds as residues, and that each amide group thus assigned is held in rigid helical array by at least one hydrogen bond. Provided that the appropriate weight factor, $M_{R}$, is used in treating the rotatory data, either chemical group can thus serve as a basis for predicting the helical content of myoglobin in solution. Since the rotations of myoglobin are corrected for its mean residue weight in conformity with the practice for all polypeptides, the crystalline count of 118 helical residues, that is, 77 per cent, is the value against which to judge the results of rotatory dispersion. In spite of the equivalence of amide bonds and residues in defining the secondary structure of this protein, the shortness of some of its helical segments may lead to an underestimate of helical content by optical methods. This additional complication will be assessed after an attempt has been made to compute the amount of helix in the native molecule in solution.

The rotatory dispersion between 224 and 440 m$\mu$ of sperm whale metmyoglobin in 0.1 M phosphate buffer at pH 7.0 is illustrated in Fig. 4. Rotatory data were obtained and treated in the same manner as described above for the synthetic polypeptides. Measurements of protein concentration were based on microKjeldahl determination for total nitrogen and the analytical data of Edmondson and Hirs (1962). Although the mean residue weight of the globin chain is 112.4, a value of 116.4 is appropriate to rotatory data for which concentration measurements include the presence of the heme group. Before the data are treated according to the Moffitt equation (2), they should be corrected for the contribution of the Cotton effect located near the Soret band at 409 m$\mu$ (Fig. 4). Given the wavelength, amplitude, and half-band width of this Cotton effect, rotations at other wavelengths which arise from this optically active transition can be calculated (Moscowitz, 1960). For myoglobin data between 244 and 302 m$\mu$, this correction is found to affect the final estimate of helical content by less than 2 per cent. Moffitt plots for the corrected rotations at several values of $\lambda_0$ (Fig. 5) show a striking similarity to those for a helical polypeptide (Fig. 3) in that a distinct curvature exists at a $\lambda_0$ of 212 m$\mu$, becomes less at 216 m$\mu$, and is finally eliminated when $\lambda_0$ is set to 220 m$\mu$. As in the case of the synthetic models, statistical
treatment of the dispersion by the method of Sogami et al. (1963) selects this wavelength as that best suited to make linear the Moffitt plot. This agreement between metmyoglobin and polypeptides known to be fully helical is itself an index of high helical content, for, as noted above, disordered chains are quite insensitive to changes in this parameter.

![Optical rotatory dispersion of sperm whale metmyoglobin, 224 to 440 mÅ.](image)

**Figure 4.** Optical rotatory dispersion of sperm whale metmyoglobin, 224 to 440 mÅ. Data for the native protein in 0.1 M phosphate buffer at pH 7.0 are indicated by a solid line, while the dispersion of the denatured form in 8 M urea, 0.1 M phosphate buffer at pH 7.0, is indicated by a dashed line (from Urnes, 1963; Urnes and Doty, 1965).

The value of \(\beta_0^{obs}\) at a \(\lambda_0\) setting of 220 mÅ for corrected data between 244 and 302 mÅ is \(-266 \pm 3\). When this value is placed on the scales given in Table II and interpreted according to equation (7), a helical content of 77 per cent is obtained. Although this is in precise agreement with the number of helical residues in the crystalline protein, an uncertainty of about \(\pm 5\) per cent must attach to this result in view of the variation in \(\beta_0^{pol}\) among the three helical polypeptides that here serve as conformational models for myoglobin. If poly-L-glutamic acid alone were taken as the reference, the helical content would drop to 71 per cent; given only the parameters of the copolymer containing tyrosine, it is 82 per cent. Although it is clear that the measurement of helical content in this protein will become more precise as the role of additional species of side chain is studied, a procedure based upon averaging the rotatory
parameters of synthetic polymers that are available seems to be the most reasonable course at the present time. For the denatured protein in 8 M urea, a \( b_0 \) value of \(-26 \pm 3\) at a \( \lambda_0 \) setting of 220 m\( \mu \) indicates a helical content of about 20 per cent.

![Graphical treatment according to the Moffitt equation (2) of the rotatory dispersion of native sperm whale metmyoglobin after correction for the contribution of the Cotton effect at the Soret band (Fig. 4). Three values of \( \lambda_0 \) are represented, 212, 216, and 220 m\( \mu \). The wavelength range is 236 to 365 m\( \mu \), and the solvent is 0.1 M phosphate buffer, pH 7.0 (from Urnes, 1963; Urnes and Doty, 1965).](image)

These findings are in accord with an earlier estimate of helical content in metmyoglobin in which the compromise value of \( \lambda_0 \), 216 m\( \mu \), and its associated parameters for a single model polypeptide were employed (Urnes et al., 1961). Using a \( \lambda_0 \) value of 212 m\( \mu \) over the same low wavelength range, Beychok and Blout (1961) report that the helical content based on \( b_0 \) is about 80 per cent; since their observed rotations are from 10 to 15 per cent smaller than those reported here, the agreement in result must be considered fortuitous. More recently, Samejima and Yang (1964) report \( b_0 \) values which indicate that their observed rotations are intermediate between these two other sets of measurements.
The value of \( a_{0}^{\alpha \beta} \) for native metmyoglobin, \(-23 \pm 8\), cannot be as readily interpreted as that of \( b_{0}^{\alpha \beta} \). It is more positive than the average \( a_{0}^{\alpha \beta} \) for the three helical models, \(-153\), but this finding agrees with other evidence that the myoglobin chain gives more positive rotations than do synthetic polypeptides. Since a relatively positive intercept persists in the denatured globin chain free of the heme group (Imahori and Doty, 1957), it may arise from some peculiarity in the amino acid composition of the chain. If it were certain that the globin chain was fully unfolded, its intercept could set the value of \( a_{0}^{\alpha \beta} \) for use in equation (6). At present, however, one can only say that according to the \( a_{0}^{\alpha \beta} \) scale in Table II, native metmyoglobin is about 70 per cent more helical than it is in 8 M urea, for which \( a_{0}^{\alpha \beta} \) is \(-409 \pm 10\).

A third rotatory parameter, the value of \([m']_{233}\) at 233 m\(\mu\), has been used to estimate helical content, for a sharp minimum characteristic of the helical conformation has been established at this wavelength (Simmons et al., 1961; Blout et al., 1962). Although the dispersion of myoglobin exhibits a distinct trough at 233 m\(\mu\) for which \([m']_{233}\) is about \(-9900^\circ\), the data are not sufficiently reliable to warrant a strict interpretation in terms of equation (8). There is, moreover, a considerable variation in the reference values that have been reported for fully helical polypeptides. If \([m']_{233}\) for poly-L-glutamic acid is \(-12,700^\circ\) while \([m']_{233}\) is \(-1800^\circ\) (Simmons et al., 1961), then the helical content of myoglobin is 75 per cent; if \([m']_{233}\) is instead \(-18,000^\circ\) (Yang and Samejima, 1963), then the helical content of the protein comes to about 50 per cent. The extent to which rotatory artifact and different experimental procedures contribute to this variation is not clear. Further, the fact that myoglobin rotations appear to be relatively positive, as discussed above, might make rotations at single wavelengths difficult to compare with standard synthetic polypeptides. Despite the uncertainty in interpretation of both \(a_{0}\) and \([m']_{233}\), their values give some support to the more firm conclusion based on \(b_{0}\) that the helical content of metmyoglobin in solution is high.

Important corroboration of this finding is provided by recent measurements of circular dichroism at low wavelength (Holzwarth and Doty, 1965). This phenomenon, which is the difference in absorption of right and left circularly polarized light, is the correlate of rotatory dispersion and hence can serve as an independent measure of optical activity. As illustrated for the helical copolymer of glutamic acid, lysine, and alanine in Fig. 6, helical polypeptides exhibit a band of positive dichroism near 192 m\(\mu\) and two closely placed negative bands at 206 and 222 m\(\mu\). Since the dichroism of disordered chains is radically different, having a slight positive region near 220 m\(\mu\) and a large negative band near 200 m\(\mu\), the observed difference can be used as a basis for estimating helical content in proteins. Although the data are not as precise as those for rotatory dispersion, the intensity of the circular dichroism in metmyoglobin averages to about 70 per cent of that found for fully helical chains.
Further, the strong similarity in dichroic pattern between metmyoglobin and helical polypeptides is good evidence that the optical activity of the protein originates largely with the amide groups of the backbone and not with particular side chains or the heme group. Measurements of circular dichroism thus strengthen the interpretation of the rotatory dispersion of myoglobin at higher wavelengths in terms of helical content.

One may conclude from these various measures that no detectable change in helical content occurs as crystalline myoglobin enters solution. Additional evidence for this conclusion is given by the fact that the rotatory dispersion from 240 to 313 m\(\mu\) is constant with respect to ionic strength, showing little alteration on going from distilled water to 2.8 m phosphate buffer, that is, to a concentration of salt at which crystals begin to form.

Before we attempt to use this conclusion in further correlation of the structure of myoglobin in crystal and solution, it is worth noting that it suggests answers to three questions concerning the structure and optical activity of helical polypeptides, thus in this respect reversing the role which they play.

![Graph showing circular dichroism](image-url)

**Figure 6.** Circular dichroism of a helical synthetic polypeptide and native sperm whale metmyoglobin. The polypeptide is a copolymer of L-glutamic acid (42 per cent), L-lysine (28 per cent), and L-alanine (30 per cent) in 0.1 M NaF at pH 3.1. The protein is in 0.02 M phosphate buffer at pH 7.0 (from Holzwarth and Doty, 1965).
as simple models for the more complicated protein. The first is the question of whether standard polypeptide helices consisting of L-amino acids are right- or left-handed (Urnes et al., 1961). Since the optical rotatory dispersions of both this protein and standard polypeptides have the same sign, becoming more negative at lower wavelength, their helices have the same sense. That this sense is almost certainly right- rather than left-handed follows from the existence of right-handed helices in crystalline myoglobin together with the persistence of high helical content in solution, for it is most unlikely that the chain would unfold and then form 77 per cent left-handed helix upon dissolving. The second issue, which is discussed by Holzwarth and Doty (1965), is whether standard helical polypeptides have the coordinates of the \( \alpha \)-helix or those of the more elongated \( 3_{10} \) helix, as suggested by Luzzati et al. (1961). Again it is unlikely that myoglobin would retain its high helical content in solution and yet shift each helical segment into the \( 3_{10} \) conformation. Then, on grounds of the similarity of electronic energies and band widths that is revealed by circular dichroism, it is possible to argue that standard helical polypeptides are specifically \( \alpha \)-helical in structure. The third question is that of the extent to which helical content in very short helices is underestimated by optical methods, that is, the difficult problem of end effects. Since there appears to be no serious underestimate of helical content in myoglobin, even though three of its helical segments in the crystal have only seven to nine residues, the results suggest that end effects are small. Although it is conceivable that myoglobin in solution actually contains more than 77 per cent helix, thus allowing more room for end effects, this possibility is lessened by the fact that the globin chain of horse myoglobin is about 80 per cent helical in 2-chloroethanol, a helix-promoting solvent (Imahori and Doty, 1957), so that the native chain may already be near the maximum helical content permitted by its amino acid composition.

**Correlation of Structure in Crystal and Solution**

Let us now return to the problem of whether the structure of metmyoglobin in solution is the same as that in the crystal. Although the finding that 77 per cent of the chain is helically folded eliminates many possible arrangements of the polypeptide backbone, optical methods are blind to the actual number, distribution, and mutual orientation of helical segments along the chain. Is there any way, then, in which one can limit the high helical content observed in solution to the eight segments seen in the crystallographic model?

One argument hangs on the way in which individual amino acid residues are known to determine the course of the chain. Specifically, four of the helical segments in the model are terminated by proline, the steric properties of which prevent the continuation of helix beyond these points. Thus, if the net helical content of the crystalline protein should become redistributed as the molecule
enters solution, it is safe to say that the existing helical segments cannot grow beyond these four prolyl residues. A related point may be made by referring again to the fact that the globin chain of horse myoglobin appears to become only about 80 per cent helical in a helix-promoting solvent. If this value represents the maximum helix that this particular sequence of amino acids will permit, it then suggests that the only change possible is a loss of helix rather than an increase or redistribution of the segments. From this it would follow that the helical segments of the native molecule in solution are in fact confined to the same stretches of chain as in the crystal.

A better way of attempting to limit the helical segments existing in solution to those present in the crystalline protein is based upon the identity of heme spectra in both crystal and solution. In the course of measuring ratios of absorbance along different axes of single crystals of horse and human hemoglobin and horse myoglobin, Perutz (1953 a, b) has established that the wavelengths of visible absorption bands match exactly in crystal and solution and that the intensities agree to within the error of measuring crystal thickness. This correspondence is found for reduced, oxygenated, and methemoglobin, reduced and metmyoglobin, and the azide and imidazole derivatives of methemoglobin. For crystals of finback whale metmyoglobin, Kendrew and Pauling (1956) report measurements of dichroic ratio at absorption bands at 500 and 633 mμ, wavelengths which are characteristic of metmyoglobins in solution. Although no experimental details are given, Kendrew and Parrish (1957) cite dichroic ratios for the reduced and met-forms of crystalline sperm whale myoglobin.

That visible absorption spectra are sensitive indicators of the structural surroundings of the heme group is shown both by the radical changes observed with denaturation and by the more subtle differences between myoglobin and hemoglobin, which can be plausibly traced to the slight variation among species of side chain that provide the immediate environment of the prosthetic group. As illustrated by Kendrew (1962), a very high degree of homology exists among the eleven side chains that are in direct contact with the heme group in the crystalline structures of myoglobin and both the α- and β-chains of hemoglobin, a fact that probably reflects the precise functional requirements placed upon heme in these proteins. If, now, we agree that the identity of heme spectra in both crystal and solution is evidence that the immediate environment of this planar ring does not differ in the two states, then the relative positions of these eleven groups are fixed. Since five of the eight helical segments in myoglobin contribute side chains to this environment, the requirement that these particular residues be fixed in space places a great restriction on how the high helical content observed in solution must be distributed along the chain and oriented with respect to the heme group. In fact, upon examination of the crystallographic model of myoglobin, it would
appear that there are few ways in which one can unfold the molecule and yet maintain both the heme contacts and the high helical content.

Two other types of evidence for eliminating structural alternatives in myoglobin may be mentioned. One depends upon the reactivity of specific side chains, as has already been cited. For example, spectroscopic titration of tyrosyl groups in carboxymyoglobin indicates that two are free to ionize in the usual manner while the third is not (Hermans, 1962). In the crystallographic model of metmyoglobin the phenolic hydrogen of tyrosine H 22 in fact forms a bond with a carbonyl group of the main chain and provides a bridge between the G and H helices. If one is permitted to identify tyrosine H 22 with the unreactive residue in solution, then it can serve to limit further the spatial arrangement of helical regions. The finding of Gurd and his co-workers (Breslow and Gurd, 1962; Banaszak et al., 1963) that about half of the twelve histidines and all the nineteen lysines are exposed to solvent while neither of the two methionines can react is again in excellent correspondence with the crystallographic model.

A different kind of structural assessment is provided by measurements of size and shape in solution. The available data on the sedimentation, viscosity, and dielectric behavior of myoglobin can be interpreted according to the treatments of Oncley (1941) and of Scheraga and Mandelkern (1953). Although the data are compatible with a particle having an axial ratio of about 1.6 and a hydration of 0.4 gm per gm protein, they do little more than rule out any gross unfolding of the chain. They would, however, help determine the disposition of helical portions of the chain that are not in actual contact with the heme group, for example, the A and B helices, which probably could not float freely from the rest of the molecule without affecting its hydrodynamic properties. One last piece of evidence against any large unfolding of the chain outside the crystalline lattice is provided by electron micrographs of single myoglobin molecules, which have dimensions in good agreement with those of the crystalline protein (Levin, 1963; Deutsch, 1963).

It is important to note that studies in which myoglobin is denatured by urea, extremes of pH, the binding of metal ions, or removal of the heme group can serve as controls for the investigation of the native molecule in solution. They in fact indicate that each of the methods employed is capable of detecting structural change, in this case that which is brought about intentionally. Denaturation also gives additional evidence that the native structure in solution does not depart widely from the compact crystalline form, for it is possible to achieve a substantial unfolding of the chain by the deliberate use of non-physiological agencies.

Thus, to summarize the structural correspondence of metmyoglobin in crystal and solution, a great many alternative arrangements of the chain in solution can be ruled out by its high helical content together with the postulate
that the immediate environment of the heme group remains the same in both phases. Other information concerned with the reactivity of specific groups and the size and shape of the molecule serves to restrict further the range of structural possibilities. Indeed it is fair to say that none of the evidence at hand gives any indication of a definite difference in structure between crystal and solution.

Let us now turn briefly to the two remaining questions with which this discussion opened, one concerning the physiologically active forms of myoglobin and the other dealing with more general arguments that bear on the crystal-solution problem.

**THE OPTICAL ROTATORY DISPERSION OF REDUCED AND OXYMYOGLOBIN**

Solutions of reduced myoglobin suitable for spectroscopic study at low wavelength may be obtained by reduction of metmyoglobin with ferrous pyrophosphate in the absence of oxygen and then removal of the reductant, which absorbs in the ultraviolet spectrum, by passage of the protein across a sephadex column, again in the absence of oxygen. The product is sufficiently stable in 0.1 M phosphate buffer at neutrality for measurements of both optical rotation and absorption lasting some hours to be carried out. The same solution can then be rapidly oxygenated to produce oxymyoglobin without change of protein or cell, thereby permitting a relatively precise comparison to be made. The extent of reduction and oxygenation may be judged from visible absorption spectra. As metmyoglobin is again formed by autooxidation, its optical properties can also be directly compared under these favorable conditions.

When ultraviolet spectra of the three forms are compared, distinct differences are observed in the vicinity of aromatic absorption near 280 m\(\mu\). Of more immediate importance, however, are measurements of rotatory dispersion from 240 to 313 m\(\mu\), which are illustrated in Fig. 7. Although differences among the three forms are observed at higher wavelengths, the rotatory power at shorter wavelengths, where it is most sensitive to helical content, remains unchanged, as does the basic pattern of the entire dispersion. The small changes in the region of aromatic absorption occur near the cross-over point between the dispersions of the native and denatured forms (Fig. 4) and hence very probably do not directly represent an alteration in helical content. More recent measurements of circular dichroism assign these changes to transitions in the region from 260 to 290 m\(\mu\) rather than to those of the helical backbone at lower wavelength. One may thus conclude from the identity of their rotatory dispersions at low wavelength that the helical content of reduced, oxy- and metmyoglobin in solution is the same. Samejima and Yang (1964) reach a similar conclusion for these and other derivatives of this protein.
For those functional forms of myoglobin which have been compared in crystal and solution, the visible absorption spectra are the same. Therefore a principal piece of evidence for limiting the high helical content in solution to specific regions of the chain is at hand. The x-ray diffraction patterns of reduced and oxymyoglobin have been reported to be essentially the same as that of metmyoglobin (Perutz, 1960, 1962), so that only minor differences among the crystallographic models may be expected to emerge. Although

---

**Figure 7.** Optical rotatory dispersion of reduced and oxymyoglobin of sperm whale measured upon a single sample of the protein, 0.066 per cent in 0.1 M phosphate buffer at pH 7.0 in a path of 1 cm. These forms were obtained by reduction of metmyoglobin with ferrous pyrophosphate, separation of the protein from the reducing agent by sephadex filtration, and then oxygenation of the reduced myoglobin. The rotatory dispersion of oxymyoglobin after partial oxidation to metmyoglobin is also illustrated. The dashed line indicates the dispersion of fully oxidized metmyoglobin as taken from the data represented in Fig. 4 (from Urnes, 1963; Urnes and Doty, 1965).
the structural study of reduced and oxymyoglobin has only begun, it is possible to argue that no fundamental rearrangement of the molecule occurs, either upon dissolving, reduction, or reversible combination with oxygen. The crystallographic model of myoglobin now available can thus be used in attempts to account for its physiological activity in solution.

**THE CRYSTAL-SOLUTION PROBLEM OF OTHER PROTEINS**

The crystal-solution problems presented by other species of protein will be more difficult to undertake in a strictly structural manner than is that of myoglobin, and one might add, hemoglobin, for which a parallel case can almost certainly be made. As judged by optical rotatory dispersion, the helical content of most globular proteins is considerably less than in these heme proteins, and many further lack a prosthetic group with its known structure, numerous contacts, and characteristic spectra. Existing methods for assessing protein structure in solution may hence be unable to eliminate as many alternative arrangements of the main chain. Is there, then, any way in which we can judge the persistence of a crystalline structure in the dissolved state without being dependent upon structural information in solution?

**Functional Correspondence**

One approach to the crystal-solution problem of a biologically active protein that does not hang on a knowledge of its structure is to compare its functional capacity in crystal and in solution. As excellent evidence of this type, the measurements of heme absorption by single crystals of myoglobin and hemoglobin in various functional states may again be cited. In the case of an enzyme, crystalline ribonuclease will hydrolyze uridine 2', 3' cyclic phosphate, a compound related to its natural substrate (Doscher and Richards, 1963). Critical though this type of evidence may be, a failure to obtain reactivity would not necessarily indicate a difference in structure between crystal and solution, for one can easily imagine that some substrates are too large to enter the crystalline lattice or that the active site is sequestered at some point by a neighboring protein molecule. Tests of the most sensitive kind, such as the kinetics of reaction, may be difficult to interpret or impossible to follow because of a conformational change that leads to disintegration of the crystals or slow lattice changes, as is the case with hemoglobin (Perutz, 1953 a; Perutz et al., 1964). If an identity in function for both crystal and solution can be established, it does imply that the structure of the active site is the same in the two phases. However, it implicates the rest of the molecule only by inference. As in papain, in which catalytic activity persists in only one-third of the chain (Smith, 1957), this inference may be weak, thus allowing for the possibility that parts of the structure unessential to function change in the transition from
crystal to solution. For these reasons, despite their specificity, tests of functional correspondence enjoy no special advantage over strictly structural methods in facing the crystal-solution problem.

The Mechanism of Crystal Formation

It is rather upon the mechanism by which protein crystals are formed that we may focus, for it appears that fundamentally the same factors govern the equilibrium of all proteins between crystal and solution. Protein crystals are generally obtained from solution by salting out, that is, by increasing the concentration of added electrolyte, usually ammonium sulfate or phosphate buffer. As illustrated in Fig. 8 for a variety of proteins, including myoglobin and hemoglobin, the solubility of a protein decreases logarithmically as the ionic strength of the solution is raised. This regular dependence of solubility upon ionic strength is also characteristic of the crystallization of amino acids, a similarity which has been analyzed in detail by Cohn and Edsall (1943) and

\[
\log S = \log S_i - K_s \omega
\]

in which \(S_i\) is the hypothetical solubility extrapolated to zero ionic strength and \(K_s\) is the slope of the plot. For a given combination of protein and precipitating salt, \(S_i\) is dependent upon temperature and pH whereas \(K_s\) generally remains independent of these variables (from Cohn and Edsall, 1943, p. 602; and Edsall and Wyman, 1958, p. 276).
by Edsall and Wyman (1958) and which is strikingly demonstrated in studies upon carboxyhemoglobin and cystine. In principle, the salting out of a polar solute, which is a phenomenon by no means restricted to proteins and amino acids, can be explained by a transfer of water from the solute to the added ions, with a consequent tendency for the dehydrated polar molecules to interact with one another. In the theories of Debye (1927) and Debye and McAulay (1925), this shifting of the equilibrium between crystal and solution proceeds by essentially electrostatic mechanisms. What, therefore, is the evidence that dehydration of polar groups is the critical structural alteration that occurs in the crystallization of both proteins and amino acids?

If we examine the bonds actually formed in crystals of sperm whale myoglobin, x-ray diffraction reveals that each molecule within the ammonium sulfate crystal makes six polar linkages with its neighbors (Kendrew, 1962). These include several interactions of carboxylate groups of glutamic acid with ammonium groups of lysine, one of a guanidinium group of arginine with two carbonyl groups of a neighboring protein molecule, and one in which a sulfate ion forms a tetrahedral complex with four imino groups, two from each adjacent molecule. That these specific groups are almost certainly hydrated in solution can be argued by the fact that all other polar groups on the surface of crystalline myoglobin, including these same species of side chain, are seen to bind water. Therefore the simplest picture of forming the myoglobin crystal is that these particular groups exchange water for polar side chains of molecules already incorporated into the crystalline lattice. Although the intermolecular bonds formed in crystals of other proteins have not been identified, the basic architecture of protein crystals is the same. From both the penetrability of crystals to small molecules (Adair and Adair, 1936; McMeekin and Warner, 1942; Low and Richards, 1954) and x-ray analysis of the space available to added salt (Perutz, 1946; Bragg and Perutz, 1952), one may infer that much of the surface of a crystalline protein is in contact with solvent and that it is highly hydrated. Further, the electrophoretic and titrimetric properties of native proteins indicate that charged groups like those of myoglobin exist at their surfaces.

The intermolecular bonds formed in crystals of amino acids have been established by x-ray analysis. As described for a variety of amino acids by Low (1953), they involve precisely the type of oriented hydrogen bond seen in myoglobin. Since amino acids are dipolar ions and hence are hydrated in solution (Cohn and Edsall, 1943), it follows from their crystal structures that individual molecules give up water upon entering the crystal lattice. Given the similarity of amino acids and proteins in salting out behavior, the identity of crystalline bonds formed, and the occurrence of dehydration in this process, both as a matter of theory and of structural necessity, it is possible to argue that the common structural event to accompany crystallization in both classes of molecule is the loss of water from polar groups which then form inter-
molecular hydrogen bonds. From this aspect, a protein molecule is simply a bulky amino acid in which the critical polar groups are widely spaced on its surface. Is there evidence, however, that no change beyond that held in common with an amino acid takes place as a protein crystallizes?

Thermodynamic support for an identity in mechanism of crystal formation in both amino acids and proteins is given by the similarity in their heats of solution. As tabulated by Cohn and Edsall (1943), $\Delta H$ values for amino acids are almost all positive but range from $-4$ kcal/mole for lysine to about $+6$ kcal/mole for aspartic acid, while those for proteins vary from $-18$ kcal/mole for carboxyhemoglobin to $+7.8$ kcal/mole for casein. In contrast, $\Delta H$ for the denaturation of chymotrypsin, a process in which the polypeptide chain is largely unfolded, is estimated by Schellman (1958 c) to be $+110$ kcal/mole, while that for $\beta$-lactoglobulin is about $+600$ kcal/mole (Schellman, 1958 a).

In the crystallization of proteins, then, one may not only infer that specific polar groups at the surface exchange water for intermolecular hydrogen bonds, as argued above, but also suggest on analogy with amino acids that this process alone is sufficient to account for the observed thermodynamic properties. From this it would follow that the internal structure of a protein is unaltered in the reversible transition between crystal and solution. Although the observations are too diverse to summarize here, there is in fact a substantial body of crystallographic evidence to show that the internal structure of a protein is unaffected by factors that determine the equilibrium between crystal and solution, that is, ionic species and strength, extent and site of dehydration, and the identity of particular side chains exposed at the protein surface.

As a general approach to the crystal-solution problem, one that is independent of structural assessment in solution, this argument hangs on the fundamental similarity in the salting out of all proteins. Since salting out implicitly stipulates a condition of equilibrium between crystal and solution, conclusions based upon an analysis of its mechanism do not necessarily apply to solutions that are more typical of a physiological environment. Once a protein molecule is in solution, however, even under the restricted condition of equilibrium, we have at our disposal a battery of techniques for detecting whether structural change occurs as physiological circumstances are more closely approached. For justifying the use of the crystalline structure of any protein in attempting to explain its function in solution, it is perhaps sufficient if salting out can ensure structural continuity solely in the first and reversible step from crystal to solution.

**CONCLUSION**

In closing, let us return to the problem with which we began, that of sperm whale myoglobin. As we have seen, two separate lines of evidence indicate
that the crystalline structure of this protein persists as it dissolves. One is based on the high degree of structural correspondence that can be demonstrated between specific features of the crystallographic model and properties of the protein in solution. Given an agreement in helical content of the chain and an identity of heme spectra in crystal and solution, together with ancillary measures of reactive side chains and over-all size and shape, it is possible to conclude that the structure of metmyoglobin in solution is very largely the same as that in the crystalline state. This structure is, furthermore, of physiological significance since reduced and oxymyoglobin do not appear to differ from it, either in crystal or in solution.

The other, more general, line of evidence is concerned with an analysis of the mechanism by which protein crystals are formed. Beginning with the similarity in salting out behavior of both proteins and amino acids, it points to the plausible conclusion that no change other than the hydration of polar groups at the surface of the molecule is involved in the reversible transition from crystal to solution. At equilibrium between crystal and solution, one may thus expect the internal structure of protein molecules to be the same, not only for myoglobin but for other species of protein as well.

The valuable comments of Professor Paul Doty on this manuscript and his generous support of the work reported in it are gratefully acknowledged.

REFERENCES

Benson, E. S., and Lindström-Lang, K., 1958, Biochim. et Biophysica Acta, 22, 579.
DEBYE, P., and McAuLay, J., 1925, Physik. Z., 26, 22.
KAUZMANN, W., 1959, Advances Protein Chem., 14, 1.
Urnès, P., and Doty, P., 1965, data to be published.