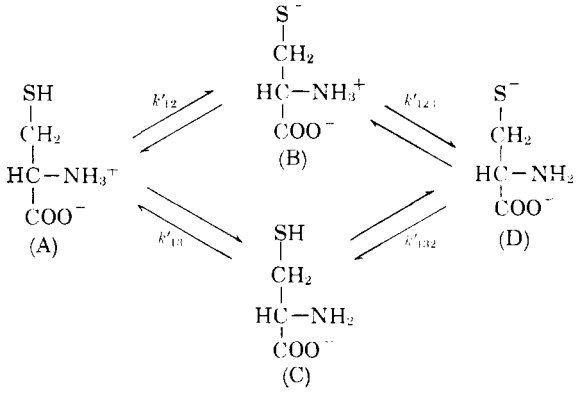
The Ionization of Cysteine and the Possible Role of Hydrogen Bonding by the Thiol Group*

John T. Edsall

ABSTRACT: The ionization of the thiol and amino groups in cysteine requires for its description a formulation of equilibria between at least four microscopically distinct forms of the amino acid. A fifth form, involving a hydrogen bond between the un-ionized thiol and the uncharged amino group, has also been proposed as an intermediate in ionization. After consideration of general structural evidence, and of the major features and the *p*H dependence of the ultraviolet and Raman spectra of cysteine, it is concluded that the concentration of this hydrogen-bonded intermediate is so low that it is virtually undetectable.

Macroscopic and Microscopic Ionization Constants

The calculation of the ionization constants of the amino and the thiol groups in cysteine, from the experimental titration data, has given rise to considerable controversy. A good brief summary is given in the recent review by Cecil (1963). A direct *p*H titration of isoelectric cysteine with alkali yields two "macroscopic" pK values: $pK'_2 = 8.30$, $pK'_3 = 10.40$ (ionic strength 0.15, 25°), but it is an oversimplification to assign pK'_2 to one group and pK'_3 to the other. Several investigators (for instance, Grafius and Neilands, 1955; Edsall and Wyman, 1958) have formulated their results in terms of a set of microscopic constants,[1] according to Scheme I.



SCHEME I

The last numeral in the subscript of each k' value denotes the ionizing group; the preceding numerals denote the acidic groups that have already lost their protons before the ionization in question takes place. According to the convention adopted by Edsall and Wyman (1958) the carboxyl group is denoted by subscript 1, the —SH group by 2, and the NH_3^+ group by 3.

The microscopic constants are related to the aforementioned macroscopic pK values by the equations:

$$K'_2 = k'_{12} + k'_{13}$$

$$1/K'_3 = 1/k'_{123} + 1/k'_{132}$$

Benesch and Benesch (1955) calculated the four microscopic constants, k'_{12}, k'_{13}, k'_{123}, and k'_{132} from the variation with *p*H of the ultraviolet absorption of cysteine in the region between 230 and 240 mμ. They assumed that the intensity of absorption in this region of the spectrum was determined by the concentration of ionized thiol groups in the solution, that is, by the sum of the concentrations of species B and D in Scheme I, and deduced the following pK values for cysteine: $pk'_{12} = 8.53$, $pk'_{13} = 8.86$, $pk'_{123} = 10.36$, and $pk'_{132} = 10.03$. According to this scheme, therefore, the thiol and the charged amino group in cysteine are acids of comparable strength, although the thiol group is slightly the stronger of the two, since $k'_{12} \cong 2k'_{13}$. The acidic strength of either group is of course profoundly affected by the presence or absence of an acidic proton on the other neighboring group, as the above pk' values show. Microscopic pk values for the amino and thiol groups of glutathione have been reported by Martin and Edsall (1958).

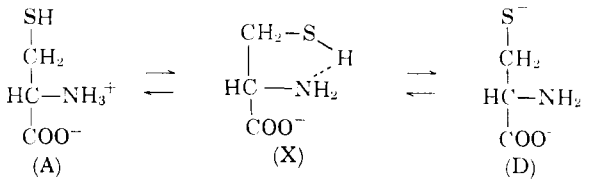
Garfinkel and Edsall (1958) and Elson and Edsall (1962) studied the Raman spectrum of cysteine as a function of *p*H, and interpreted the results in terms of Scheme I. From the variation of the intensity of the S—H stretching frequency at 2580 cm^{-1}, as a function of *p*H, Elson and Edsall (1962) derived a set of micro-

* From the Biological Laboratories, Harvard University, Cambridge, Mass. *Received October 12, 1964.* The research that provided the background for the discussion given in this paper was supported by grants from the National Science Foundation (GB-1255) and the U.S. Public Health Service (HE-03169).

[1] Scheme I involves only the microscopic forms in which the carboxyl group is ionized. A scheme for the ionization of the other four microscopic forms, in which the carboxyl group is un-ionized, has been presented by Edsall and Wyman (1958, p. 502). However, the latter forms are present in negligible amounts, and may be ignored in the present discussion.

scopic constants close to those of Benesch and Benesch (1955).

In 1956 De Deken *et al.* measured the absorption spectra of cysteine and other thiol compounds, as a function of *p*H, with experimental results very similar to those of Benesch and Benesch (1955). Their interpretation of the data, however, was radically different. They proposed the following scheme for the ionization of the thiol and ammonium groups in cysteine:



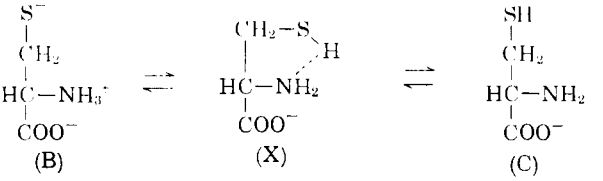
SCHEME II

According to Scheme II, the initial addition of OH^- ions to the solution of isoelectric cysteine removes a proton from the $-NH_3^+$ group, and the thiol group then forms a hydrogen bond with the resulting uncharged amino group. Only after the formation of this hydrogen-bonded intermediate (X) is completed does the thiol group then proceed to ionize, on the addition of more OH^- ions, according to Scheme II. In other words, this scheme identifies the transition from A to X as corresponding to the macroscopic constant pK'_2, and the transition from X to D as corresponding to the constant pK'_3.

The argument of De Deken *et al.* (1956) in favor of this scheme is primarily based on the fact that the maximum in the ultraviolet-absorption spectrum of ionized cysteine becomes displaced to longer wavelengths as *p*H increases. This maximum lies close to 230 mμ at *p*H near 7 and shifts gradually upward to about 240 mμ as the *p*H rises to values near 12. In contrast, the ultraviolet absorption of simple ionized thiols, such as 2-propanethiol, shows an absorption maximum near 240 mμ; its position is independent of *p*H and its intensity simply increases progressively as the S—H group ionizes. De Deken *et al.* (1956) also showed that the absorption maxima of cysteamine, of thiomalic acid, and of glutathione near 235 mμ show variations with *p*H similar to those of cysteine. They interpreted these shifts in the absorption spectrum with increasing *p*H in terms of the formation, and subsequent breakage, of an internal hydrogen bond between the thiol and an uncharged amino group, or, in the case of thiomalic acid, of an ionized carboxyl group.

A Critical Examination of the Hypothesis of Thiol Hydrogen Bonding

The hydrogen-bonding hypothesis set forth in Scheme II obviously differs decidedly from Scheme I. The two schemes, however, are not in principle mutually exclusive. The four different microscopic forms that are indicated in Scheme I must exist in cysteine solutions, and the hydrogen-bonded form X in Scheme II may be present also. Thus we can in principle formulate an equilibrium between B and C of Scheme I and X of Scheme II, all of which contain the same number of protons:



SCHEME III

Obviously, the equilibrium constants involved, i.e., the concentration ratios B:X:C, are independent of *p*H, although they do, in general, vary with temperature, ionic strength, and dielectric constant of the medium. The hypothesis of De Deken *et al.* (1956) assumes that, of these three forms, X is predominant and B and C are relatively negligible. On the other hand, Benesch and Benesch (1955) and Elson and Edsall (1962) neglected form X completely in the formulation of Scheme I. Which point of view is nearer to the truth?

In evaluating this problem we may consider several lines of evidence. (1) It is universally agreed that hydrogen bonds which involve sulfur are much weaker than those involving nitrogen, oxygen, or fluorine. On the Pauling electronegativity scale (Pauling, 1960) the electronegativity of sulfur is only 2.5, a value identical to that for carbon and well below the values for nitrogen, fluorine, and even chlorine, which all lie between 3 and 4. Since high electronegativity is a prerequisite for strong hydrogen bonding, this clearly indicates that hydrogen bonds involving thiol groups must be relatively weak.

(2) Evidence from crystal structures of molecules containing thiol groups is highly relevant, but the data are few. The study of Wright (1958) on the structure of glutathione demonstrated that the thiol group is not involved in intramolecular hydrogen bonding, and apparently not in intermolecular hydrogen bonding either.

(3) The hydrogen bond postulated for form X of Scheme II would be part of a 5-membered ring. Examination of a molecular model of cysteine indicates that the SH $\cdots$ N bond angle would be of the order of 110°, a long way from the value of 180° which is the most favorable angle for hydrogen-bond formation.

(4) The shift of the ultraviolet-absorption maximum in cysteine from near 230 to about 240 mμ had already been observed by Benesch and Benesch (1955) before the study of De Deken *et al.* (1956). Benesch and Benesch interpreted this shift on the assumption that the microscopic form $^-SRNH_3^+$ (B) has an absorption maximum at slightly shorter wavelengths than the microscopic form $^-SRNH_2$ (D). The removal of a proton from the amino group must certainly lead to electronic displacements in the molecule, which should alter the energy levels associated with the transitions

giving rise to the absorption band near 235 mμ. Although no quantitative calculations of shifts in energy levels are available, this interpretation appears at least as reasonable as that involving the hydrogen-bonded intermediate X.

(5) The data that De Deken *et al.* (1956) have themselves reported for the absorbance of cysteine solutions in the 235-mμ region, as a function of *p*H, also appear to furnish strong evidence against Scheme II. Figure 11 of their paper, in close agreement with the earlier findings of Benesch and Benesch (1955), shows a large increase in absorbance at 235 mμ between *p*H 7 and 9; indeed, approximately half the total absorbance increase between *p*H 6 and 12 occurs in this interval. This is the *p*H region in which, according to Scheme II, the thiol group remains un-ionized, although hydrogen bonded, while a proton is being removed from the charged NH_3^+ group. The removal of a proton from the NH_3^+ group should give rise to no absorption at 235 mμ; no aliphatic amino group, whether protonated or not, shows absorption in this range of wavelengths. Hydrogen-bond formation by the S—H group, without actual ionization, should not give rise to absorption in this range. The data would indicate clearly that about half the molecules, or slightly more than half, have lost protons from their S—H groups by *p*H 9. These experimental findings fit Scheme I very well; indeed, they furnished the basis for this scheme in the treatment of Benesch and Benesch (1955). In Scheme II they appear inexplicable.

(6) The study of vibrational spectra furnishes a powerful and sensitive test for the presence of hydrogen bonding. For thiol compounds the S—H stretching frequency in the Raman spectrum is sharp and highly intense. (In the infrared spectrum it is relatively much weaker.) In aliphatic thiols it lies in the region 2570–2580 $cm.^{-1}$, a region of the spectrum that is virtually free of other Raman bands.

The effect of hydrogen bonding on OH or NH stretching frequencies is to displace these vibrations to lower frequencies and also to broaden them markedly. Such effects are striking and well known in the Raman and infrared spectra of water, alcohols, and phenols, either in the pure liquid state or in solutions containing other molecules that can act as acceptors in hydrogen bonding.

No such effect is observed with cysteine. The S—H stretching frequency is at 2580 cm^{-1}; it is sharp and well defined (Garfinkel and Edsall, 1958; Elson and Edsall, 1962). For comparison, we note that the Raman spectra of various aliphatic mercaptans in the pure liquid state show S—H stretching frequencies between 2568 and 2573 cm^{-1} (Kohlrausch, 1943). Thus the S—H frequency of cysteine in aqueous solution is essentially identical in position with that found in the pure aliphatic mercaptans; it is not broadened, nor is it shifted to lower frequencies. These data therefore appear to provide no evidence for hydrogen bonding by the thiol groups in cysteine.

(7) The preceding discussion is concerned with equilibrium relations, not with kinetics. Kinetically, the interconversion of the two isomers B and C of Scheme I may proceed by exchange of protons between the thiol and amino groups and the surrounding solvent molecules; in this case the microscopic forms A and D are transitory intermediates between B and C. On the other hand, B may be converted to C, and vice versa, through an intramolecular proton transfer, as indicated in Scheme III. As yet we lack the data to appraise the quantitative significance of these two routes of exchange. Eigen and De Maeyer (1963) report an intramolecular proton exchange rate in β-mercaptoethylamine (cysteamine) of 2×10^7 sec^{-1} at 20°, but the experimental details of the work are not yet available. Even if the intramolecular transfer route according to Scheme III should prove to be predominant, the equilibrium data discussed above indicate that the hydrogen-bonded intermediate X is present in very small amount at any moment; the proton leaps back and forth between states B and C, and the intermediate state is almost undetectable.

The presence of a small amount of the hydrogen-bonded intermediate X is not, of course, completely excluded by the existing data.[2] However, the weight of evidence summarized above points clearly to the conclusion that this intermediate must be present in very low concentrations, if at all. The author concludes that Scheme I remains an adequate formulation of the ionization of cysteine, and presumably of other related thiol compounds as well. These considerations may be of importance in the interpretation of the behavior of proteins containing thiol groups.

References

Benesch, R. E., and Benesch, R. (1955), *J. Am. Chem. Soc. 77*, 5877.

Cecil, R. (1963), *Proteins 1* (2nd ed.), 391.

De Deken, R. H., Broekhuysen, J., Bechet, J., and Mortier, A. (1956), *Biochim. Biophys. Acta 19*, 45.

Edsall, J. T., and Wyman, J. (1958), Biophysical Chemistry, Vol. I, New York, Academic, pp. 496–504.

Eigen, M., and De Maeyer, L. (1963), *in* Technique of Organic Chemistry, 2nd ed., Vol. VIII, Part II, Friess, S. L., Lewis, E. S., and Weissberger, A., eds., New York, Academic, p. 1040.

[2] Insofar as the thiol group acts as a hydrogen-bond donor, the oxygen atoms of neighboring water molecules can also act as acceptors. Any SH···O hydrogen bonding of this type need not be indicated explicitly in Scheme I. Insofar as it exists, and there is probably little of it, for reasons already given, it is implicitly included in the formulas of Schemes I and II, which denote the total concentrations of the various species in solution, including their relations to the surrounding water molecules, which need not be specified in detail. The arguments in this paper are not to be construed as totally denying the existence of SH···N or SH···O hydrogen bonds. The work of Josien, in particular (see, for instance, Josien *et al.*, 1957) has clearly shown downward displacements of S—H stretching frequencies in benzenethiol by 30–40 cm^{-1}, in the presence of basic solvents, and even larger shifts for H_2S in such solvents. Our present conclusion is simply that such hydrogen bonding is very weak or absent for cysteine and related aliphatic thiols dissolved in water.

Elson, E. L., and Edsall, J. T. (1962), *Biochemistry 1*, 1.
Garfinkel, D., and Edsall, J. T. (1958), *J. Am. Chem. Soc. 80*, 3823.
Grafius, M. A., and Neilands, J. B. (1955), *J. Am. Chem. Soc. 77*, 3389.
Josien, M. L., Dizabo, P., and Saumagne, P. (1957), *Bull. Soc. Chim. France*, 423.
Kohlrausch, K. W. F. (1943), Ramanspektren, Leipzig, Akademische Verlagsgesellschaft p. 211.
Martin, R. B., and Edsall, J. T. (1958), *Bull. Soc. Chim. Biol. 40*, 1763.
Pauling, L. (1960), The Nature of the Chemical Bond, 3d ed., Ithaca, N. Y., Cornell Univ. Press, p. 90.
Wright, W. B. (1958), *Acta Cryst. 11*, 632.

Polymerization-Depolymerization of Tobacco Mosaic Virus Protein. IV. The Role of Water*

Charles L. Stevens and Max A. Lauffer

ABSTRACT: Previous investigations have shown that the polymerization of tobacco mosaic virus protein is a spontaneous endothermic process, presumably because of an interaction between the protein subunits and solvent. Specifically, it is an increase of entropy associated with the loss of a bound solvent component which appears to be decisive. Utilizing a new quartz spring-balance, we have measured directly the amount of water lost by the protein upon polymerization and have found it to be about 150 moles of water from each 1×10^5 g of polymerizing material. This amount may be sufficient to explain the entire observed entropy of polymerization. In addition, we find an increase of about 7 moles of salt bound to each 1×10^5 g of protein and an increase of 0.0060 ml/g in the partial specific volume. The data seem to be consistent with electrostrictive binding of water at zwitterionic groups, but the loss of icelike domains of water about hydrophobic groups on the protein is an alternative hypothesis.

Protein from tobacco mosaic virus (TMV),[1] even when freed from nucleic acid, has the remarkable property of polymerizing spontaneously into a rod similar to the intact virus. The apparent physical differences between the polymerized protein and the intact virus are minor (Klug and Caspar, 1960); polymerization of protein in the absence of nucleic acid, however, does not produce infective particles.

Polymerization is thermodynamically reversible (Lauffer *et al.*, 1958; C. E. Smith and M. A. Lauffer, data to be published). Positive increments of temperature, ionic strength, and protein concentration and a decrease of *p*H favor polymerization. Protein in 0.1% solution buffered at *p*H 6.5 by 0.1 ionic strength sodium phosphate exists predominantly in the polymerized state at room temperature and is depolymerized at refrigerator temperature. Although the basic subunit has a molecular weight of about 17,500, the polymerizing unit under the conditions mentioned seems to have a molecular weight of about 50,000–100,000 (Schramm and Zillig, 1955). Complete dissociation is attained probably only under conditions of extreme dilution (Ansevin and Lauffer, 1963), extremes of *p*H, or by chemical modification or denaturation of the subunit.

Lauffer *et al.* (1958) found that the polymerization reaction is endothermic and has a positive entropy change. The last result might be regarded as unexpected because polymerization involves the ordering of protein units into a highly specific structure. Accordingly, the authors state that the polymerization of protein from TMV can be understood only in terms of interaction between the protein and solvent; they propose that the protein monomer loses water of hydration upon polymerization. Thus, the over-all reaction is: hydrated monomer yields polymer plus water with a net increase of enthalpy and entropy.

The purpose of this investigation is to provide a direct test of the hypothesis that water is released from

* From the Department of Biophysics, University of Pittsburgh, Pittsburgh, Pa. *Received September 25, 1964.* Publications I, II, and III of this series are Ansevin and Lauffer (1963), Lauffer (1964), and Ansevin *et al.* (1964), respectively. This is publication No. 111 of the Department of Biophysics, University of Pittsburgh. Work was supported by a U.S. Public Health Service grant (GM 10403-02). The data presented in this paper are taken in part from the dissertation presented by C.L.S. to the University of Pittsburgh in partial fulfillment of the requirements for the degree of Doctor of Philosophy, April 18, 1962.

[1] Abbreviation used in this work: TMV, tobacco mosaic virus; RNA, ribonucleic acid.