THE USE OF SOLUBILITY AS A CRITERION OF PURITY OF PROTEINS

I. APPLICATION OF THE PHASE RULE TO THE SOLUBILITY OF PROTEINS.
II. THE SOLUBILITY CURVES AND PURITY OF CHYMOTRYPSINOGEN

By J. A. V. BUTLER*

(From the Laboratories of The Rockefeller Institute for Medical Research, Princeton, New Jersey)

(Received for publication, June 20, 1940)

I. Application of the Phase Rule to the Solubility of Proteins

Since proteins cannot be heated and can only be treated with a limited range of solvents without injury, it has been inevitable that the methods of separation and characterization of different proteins should depend mainly on their behavior in aqueous solutions of varying acid and salt concentrations. In the early studies of Hardy (1) and Mellanby (2) it was found that the solubility of serum proteins varied with the amount of the saturating body, and Sørensen and his collaborators also found this kind of behavior with egg (3) and serum albumin (4) and casein (5). Sørensen seems to have been the first to apply the phase rule in the sense that for a single solute the solubility should be independent of the amount of the solid phase, and his inability to obtain fractions obeying this criterion led to his theory of "reversibly dissociating systems," according to which the real independent components of a protein were polypeptides which formed complex aggregations varying with the circumstances.

Since then a number of cases have been found in which the solubility is approximately independent of the amount of the solid phase (6), at least when there is a considerable excess of the latter. Northrop and Kunitz (7) have, however, shown that constant solubility in the presence of a large excess of solid is an insufficient criterion of purity, but if solubility curves (i.e., solubility against total amount of saturating material) are determined in the region in which the excess of solid is small it may be possible to distinguish the solid phases present and in some cases to differentiate between mixtures of crystalline phases and solid solutions (8).

Few investigations of the purity of proteins have been made which satisfy

* Fellow of The Rockefeller Foundation.

189
these conditions and the object of the present work was to make a more exhaustive study of the solubility method as a criterion of purity of proteins.

**Application of Phase Rule to Protein-Acid-Salt Solutions**

In the first place it may be well to point out exactly what can be deduced from the phase rule. In the usual form this is written as \( P + F = C + 2 \), where \( P \) = number of phases, \( F \) = number of degrees of freedom, and \( C \) = number of components. We have first to choose as components the least number of substances by which we can express the composition of the solid protein phase or phases and the solutions which can be formed from them. Since the solutions with which we shall be concerned contain a buffering system, the least number of components will be four; *viz.*, protein, salt, acid, and water. Additional components may of course be introduced into the system at will, *e.g.* there may be two salts, such as the buffering salt and a neutral salt like ammonium sulfate. But it is unnecessary to regard a salt formed by the combination of the protein and the acid, even if the solid is present partly or completely as such, as a separate component, since its composition can be completely specified by the amounts of protein and acid contained in it. We can take as the component protein, the protein in a specified invariable condition and the amount of protein in any phase is determined by a suitable analysis; *e.g.*, of the amount of "protein nitrogen."

By the phase rule, this system of four components has five degrees of freedom in one phase, and four degrees of freedom when present in two phases. Four of these are the temperature, pressure, and concentrations of salt and acid respectively. When these four variable factors are fixed there remains one degree of freedom for a single phase and none for a system of two phases. In a homogeneous solution of a single protein the concentration of protein may vary, but if a solid phase is present the concentration of the dissolved protein is fixed. If it is not fixed then an additional component must be present. If it should happen that two distinct solid phases are present (*e.g.*, if a new phase were formed by reaction of the protein with the salt solution) the number of degrees of freedom would be further reduced by one; *i.e.*, it could only occur at a particular acid or salt concentration.

The characteristics of the possible systems which may be formed from one or two proteins together with water, acid, and salt are shown in Table I. When two proteins are present, a single solid phase may consist of either one protein, or a solid solution of the two proteins. In either case, since there are now five components in two phases, the number of degrees of freedom
under the stated conditions is one. Thus if the solid phase is one of the proteins, the concentration of the other in the solution is variable, while if the solid phase is a solid solution of both proteins, the composition of the solution will depend on the composition of the solid solution, but if the composition of the solid solution is fixed, the composition of the aqueous solution in equilibrium with it is also fixed. Lastly, if there are two solid phases the system again becomes invariant and the composition of the solution, when the temperature, pressure, and concentrations of acid and salt are fixed, is definite.

If additional components such as salts are present, each new component introduces a new degree of freedom; but if the concentrations of the new components are fixed, the degrees of freedom are unchanged.

The case of a solid solution of two proteins may be discussed in greater detail. Such a solution will be in equilibrium with an aqueous solution of fixed composition, but the ratio of the proteins in the aqueous solution may be either the same or different from that in the solid solution. In the first case an equilibrium solution can be formed by simple dissolution of the solid phase and therefore without any change in its composition. Therefore the solubility will be independent of the amount of the solid. In the second case the formation of the equilibrium solution will result in change in the composition of the solid, and as shown by Northrop and Kunitz (9) the effective solubility will then vary with the amount of the solid phase. The following is a simplified derivation of this conclusion, which does not depend on the particular assumptions previously made.

<table>
<thead>
<tr>
<th>Components</th>
<th>Degrees of freedom at constant temperature, pressure, and concentrations of acid and salt</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 = water, salt, acid, protein</td>
<td>1</td>
<td>A homogeneous solution</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>A solid phase + saturated solution</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Two solid phases + saturated solution. (Only possible at a particular concentration of acid or salt.)</td>
</tr>
<tr>
<td>5 = water, salt, acid, two proteins</td>
<td>1</td>
<td>A homogeneous solution</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>One solid phase containing (a) a single protein, or (b) a solid solution of two proteins, in a solution of variable composition.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Two solid phases in solution of constant composition</td>
</tr>
</tbody>
</table>
Suppose that the relation between the composition of the solid and that of the solution in equilibrium with it is as represented in Fig. 1, where $X$ is the composition of the solid phase in equilibrium with a solution in which the ratio of the two proteins is represented by $Y$. In this illustration the proportion of $A$ entering the solution is greater than that in the solid and the solid is therefore left with a smaller proportion of $A$. We will now consider the following cases:

1. If the quantity of solid is large compared with the amount dissolved, a solution of composition $Y$ is obtained with a very small change in the composition of the solid and the solution obtained from the original crystals will effectively have the composition $Y$.

2. If the solid is nearly all dissolved, the composition of the solution must approximate to that of the original crystals and will be represented by $Y'$. The solid which is in equilibrium with this solution has the composition $X'$.

If we add increasing amounts of the solid phase $X$ to a definite quantity of the solvent, on the assumption that the solid phase can adjust itself to equilibrium with the solution, we should find that when excess of solid first appears it has the composition $X'$ and a solubility which corresponds to the point $Y'$. As the excess of solid increases the solid phase remaining will change in composition from $X'$ to $X$ and the solubility will change correspondingly. It follows that the point at which the solid ceases to give a clear solution is not necessarily the solubility of the solid phase, but it will be the solubility of a solid in equilibrium with the solution having the same composition as the original solid phase. The amount of protein dissolved will be equal to the amount of protein added, up to the point at which solid first appears, but under these circumstances the solubility may then change as the excess of solid is increased until with a great excess of solid a constant value is obtained.

It is possible that the whole of the solid phase may not readily come to equilibrium with the solution and that the adjustments of composition may be confined to the surface layers of the crystals. In that case we may get the solubility which corresponds to a small residue of the solid while an appreciable amount of the solid remains.

The amount of protein present in a solution of definite acidity and salt concentration will therefore be independent of the amount of solid in contact with the solution, from the first appearance of turbidity, in the following cases:
1. The solid is a single phase. (a) A single pure protein, and (b) a solid solution of two proteins, which dissolve in the same proportion as they are present in the solid.

2. The solid consists of two distinct phases. (a) A mixture of two proteins, present in amounts proportional to their solubilities.

To distinguish cases 1 b and 2 a from 1 a, we may observe that it is unlikely that the solubilities of two proteins would be equally affected by a change of acidity and/or salt concentration. If the solvent is changed, it is therefore probable that in the case of the solid solution the two proteins will no longer dissolve in the same proportions and the solubility will then depend on the amount of the solid phase, as described above; while in the case of two distinct solid phases, increasing the amount of solvent will give rise to a point at which one solid has completely dissolved, and here the protein concentration becomes variable. Independence of the solubility of the amount of the solid phase in several solvents can therefore be taken as strong evidence that the material is a pure protein.

On the other hand, a single protein which is present in two distinct solid forms, e.g. amorphous and crystalline, may also give the solubility curve of a mixture. If equilibrium between the two solids and the solution were established, this would be a rare occurrence, since it can only be found in particular solvents in which it happens that the two forms have the same solubility. But the approach to equilibrium may be very slow and cases may be encountered in which the two solids persist for long periods although not in equilibrium and give rise to the solubility curve of a mixture. Such cases may be distinguished by adding sufficient solvent to dissolve one phase and examining the residue. If it gives identical solutions to the original solid, or can be changed into the other form, we are dealing with a case of polymorphism. To decide the nature of the material it is therefore necessary to examine not only the solubility curves in a number of solvents, but in case the curves of a mixture are obtained, to separate the solids and determine their nature.

**Equilibration of the Protein with the Solvent**

We must now consider how far the conditions of constant acidity and constant salt concentration can be satisfied with varying amounts of the solid protein. In the previous experiments of Kunitz and Northrop, it has been considered that this condition is satisfied by previously equilibrating the solid by washing with successive portions of the solvent in which the solubility is to be measured, until the solubility of the protein is constant. Then it is taken that the addition of varying amounts of the equilibrated
solid to the solvent will not produce variations of the acidity and salt concentration in the latter. A careful examination of the validity of this process is desirable. We can distinguish the following cases:

1. Suppose that the whole of the solid protein can come into equilibrium with the buffer. It is known that protein crystals contain considerable quantities of water and ions derived from the mother solution, and it has been found that appreciable differences of density (10), and of cell dimensions (11) occur according to the nature of the solution in which the crystals are suspended. It follows that the penetration of the salt solution in which the crystals are suspended occurs at least in some cases. In such a case the protein in the interior of the crystal may be expected to come into equilibrium with the solvent both in respect to acidity and salt concentration. On mixing a quantity of the solid with the buffer solution, the equilibrium finally reached will correspond with a pH intermediate between that of the buffer and the original protein. If the solution is poured away and a fresh quantity of the buffer solution added, a second equilibrium will be reached, and it seems evident that after successive treatments in this way the solid will approach a state in which further washings produce no change in the solid. The solubility will then be constant and will be independent of the ratio of the solid to liquid phases.

This procedure would fail only when the solubility of the protein is such that it dissolves completely before equilibrium is reached. This might be overcome by increasing the concentration of the buffer in such a way that the buffering power of a given quantity of solvent is increased, without increasing the solubility of the protein; or by using as the solvent a more concentrated salt solution in which the protein solubility is less.

2. The solid may not equilibrate at all with the buffer. The only function of successive washings would then be to remove the mother liquor adhering to the crystals. A given quantity of solvent would then always dissolve the same amount of protein; which will produce a constant change of pH. The change of pH produced by the dissolution of the protein would, however, depend on the pH of the protein preparation and the solubility will therefore probably depend on the nature of the solution in which the protein crystals were formed.

3. Suppose that the buffer solution reacts with the protein to form a new phase, e.g. a protein salt. It has been shown that two solid phases of a single protein can only co-exist with a solution in which one variable concentration, e.g. acidity, is fixed. The buffer solution will therefore be changed by the reaction to a composition which corresponds to the pH fixed by the protein solids. The addition of more buffer will produce more
of the protein salt, and only when the first solid phase has been completely converted into the second can the pH change and approach the value defined by the buffer solution. In this case it is evident that a stoichiometrical excess of the buffering substances is required to bring the protein into a state in which it can exist unchanged in contact with fresh portions of the buffer solution.

II. The Solubility Curves and Purity of Chymotrypsinogen

Chymotrypsinogen from beef pancreas was chosen for these experiments on account of its stability and ease of crystallization. Good solubility curves for this material have already been obtained by Kunitz and Northrop (12) using the activity, when converted into chymotrypsin, and the protein concentration as determined by turbidity measurements as measures of the solubility. In this work we have used the total nitrogen of the solution as the measure of the solubility; a more rigorous test since nitrogenous impurities of a non-protein character are included in the determined values.

Under these circumstances ammonium sulfate solutions cannot be used as the solvent and magnesium sulfate was substituted. The first experiments showed that when recrystallized seven or eight times from magnesium sulfate, this material gave quite good solubility curves, the solubility being practically independent of the amount of solid phase from the first appearance of turbidity, but the actual solubility varied somewhat from one preparation to another, and the residue left after dissolving part of the substance had an appreciably different solubility. This suggested that the material might be a solid solution and in an attempt to effect a separation a series of fractional crystallizations was carried out on twice crystallized material.

In this first series a solution of the protein was prepared at pH 4.0 and sufficient magnesium sulfate was added to cause the crystallization of about one-third of the total protein. The addition of more magnesium sulfate gave a further crop of crystals and the remainder was precipitated from a nearly saturated solution of the salt. This process was repeated on the fractions five times, the “less soluble” precipitate from one fraction being systematically united with the “more soluble” precipitate from the next. In this way three fractions were obtained, A, B, and C; A being obtained from the first precipitates and B and C from the middle and last fractions. The three fractions were crystallized twice from a pH 4.0, 0.4 saturated magnesium sulfate solution and the solubility curves were determined with this solvent. The curves obtained are shown in Fig. 2. Both B and C
gave very good solubility curves, the solubility being practically constant from the first appearance of turbidity, but the solubility of C is slightly greater than that of B. The curve of A has a distinct break and the solubility continues to rise after the first appearance of turbidity at P to the region of Q, indicating that this material is complex.

The whole fractionation was repeated with new material in a somewhat simplified form in the course of which the middle fraction (B) was divided.
between the two end fractions. In this way two fractions $A'$, $C'$ were finally obtained. It was found that the solubility curve of $A'$ was identical with that of $A$ and $C'$ with the previous fraction $C$.

Since $B$ and $C'$ were obtained by adding more magnesium sulfate to a saturated solution of protein in approximately the same solvent as that in which the solubility was determined, it was possible that these fractions might contain two substances present in the same proportion as their solubility in this solvent. Solubility curves of $C'$ were therefore deter-

![Graph showing solubilities of crystals formed at pH 4.0 and pH 5.0. Triangles refer to crystals formed at pH 4.0, and open circles to crystals formed at pH 5.0. Closed circles represent successive concentrations in supersaturated solutions. The upper set of points is for solubilities in 0.19 saturated magnesium sulfate, pH 5.0 solvent; the lower set of points is for solubilities in 0.4 saturated magnesium sulfate, pH 4.0 solvent.](image)

**Fig. 3.** Solubilities of crystals formed at pH 4.0 and pH 5.0. Triangles refer to crystals formed at pH 4.0, and open circles to crystals formed at pH 5.0. Closed circles represent successive concentrations in supersaturated solutions. The upper set of points is for solubilities in 0.19 saturated magnesium sulfate, pH 5.0 solvent; the lower set of points is for solubilities in 0.4 saturated magnesium sulfate, pH 4.0 solvent.

mined in two other solvents of pH 5.0 and 8.0, with suitable concentrations of magnesium sulfate (Fig. 2, d and e). In both cases a good solubility curve was obtained and since it is very unlikely that the ratio of the solubilities of two substances would be identical at three different pH's, this may be taken as strong evidence that $C$ and $C'$ are a single substance.

That a true equilibrium was established in these experiments was shown by the following facts.

1. The solubility of the crystals in a given solvent was independent of the pH of the solution in which the crystals were formed. Fig. 3 shows the solubilities of crystals formed in magnesium sulfate solutions at pH 4.0 and 5.0 in successive shakings with fresh portions of two distinct solvents. The
crystals made at both pH's ultimately reached the same solubility in either solvent.

2. The solubility of the crystals was independent of the concentration of magnesium sulfate in the solution in which they were formed.

<table>
<thead>
<tr>
<th>Concentration of magnesium sulfate</th>
<th>0.35 saturated</th>
<th>0.4 saturated</th>
<th>0.5 saturated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solubility in approximately 0.4 saturated magnesium sulfate at pH 4.0 (mg. N/ml.)</td>
<td>1.01</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

3. The same equilibrium was reached from supersaturated as from undersaturated solutions. The supersaturated solutions were prepared by cooling a saturated solution with excess of the crystals. More dissolves at the lower temperature and when the temperature is raised again a supersaturated solution is obtained. After several days the protein concentration returns to the original value (broken lines of Fig. 3).

The material A, which has a complex solubility curve is less soluble than B and the solubility decreases on repeated extractions. It is probable that this material contains some substance which is precipitated at the smaller salt concentration and which forms a solid solution with the chymotrypsinogen.

Apart from the solubility curves and a rather greater proportion of non-protein nitrogen in A', no significant difference between the fractions A' and C' could be detected. When converted by trypsin into chymotrypsin, they had practically the same activities per milligram nitrogen, as measured both by milk clotting and by the digestion of hemoglobin, and the velocity constants of the activation process were the same within the experimental error (see Table II). Since the accuracy of these measurements is of the order of 5 per cent it can be concluded that the impurity in A and A',

*For these measurements I am indebted to Miss M. McDonald. For method see Anson (14).
which is not present in C, does not contribute more than 5 per cent of the total nitrogen, unless it is also capable of activation. It must be remembered that if the molecular weight of the impurity is low the molar fraction may be appreciable and sufficient to produce a significant change of solubility.

The impurity in this fraction is probably present in solution in the protein crystals. If it were a separate phase the solid residue in the region PQ, in which the solutions are saturated with the main constituent but the subsidiary constituent continues to dissolve (Fig. 2 a), should consist of the pure protein. Some of this residue was collected and its solubility curve, shown by the dashed line in Fig. 2 a, indicates a greater concentration of the impurity in the residue; as might be expected for a solid solution. The dialysis of solutions of the fraction A' caused a small decrease in the amount of non-protein nitrogen, but had no significant effect on the solubility curve.

The electrophoresis of solutions A' and C' and of a mixture of A' and C' in a 1/15 citrate-hydrochloric acid buffer solution at pH 3.0 after equilibration by dialysis with a large quantity of this buffer, was kindly examined by Dr. Alexandre Rothen of The Rockefeller Institute for Medical Research in New York. The electrophoretic patterns obtained by him, using Longsworth's schlieren scanning technique, are shown in Fig. 4.

In all cases only a single moving boundary was observed and the cathodic boundaries in every case were extremely sharp. The mixture of A' and C' showed no signs of any resolution even after more than 15 hours migration. The migration velocities were approximately the same in all solutions, the small differences being attributable to slight variations of the pH in the dialyzed solutions.

We conclude that no genuine fractionation of the protein has been achieved, but a small quantity of impurity has been concentrated in the fractions first precipitated. This is insufficient in quantity to produce any appreciable difference in the enzymatic properties and it does not produce a visible boundary in the electrophoretic pattern. But if the molecular weight of the impurity is comparatively low, its molar fraction in the crystals may be sufficient to produce the observed diminution of solubility of the protein and the break in the solubility curve. The final fractions satisfy every test of a pure substance which has been applied.

EXPERIMENTAL METHODS

Preparation of the Material and Methods of Crystallisation.—The chymotrypsinogen was prepared from beef pancreas by the method described by Kunitz and Northrop (13). It was recrystallized from magnesium sulfate by dissolving in 2.5 volumes of 0.2 m acetic acid and then sufficient sodium acetate was added to produce a buffer of pH 4.0
FIG. 4. Electrophoresis patterns of fractions $A'$ and $C''$ and mixtures of both (Dr. A. Rothen).

Solution $A'$; pH 2.98, velocity in cathode compartment $0.675 \times 10^{-4}$ cm.$^2$/sec. volt; gradient 4.09 volt/cm.
Solution $C''$; pH 2.98, velocity in cathode compartment $0.667 \times 10^{-4}$ cm.$^2$/sec. volt; gradient 4.09 volt/cm.
Solution $A' + C''$; pH 3.0, velocity in cathode compartment $0.608 \times 10^{-4}$ cm.$^2$/sec. volt; gradient 4.09 volt/cm. for 188 minutes, 2.04 volt/cm. for additional time.
(22.6 ml. 0.2 M sodium acetate to 100 ml. 0.2 M acetic acid), followed by 2 volumes of saturated magnesium sulfate. Nearly the whole of the protein crystallized out in the course of a few hours. In the fractionations the protein was dissolved in 2 volumes of acetic acid and the same proportion of sodium acetate; then sufficient magnesium sulfate was added to bring the concentration to one-third saturation. The crystals formed were filtered off and the magnesium sulfate concentration was raised to about 0.45 saturation, when a further crop of crystals was obtained. The remaining protein was precipitated by adding solid magnesium sulfate to saturation. In Series II, after the first crop of crystals, the magnesium sulfate concentration was raised at once to about 0.7 saturation.

**Determination of the Solubility.**—The protein was placed in a test tube with a marble or ball-bearing and the tube was completely filled with the solvent up to the rubber stopper. It was rocked in a thermostat at 19.5°C ± 0.1°C for 2 days. The tubes were now centrifuged and a portion of the clear supernatant liquid was carefully removed by pipette for analysis. These operations were carried out in a constant temperature room at 19.5°C ± 0.5°C. The total nitrogen in the liquid was determined by a semi-micro Kjeldahl method using 1 ml. samples. The material was shaken with fresh portions of the solvent until two successive solubilities were the same and then a “solubility curve” was obtained by distributing the suspension in varying amounts among tubes filled with solvent.

**SUMMARY**

1. The conditions under which the phase rule may be applied to systems containing proteins are formulated.

2. An attempt was made to fractionate chymotrypsinogen, by crystallization in stages with increasing concentration of magnesium sulfate. No significant fractionation of the protein was achieved, but a small amount of impurity which affects the solubility, while having little influence on other properties of the material, was concentrated in the fractions first precipitated.

3. The solubility of the final fraction was independent of the amount of the saturating solid, from the first appearance of a solid phase, in solvents of three different pH's. The solubility was independent of the environment in which the crystals were formed (within the limits in which crystallization can be carried out) and the same value was reached from the supersaturated as from the undersaturated side. This material, therefore, conforms closely with the phase rule criteria of a pure protein.

The author wishes to express his thanks to Dr. John H. Northrop for granting him the facilities of his laboratory and for much valuable advice.

**LITERATURE**