PROTEIN CONSTITUENT ANALYSIS BY THE SOLUBILITY METHOD¹

STANFORD MOORE, WILLIAM H. STEIN, AND MAX BERGMANN

Laboratories of the Rockefeller Institute for Medical Research, New York, New York

Received January **2,** *1948*

The principles of the solubility-product method and their application to the determination of amino acids in protein hydrolysates are outlined. The methods for leucine and glycine serve as specific examples. The properties of aromatic sulfonic acids as reagents for amino acids are included to illustrate their use in solubility analysis and in the qualitative fractionation of mixtures of amino acids.

I. INTRODUCTION

Protein chemistry requires as a fundamental basis the accurate establishment of the amino acid composition of the various proteins. The problem is analogous in principle to the establishment of the composition of simpler molecules in terms of their constituent atoms. In practice, many diverse approaches have been made to the problem of analysis for amino acid constituents. Yet the relative difficulty of the task is evidenced by the fact that we cannot write today the empirical formula for the amino acid composition of even the simplest protein, and among many of the more complex proteins 40 per cent or more of the amino acids in the molecule remain completely unaccounted for.

Fractionation procedures, precipitation methods, and colorimetric methods have all been applied, with varying degrees of success, to the analysis of complicated mixtures of amino acids. For a few of the amino acids, results of high precision have been attained; for others, however, the accuracy has been less satisfactory, and for many of the constituent parts of the protein molecule there has existed no method which could be classified as a quantitative determination. The results acquired with the aid of the existing quantitative methods, together with the significant contributions of the qualitative methods, have been of immeasurable assistance in the progress of protein chemistry. Severtheless, the history of this field for the past fifty years seems to point to the conclusion that a major advance in the *quantitative* knowledge of protein structure (complete analysis) can be expected only through the introduction of some fundamentally new approach.

The solubility method represents one attempt at a new solution of this problem **(3,** 11). Another approach is provided by the isotope dilution method of Rittenberg and Foster (8, 9). The solubility method, which is the subject of this discussion, has originated from work on the gravimetric determination of amino acids and represents an attempt to extend the range of this type of analytical approach.

Presented at the Symposium on Physicochemical Methods in Protein Chemistry, which was held under the joint auspices of the Division of Physical and Inorganic Chemistry and the Division of Biological Chemistry at the 102nd Meeting of the American Chemical Society, Atlantic City, New Jersey, September 8-12, 1941.

In gravimetric analysis a familiar procedure consists in the precipitation of a substance in the form of a sparingly soluble salt. The solubility product of the ions involved in the equilibrium is the principal factor determining the completeness of precipitation. This feature is, of course, common to inorganic as well as organic analysis,—to the determination of barium as sulfate as well as to the determination of arginine as flavianate. In the amino acid field the difficulty of attaining quantitative isolation is evidenced by the fact that in only a very few instances are reagents known which are capable of quantitatively precipitating one and only one amino acid from a mixture so complex as a protein hydrolysate. **A** second procedure, which has of necessity been followed in some cases, has been the use of reagents which precipitate only the major part of the amino acid, supplemented by the application of corrections for the amount of amino acid remaining in solution in the hydrolysate. There is no sound method for determining this correction factor, however, since the absolute value of the solu-

FIG. **1.** The solubility-product method applied to the analysis for an amino acid under conditions of partial precipitation.

bility of a compound is a function of the composition of the solvent. The solubility determined in water or normal hydrochloric acid may be far different from the solubility in a protein hydrolysate. At this point in the problem the solubility-product method was introduced in order to provide a sound principle for the calculation of the total amount of an amino acid in a solution from data obtained by partial precipitation of the amino acid.

11. PRINCIPLES OF THE SOLUBILITY-PRODUCT METHOD

The first formulation of the method **(3,** 11) is illustrated in figure 1. Two aliquots of a solution are taken, each containing *A* moles of the amino acid to be determined. To the first aliquot is added R_1 moles of a reagent which causes the partial precipitation (50 to 60 per cent) of the amino acid as a salt. If the precipitated salt $(Y_1 \text{ moles})$ is binary, containing equimolar quantities of amino acid and reagent, it may be represented by the shaded square in figure 1. Remaining in solution are $A - Y_1$ moles of the amino acid and $R_1 - Y_1$ moles of the reagent.

To a second aliquot is added a larger quantity of reagent, *Rz,* causing an increased precipitation up to, for example, **70** to 80 per cent of the amino acid present. The method of analysis is founded on the experimental observation that under conditions such as these the solubility product, that is, the product of the molar quantities of the amino acid and the reagent remaining in solution, is practically constant. In this case the solubility product for the first point is $(A - Y_1)(R_1 - Y_1)$, represented by the area K_1 , and for the second point $(A - Y_2)(R_2 - Y_2)$, or the area K_2 . If K_1 and K_2 have identical, or, as will be shown later, nearly identical values, they can be eliminated by equating the two products, and the resulting expression can be solved for the only unknown, *A.* This calculation yields the total amount of amino acid present in the sample under analysis.

FIG. 2. The solubility method applied under conditions in which the solid phase is preformed and allowed to dissolve *in* the sample instead of being formed by precipitation *from* the sample. The term *R* represents that portion of the sulfonic acid added as such (or as a sodium salt) to the second solution, in contradistinction to the amount of the same **SUI**fonic acid contributed to the solution by the salt (S) as it dissolves.

By this procedure it is possible to carry out an analysis without quantitative precipitation. The development of the method can be carried still one step further, however, and the precipitation requirement completely eliminated. The solubility-product principle may be applied under conditions in which measurement is made of the amount of an amino acid salt going *into* solution in aliquots of the sample under analysis **(2).** In this case no isolation of the amino acid from the hydrolysate is required. One might say that the method in this form permits the quantitative determination of an amino acid in solution from measurements of the extent to which its presence depresses the solubility of one of its pure salts. This technique possesses several advantages from the experimental standpoint. The working procedure of the method in its present applications to the analysis of protein hydrolysates is illustrated in figure **2.**

The determination of *l*-leucine with 2-bromotoluene-5-sulfonic acid may be

discussed as an example of this procedure. An excess of a known pure sample of 1-leucine 2-bromotoluene-5-sulfonate is added to an aliquot of the hydrolysate containing an unknown quantity *(A* millimoles) of I-leucine, and the salt allowed to dissolve to the point of saturation. The binary salt dissolved $(S_1$ millimoles) contributes equimolar amounts of 1-leucine and 2-bromotoluene-5-sulfonic acid to the solution. At equilibrium the total *l*-leucine present is $A + S_1$ and the total sulfonic acid is S_1 , giving the solubility product $K_1 = S_1(A + S_1)$. Simultaneously, the I-leucine salt is added to a second aliquot of the hydrolysate in which *R* millimoles of sodium 2-bromotoluene-5-sulfonate has previously been dissolved. At equilibrium a smaller quantity of the amino acid salt (S_2) dissolves in this second solution as a result of the common-ion effect of the added sulfonic acid. The total *l*-leucine in solution is $A + S_2$ and the total sulfonic acid is $R + S_2$, giving the solubility product $K_2 = (R + S_2)(A + S_2)$. If the two solubility products are equal, solving for *A* gives the amount of 1-leucine in the sample in terms of the three known quantities S_1 , S_2 , and R (equation I, figure **2).** This equation is applicable to the amino acid determinations in which *K* is found to be constant. In other determinations, such as that of glycine as nitronaphthalenesulfonate, solubility products of satisfactory constancy are not obtained. In these cases K_2 is consistently greater than K_1 . This deviation, however, is small and reproducible, and, as will be demonstrated later, the introduction of a factor *F* in the equation adequately accounts for this increase (equation 11, figure **2).**

Experimentally, the method involves the measurement of the solubility of a given solid phase in two aliquots of the solution of unknown amino acid content. This measurement can be made with accuracy gravimetrically. A weighed amount of the amino acid salt is added to a sample of the solution. The solubility is determined by difference after equilibrium has been attained at 0°C. The separation of the two phases without temperature change is accomplished by centrifugation at 0° C. through sintered glass (6). The undissolved amino acid salt, together with a small amount of adhering solution, is retained by the sintered glass. The filter assembly is weighed before and after drying and the loss on drying provides a measure of the adhering solution, making it possible to calculate the true net weight of the solid phase.

The application of the solubility method under these conditions permits the purity of the solid phase to be readily insured. The initial amino acid salt represents a known preparation of established purity. As it dissolves, however, it contributes to the solution a quantity of sulfonic acid which might cause the precipitation of some other salt from the unknown. To establish the absence of interference, the first step is the addition of slightly more than $R + S_2$ millimoles of sulfonic acid to a separate sample of the hydrolysate. This concentration of sulfonic acid should not be sufficient to cause the precipitation of any compound from the unknown, and a clear solution should be obtained. The second and final check is made by adding a small amount of the pure amino acid to the *A* millimoles present and determining the precision of recovery.

This proof of purity of the solid phase establishes as a corollary principle the *specificity* of the determination. It is obvious that if *l*-leucine is the only amino acid in the solid phase, then it is the only amino acid taking an active part in the equilibrium between solid phase and solution, and the analysis is fundamentally specific for *l*-leucine.

It will be observed that the solubility method does not demand reagents of high selectivity. For example, given a mixture of approximately equimolar quantities of leucine, phenylalanine, and tyrosine, it would theoretically be possible to analyze for each of the three components with the aid of a single reagent. This could be done provided the leucine, phenylalanine, and tyrosine salts had about the same solubility product. For the determination of leucine, the leucine salt would be employed as solid phase; for phenylalanine, the phenylalanine salt; and for tyrosine, the tyrosine salt. This analysis by the solubility method would be possible even though the reagent would not be considered selective for any one of the three amino acids in the usual sense of the word.

TABLE **1**

Determinations of *1-leucine* Reagent: sodium 2-bromotoluene-5-sulfonate s_1 romotoluene-5-sulf
 $S_1^2 - S_2(R + S_2)$

l -Leucine =		$S_1 - S_2(K + S_2)$			
			$(R + S_2) - S_1$		

 $* K_1$ and K_2 have been calculated on the basis of moles per liter.

111. APPLICATION TO THE DETERMINATION OF AMINO ACIDS

A, Leucine

The establishment of the precision of the solubility-product principle for a given amino acid and reagent combination is illustrated for 1-leucine 2-bromotoluene-5-sulfonate in table 1. The determinations are carried out in hydrochloric acid solution $(1.0 \pm 0.2 \text{ normal})$. The amount of sodium 2-bromotoluene-5-sulfonate added to the second aliquot in each case is about 0.1 millimole. The agreement between K_1 and K_2 is satisfactory, and equation I gives accurate recoveries for 1-leucine. In protein hydrolysates, of course, the solubility product will be higher than in a pure leucine solution. When the value of the solubility product is increased from 9×10^{-4} up to 12 or 13 $\times 10^{-4}$ by the addition of an organic solvent, the analysis is equally satisfactory. When the solubility product is similarly increased by the addition of arginine, glutamic acid, glycine, and tyrosine, the accuracy is maintained. An important char-

428 STANFORD MOORE, WILLIAM H. STEIN **AND MAX BERGMANN**

acteristic of the method is illustrated in these data. As has been pointed out, in contrast to the usual gravimetric techniques, the solubility method is in principle independent of the absolute magnitude of the solubility of the salt in the solvent or unknown hydrolysate. The relative constancy of the two *K's* in any individual leucine determination is the fundamental essential.

From the stereochemical standpoint, the analysis under these conditions is specific for *l*-leucine, and small amounts of the *d*-isomer, if present, do not enter into the equilibrium. Analysis for d-leucine can be made by use of the d -salt as solid phase. Analysis for both of the isomers in mixtures of *d-* and I-leucine may be accomplished by the use of the racemic mixture dl-leucine 2-bromotoluene-5-sulfonate as solid phase. Theoretically, if a dl-salt is a racemic *mixture,* not a racemic *compound,* K_{d} is equal to twice K_l in the ideal case. This relationship holds for the *dl*-leucine salt of this reagent, but does not hold, of course, for all dl-amino acid salts, many of which are racemic compounds.

B. Glycine

in table 2. In the first example, K_1 is 7.33×10^{-4} , and when about 0.1 millimole **A** series of analyses for glycine, using **nitronaphthalenesulfonic** acid, is listed of nitronaphthalenesulfonic acid has been added, K_2 is 7.83 \times 10⁻⁴. If solved by equation I, as used for leucine, this analysis would give a recovery of about **80** per cent, the low value being the result of the small but measurable difference between K_1 and K_2 . In the first determination it will be observed that K_1 is 93.6 per cent of K_2 . As the magnitude of K_1 is increased over a relatively wide range, namely, from 7×10^{-4} up to 17×10^{-4} , the ratio K_1/K_2 remains satisfactorily constant. This relationship holds also for analyses in the presence of glucose, of furfural, and of arginine, glutamic acid, leucine, and tyrosine. If an average value of 0.939 is taken for the ratio K_1/K_2 and this figure substituted for *F* in equation 11, the recoveries of glycine calculated by this equation all fall within the desired range of precision.

The change in the solubility product occurring in this determination of glycine as nitronaphthalenesulfonate may be considered to represent more closely the general case rather than an exception to the rule. The constancy of $K (K_1/K_2 = 1.00)$ in the leucine analysis is an experimental fact. It is known, however, to be the result of compensating factors which in the glycine case fail to compensate as completely. By independent experiments it can be shown that changes in amino acid concentration, sulfonic acid concentration, or sodium chloride content (if the sodium salt of the reagent is used) all exert small but measurable effects on the solubility product. But the addition of a given amount of a sulfonic acid reagent, a factor which can be controlled, produces nearly the same series of concentration changes in each case. It has been shown for the leucine and glycine determinations that the ratio K_1/K_2 can be maintained constant over a range of K_1 values. A constant K_1/K_2 ratio is equally as satisfactory mathematically as a constant solubility product. These effects do impose the requirement that for accurate analysis the concentration of the protein hydrolysate must be adjusted so that the initial amino acid concentration is close to the value used in these experiments. As is true of many analytical methods, the first analysis on an unknown generally serves to indicate the optimum working range for subsequent accurate determination.

 $* K_1$ and K_2 have been calculated on the basis of moles per liter.

TABLE 3

Determination of glycine and l-leucine in a synthetic mixture of amino acids

The validity of equations I and II for leucine and for glycine has been checked on a number of known mixtures other than those already listed. One example is summarized in table 3. A mixture composed of thirteen constituents, approaching in complexity a protein hydrolysate, has been analyzed with satis-

429

factory precision for both leucine and glycine by the solubility method. The final step is from synthetic mixtures of this type to protein hydrolysates. Two examples may be sufficient to demonstrate the practical application of the method (table 4). In the analysis for glycine and leucine in hydrolysates of silk fibroin and egg albumin, both the reproducibility and the recovery are comparable to that attained with synthetic mixtures. It should be pointed out, however, that experimentally the values in table **4** are based on the amino acid contents of the *hydrolysates.* It is incorrect to conclude that each value necessarily represents the total glycine or leucine content of the protein preparation employed. Before such a conclusion is justified, further information is required on the accuracy of the procedures generally used for the hydrolysis of proteins and (except in the case of glycine) on the extent of racemization during hydrolysis. The solubility method is being employed to investigate both of these problems, each of them being fundamental to the final determination of protein composition.

SILK FIBROIN		EGG ALBUMIN					
	ber cent		per cent				
	44.0		9.11				
	44.6		9.12				
	43.2		9.00				
	43.3	l -Leucine	8.95				
	44.4		9.11				
	44.6		9.12				
	Av. 44.0		Av. 9.07				
Recovery on protein hydrolysate		Recovery on protein hydroly- sate plus added <i>l</i> -leucine 100.5					

TABLE 4 Determination of glycine and 1-leucine in protein hydrolysates

Of the amino acids which come within the present scope of the method, leucine and glycine have been singled out for illustration in this discussion. Leucine, with its lack of characteristic functional groups, has previously been indeterminable. Glycine has similarly been a difficult amino acid to estimate, especially when present in relatively small amounts. The method to date has also been applied to phenylalanine, tyrosine, and proline. It may be mentioned that this analytical approach is not limited to amino acids, being in theory applicable to the determination of any substance possessing an acidic or basic group which forms sparingly soluble, dissociable salts.

IV. REAGENTS

The practical value of the solubility method is a function of the availability of reagents. The development and selection of reagents for a given problem is a special study in itself. Only a brief outline can be given here of the manner in which this problem has been approached. The investigation has involved primarily aromatic sulfonic acids *(5),* although other types of reagents have also

been shown to be applicable (3). A few sulfonic acids have already been recommended, in the earlier literature (1, 13), for use in amino acid analysis, flavianic acid $(7, 12)$ being the outstanding example. The general tendency of aromatic

TABLE 5 Solubility products of amino acid salts of aromatic sulfonic acids (5) Approximate molar solubility products determined in N HCl at 0° C. Where no value is given, the solubility product of the salt was found to be greater than about 4×10^{-2}

	SULFONIC ACIDS											
AMINO ACIDS	p-Chlorobenzene-	p-Bromobenzene-	p-Iodobenzene-	2,5-Dibromobenzene-†	2, 5-Dichlorobenzene-	3,4-Dichlorobenzene-	4-Nitrotoluene-2-	2-Bromotoluene-5-	2,6-Diiodophenol-4-	y-Tosyl-p-phenol-	(Flavianic acid) 2, 4- Dinitro-1-naphthol-7-	5-Nitronaphthalene-1-
	$X10^{-4}$											
l -Alanine l -Arginine [*] l-Aspartic acid l -Cysteine l -Cystine* l -Glutamic acid Glycine l -Histidine* l -Hydroxyproline	100	50	20	0.2 0.2	10 ¹	0.008 0.002	280	20 $\mathbf{2}$	0.08 0.002 85	0.06 9 0.06 0.01	0.0002 0.01 25 0.05	0.02 5 0.6 9
l -Isoleucine l -Leucine l -Lysine*	330	120	28	57	100	26		9	17	16 0.3 0.1	3 0.01	130 0.2
dl -Methionine dl -Phenylalanine l -Proline	100 120	100 35	16 $\overline{4}$	0.3	50 $4\ddagger$	16 5	20	10 [†]	2 68	0.8 0.1	1‡	İ
dl -Serine l -Tyrosine l -Valine	51	190		17	82		28		8	$\overline{2}$ 9	0.03	

* The arginine, histidine, lysine, and cystine salts contain 2 moles of sulfonic acid per mole of amino acid. For these ternary salts, $K = [\text{sulfonic acid}]^2 \times [\text{amino acid}]$. Therefore, a comparison of the solubility product of a ternary salt with that of a binary salt is not a direct measure of their relative solubilities. For example, the *l*-arginine and *dl*-phenylalanine salts of 2,5-dibromobenzenesulfonic acid have about the same solubility products. A saturated solution of the ternary arginine salt, however, contains about 0.02 mole of arginine per liter, whereas a saturated solution of the binary phenylalanine salt contains only 0.005 mole of phenylalanine per liter.

† The salts of l-cysteine, l-cystine, and dl-methionine were not investigated in these cases.

t l-Phenylalanine salts.

sulfonic acids to form sparingly soluble salts with nearly all amino acids, however, has not been recognized until recently. Herein lies a reservoir of reagents potentially useful for the isolation, purification, and determination of amino

431

acids. In order to choose a reagent for application to any of these problems it is necessary to know first the solubility of the salts which the reagent may form with the major constituents of protein hydrolysates. The type of data which has been gathered for this purpose is illustrated in table 5. Only a few examples of the many reagents investigated are included in this list. For leucine, 2 bromotoluene-5-sulfonic acid can be seen to possess the desired degree of specificity for most present analytical applications; for phenylalanine, 2,5-dibromo- and **2,5-dichloro-benzenesulfonic** acids; for glycine, 5-nitronaphthalene-1-sulfonic acid. From data of this type, reagents for additional members of the amino acid series have been developed. Sulfonic acids also may be utilized for the isolation of amino acids and peptides from protein hydrolysates on a preparatory scale. One example is the preparation of *l*-serine and *l*-alanine from silk fibroin (10) with the aid of p -hydroxyazobenzene- p' sulfonic acid and azobenzene-p-sulfonic acid.

V. CONCLUSION

In the beginning of this discussion it was mentioned that one of the fundamental problems of protein chemistry is the establishment of the complete amino acid composition of the proteins. It may be asked, therefore, to what extent the method outlined in this review may aid in the future accomplishment of this task. It should be emphasized that reagents remain the limiting factor in this type of analytical approach. It is true that the reagent requirements have been lowered, that many additional ones have been developed, and that more quantitative knowledge can be gained today than was possible **a** few years ago. Yet the task of finding satisfactory reagents for *all* of the amino acids must be recognized as a formidable problem. The extension of the method to additional amino acids will involve considerable further work in both the analytical and the reagent fields. At the present date the method can be summarized as a new and accurate means for the determination of a *number* of the constituent parts of the protein molecule and, as such, is serving as a tool in the general approach to a fuller understanding of protein chemistry.

REFERENCES

- (1) ACKERMANN, D.: Z. physiol. Chem. 226,46 (1934).
- (2) BERGMANN, M., MOORE, S., AND STEIN, W. H.: Unpublished data.
- (3) BERGMANN, M., AND STEIN, W. H.: J. Biol. Chem. 128, 217 (1939).
- (4) BERGMANN, M., AND STEIN, W. H.: J. Biol. Chem. 129,609 (1939).
- *(5)* DOHERTY, D. G., STEIN, W. H., AND BERGMANN, M.: J. Biol. Chem. 136,487 (1940).
- (6) ING, H. R., AND BERGMANN, M.: J. Biol. Chem. 129,603 (1939).
- (7) KOSSEL, A., AND GROSS, R. E.: 8. physiol. Chem. **136,** 167 (1924).
- (8) RITTENBERG, D.: Paper presented at the Symposium on Physiochemical Methods in Protein Chemistry, which was held at the 102nd Meeting of the American Chemical Society, Atlantic City, New Jersey, September 8-12, 1941.
- (9) RITTENBERG, D., ANDEOSTER, G. L.:J. Biol. Chem. 133,737 (1940).
- (10) STEIN, W. H., MOORE, S., AND BERGMANN, M.: J. Biol. Chem. 139, 481 (1941).
- (11) STEIN, W. H., NIEMANN, **C.,** AND BERGMANN, M.: J. Am. Chem. SOC. **60,** 1703 (1938).
- (12) VICKERY, H. B.: J. Biol. Chem. 132,325 (1940).
- (13) ZIMMERMANN, W.: 2. physiol. Chem. **188,** 180 (1930).