SOYBEAN PROTEIN NOMENCLATURE: A PROGRESS REPORT

W. J. Wolf, Chairman, Soybean Protein Nomenclature Committee,

Oilseeds Division
Northern Regional Research Laboratory
Peoria, III. 61604.

RECENT YEARS HAVE witnessed an increase in food uses of soybean proteins and a need for more information about their physical and chemical properties. As a result, research on soybean proteins is increasing and new components are being isolated and characterized. Difficulties in designating the newly isolated components and in distinguishing them from previously characterized proteins led to the formation in 1967 of a Soybean Protein Nomenclature Committee under sponsorship of the Oilseeds Division. In addition to the Chairman, the Committee consists of the following members: Dr. Michael Tombs, Unilever Research Laboratory, Sharnbrook, England; Dr. Ikunori Koshiyama, Noda Institute for Scientific Research, Noda, Japan; Dr. Kazuo Shibasaki, Tohoku University, Sendai, Japan; Dr. Nicholas Catsimpoolas, Central Soya, Chicago, Illinois; and Dr. Paul Melnychyn, Carnation Company, Van Nuys, California.

Nomenclature of soybean proteins is not a serious problem as yet, because of the limited amount of work done in this area. This situation is likely to change, however, as more workers enter the field and as new techniques are applied to the fractionation and characterization of soybean proteins. For example, gel electrophoresis (1-3) can quickly lead to problems of nomenclature for various reasons. Several of the soybean proteins have quaternary structures made up of dissimilar subunits. Gel electrophoresis may be run under a wide range of conditions including variations in gel media, pH, and buffer ions, plus the presence or absence of such dissociating reagents as urea and detergents, or with or without reducing agents. Changes in many of these variables can disrupt the quaternary structure and alter the gel pattern. In many instances it may be difficult to differentiate between individual subunits and subunit aggregates. Nevertheless, research workers want to designate the various bands observed, if for no other reason than to make it easier to describe their results. However, if different workers use different names for the same protein bands, confusion results. There is, therefore, a need for a nomenclature system that will, it is hoped, prevent a multiplicity of terminology.

This report summarizes the work of the committee so far. Past nomenclature and the various proposals under consideration are reviewed briefly.

Past Soybean Protein Terminology
Names used in the past for soybean proteins are listed in Table I. The list is not all-inclusive. Enzymes are not given, since they present no special problems. Terms such as curd and whey are omitted because they have industrial connotations and refer to gross fractions rather than to individual proteins.

Meissl and Böcker (4) introduced the names casein and albumin. Since casein is well established in the milk protein field, it should be discarded. Albumin is a general term for a class of proteins and thus there is no compelling reason for keeping it either. Osborne and Campbell (5) introduced four names in their studies on soybean proteins: Glycinin is a classical name derived from the Latin name for soybeans and should be retained, although it may need redefining. Phaseolin should be discarded, since Osborne used it earlier to name the major globulin of Phaseolus vulgaris (9). Legumelin was also used by Osborne for a protein fraction from a number of seeds, including peas (10); therefore the use of legumelin as a name for a soybean protein should be discontinued. Proteose is a term similar to albumin in that it refers to a class of proteins rather than to a specific protein; it therefore seems preferable to discard it too.

Soybean trypsin inhibitor and soybean hemagglutinin are names reflecting biological activities. These names are adequate and have the advantage of being applicable to similar proteins from other seeds merely by changing soybean to the appropriate seed name.

![Fig. 1. Effects of ionic strength on the ultracentrifuge pattern for water-extractable soybean proteins at pH 7.6.](image)

The terminology used most extensively in the past 10 years was introduced by Naismith (7) and is based on sedimentation coefficients of 2, 7, 11, and 15S. Although the ultracentrifuge resolves soybean proteins into four fractions, there are more than four components present. Gel filtration reveals at least seven or eight components, several of which are in the 6S to 8S region (11). In the
ultracentrifuge the 6S to 8S components are normally observed only as a single 7S fraction.

Another difficulty with using ultracentrifuge terminology is that sedimentation properties change with conditions used in the analysis. Urea concentration, extremes of pH, heating, and variations in ionic strength all change sedimentation patterns (12-14).

An example of the effect of ionic strength at pH 7.6 is shown in Fig. 1. Analysis of the water-extractable proteins from defatted meal at 0.5 ionic strength yields four fractions in accordance with Naismith’s terminology. At 0.1 ionic strength the 7S peak decreases and a new peak, designated 9S, is apparent. The new peak is formed by reversible dimerization of a portion of the 7S fraction observed at 0.5 ionic strength. This behavior is expressed by the equilibrium reaction diagramed in the lower part of Fig. 1. The reaction is a characteristic property of a protein isolated by Roberts and Briggs (15) and by Koshiyama (16). This protein is referred to as 7S globulin but accounts for only 60% of the total 7S fraction at 0.5 ionic strength when areas under the ultracentrifugal patterns of Fig. 1 are measured (17). Use of 7S to designate more than one protein thus becomes confusing.

Osborne and Campbell (5) applied the name glycinin to a protein preparation which they believed was the major globulin of soybeans. Ultracentrifugation, electrophoresis, and hydroxylapatite chromatography (7,18,19) indicate that glycinin as defined by Osborne and Campbell is heterogeneous. A typical ultracentrifugal pattern reveals at least four components in glycinin. These results with glycinin emphasize one of the problems of soybean protein nomenclature. Should the entire protein preparation be called glycinin, even though it is not a single protein? Or, should each of the proteins present be given a separate name?

Proposals Under Study

Six proposals are under study as possible solutions to the nomenclature problem. Five of these proposals are based on the ultracentrifuge behavior of soybean proteins at pH 7.6 and 0.5 ionic strength. Proposal I is outlined in Fig. 2. The name glycinin is assigned to the 11S component because it is the major protein observed in the various glycinin preparations reported in the literature. The 11S protein also has a nitrogen content (19) that is close to the values reported for glycinin by Osborne and Campbell (5). Finally, this assignment would seem to be in the spirit of the original intent of Osborne and Campbell, that glycinin is the major globulin of soybeans. Proposal I also assigns the name conglycinin to the 7S globulin (15,16) discussed earlier. The prefix con- is used to indicate that this protein occurs with glycinin in the globulin fraction by analogy with the respective terms conarachin and concanavalin for the proteins of peanuts and jackbeans. Figure 2 applies this terminology to the ultracentrifugal patterns seen in Fig. 1. At 0.5 ionic strength the 7S fraction is recognized as consisting of conglycinin and other proteins, whereas at 0.1 ionic strength conglycinin is observed as the dimer, namely the 9S peak. This proposal provides no guidelines for naming any of the other ultracentrifuge fractions, and this omission is one of its weaknesses.

Proposal II, summarized in Fig. 3, suggests a solution to the problem of designating the proteins that have not yet been isolated and characterized. In proposal II, the 11S protein is again called glycinin, but the 2, 7, and 15S fractions collectively would be known as conglycinin. As specific proteins are isolated and characterized they would be assigned a letter from the Greek alphabet in the order that they are reported in the literature. For example, the 7S globulin is called alpha-conglycinin. Figure 3 demonstrates application of proposal II to the ultracentrifugal pattern at 0.1 ionic strength where alpha-conglycinin is present as a dimer. The cross-hatched areas and dotted lines in Fig. 3 illustrate how this nomenclature system could be extended in the future.
11S = Glycinin
7S = Conglycinin -
   -alpha-Conglycinin -
   -beta-Conglycinin -
2S = Conglycinin -
   -alpha-Conglycinin -
15S = Conglycinin -
   -alpha-Conglycinin -

Conglycinin -
   -alpha-Conglycinin - dimer

15S = Conglycinin -
   -alpha-Conglycinin -

2S = delta-Glycinin
7S = gamma-Glycinin
11S = beta-Glycinin

Fig. 4. Nomenclature proposal III.

letters in the order of their isolation. Likewise, the 15S fraction would become conglycinin-III, and proteins within the group would be assigned the letters alpha, beta, etc. Figure 4 illustrates the use of proposal III on the ultracentrifugal pattern for 0.1 ionic strength. Under these conditions conglycinin-I includes both the 7S and 9S peaks, but the latter peak is the dimer of alpha-conglycinin-I.

Proposal IV is represented in Fig. 5. It suggests calling the entire globulin fraction glycinin, but designating the sedimenting peaks by Greek letters in order of decreasing sedimentation rates. As subfractions are isolated they are assigned subscript numbers in the order of their isolation. The ultracentrifugal pattern at 0.1 ionic strength (Fig. 5) shows how proposal IV would be applied to this condition. The 7S and 9S fractions are collectively called gamma-glycinin, but the 9S peak is called gamma - glycinin dimer.

Fig. 5. Nomenclature proposal IV.

Proposal V, although similar to proposal II, requires that a protein be adequately characterized before it is given a specific name. Proposal V consists of two parts.

1. Glycinin designates the protein that:
   (a) Occurs in the protein bodies of mature cotyledons;
   (b) Is one of the major proteins of soybeans;
   (c) Has a molecular weight of about 350,000 but can exist in a number of other forms depending on conditions;
   (d) Has N-terminal residues of glycine, phenylalanine, and (iso) leucine; and
   (e) Contains little or no carbohydrate. (These are the properties of the protein designated 11S in the ultracentrifugal terminology.)

2. All other proteins are called conglycinin and are distinguished from one another by numerical additives. For example, conglycinin 1 could designate trypsin inhibitor and conglycinin 2 could serve for hemagglutinin.

In this system, before a component is named it must meet the following criteria:

(a) characterization by electrophoresis, immunochecmistry, chromatography, and ultracentrifugation; (b) chemical analysis for N-terminal and carbohydrate content; and (c) estimation of its content in soybean protein.

Proposal VI suggests calling all the proteins glycinin, as in proposal IV, but uses the ultracentrifugal terminology instead of Greek letters, which would be reserved for later designation of protein subunits. The proposal is summarized as follows, with Naismith's designations on the left and the proposed terminology on the right of the equality:

2S = 2S glycinin I, 2S glycinin II
7S = 7S glycinin I, 7S glycinin II
11S = 11S glycinin

The ultracentrifugal term indicates the relation of a protein to the ultracentrifugal pattern for the unfractionated protein mixture (Fig. 1). The Roman numeral differentiates the various proteins in each ultracentrifugal fraction.

Concluding Comments

None of the proposals under consideration are based on gel electrophoresis, although this technique has been most useful in other protein systems. There are now at least five gel-electrophoretic systems that have been applied to soybean proteins (1-3,20,21), but variations among the systems have given diverse results. More work is required before the relation between ultracentrifuge fractions and gel bands is understood. Ultimately, it may be possible to develop a nomenclature system based on gel electrophoresis. Gel electrophoresis should be especially useful in detecting and identifying subunits of proteins with quaternary structures.

The nomenclature system that is adopted should also include the relative mobility, $R_m$, of each component as determined by gel electrophoresis. The $R_m$ value is the ratio of the mobility of a given protein to the mobility of a reference protein with a high intensity of staining and high mobility. The reference protein would be given an arbitrary $R_m$ of 1.00 as was done with trypsin inhibitor by Eldridge et al. (20). An example of this notation as applied to proposal IV is beta-glycinin (0.87) to designate a component of beta-glycinin which has a $R_m$ of 0.87 under clearly specified conditions of gel electrophoresis.

Additional proposals or comments regarding the proposals under consideration are invited from interested persons. The Committee is still seeking alternatives to the six proposals under study.
NOMENCLATURE

(Continued from page 78)

Literature Cited


[Received December 5, 1968. Accepted December 23, 1968]

Reprinted from CEREAL SCIENCE TODAY, March, 1969