

Structural similarities among the high molecular weight protein fractions of oilseeds

V. PRAKASH* and M. S. NARASINGA RAO

Central Food Technological Research Institute, Mysore 570 013, India

MS received 5 June 1987; revised 16 December 1987

Abstract. Data on the physico-chemical properties of proteins from soybean, groundnut, sesame seed, sunflower seed, safflower seed, mustard seed, rapeseed and cotton seed are fairly extensive. An examination of the available data on high molecular weight proteins suggests that there are similarities in many of their properties. In this report the similarity in amino acid composition, size and shape, molecular weight, secondary structure, subunit composition, association-dissociation at high and low pH, stability towards denaturants, hydrolysis by enzymes and quaternary structure of the high molecular weight proteins is discussed. Based on these similarities a model has been proposed for the association-dissociation, denaturation and reassociation behaviour of the high molecular weight proteins of oilseeds.

Keywords. Structural similarity; oilseed proteins; high molecular weight proteins; oilseeds; model for seed protein; association-dissociation; denaturation.

Introduction

Oilseed proteins have in recent years attracted the attention of scientists in view of their importance in the plant system and their unique properties. Of the two major classes of plant proteins, namely, functional proteins and reserve proteins, the latter have been studied in great detail (Prakash and Narasinga Rao, 1986). Pernollet and Mosse (1983) have described these proteins as tropic secretory proteins able to associate and exhibit relative multiplicity and polymorphism. Their deposition in the seed is shown to be close to phosphate reserves in the cell. This article highlights the similarities in the composition and properties of these proteins. An attempt has been made to correlate structure with function in the seed, based on their solution properties.

The literature on the physico-chemical properties of some of the oilseed proteins such as those from soybean, groundnut, sesame seed, sunflower seed etc. is fairly extensive (Prakash and Narasinga Rao, 1986). In general the proteins of oilseeds can be categorised into two groups, namely, the high molecular weight (M_r) protein fraction and the low M_r protein fraction. Since extensive data is available at present on the high M_r proteins, the discussion is confined to this class.

If one carefully examines the reported data on protein fractions of various oilseeds a great similarity is apparent in (i) the number of fractions and (ii) their sedimentation coefficients. Table 1 summarises available data. All oilseeds consist predominantly of 4 protein fractions. These can be designated for the purpose of discussion as 2S (low M_r protein fraction), 7S (medium M_r protein fraction, corresponds to vicillin of legumes), 11S (high M_r protein fraction, corresponds to legumin of legumes) and 15-18S ('polymer' resulting from possible aggregation of 2S, 7S, or 11S or may be

*To whom all correspondence should be addressed.

Abbreviations used: M_r , Molecular weight; SDS, sodium dodecyl sulphate; PAGE, Polyacrylamide gel electrophoresis; GuHCl, guanidine hydrochloride.

Table 1. Sedimentation coefficient ($S_{20,w}$) values of the various protein fractions from several oilseeds.

Source	Protein fraction			Polymer
	Low M_r	Medium M_r	High M_r	
Soybean	2	7	11	15
Groundnut	2	7	11	18
Sesame seed	2	7	11	15
Sunflower seed	2	7	12	16
Mustard seed	2	7	12	—
Rapeseed	2	7	12	—
Cottonseed	2	7	11	18
Safflower seed	2	7	12	17

Data from Prakash and Narasinga Rao (1986).

inherently present in the seed). It is observed that substances such as trypsin inhibitors, hemagglutinins, polyphenols, glucosinolates, colour and bitter principles, which are inherent constituents of some of the oilseeds, are generally associated more with the 2S or 7S fraction. However one cannot rule out such association being an artifact of isolation procedures.

The high M_r protein fraction (10–12S) is the major fraction in groundnut (Prakash and Narasinga Rao, 1986), sesame seed (Prakash and Nandi, 1978), sunflower seed (Schwenke *et al.*, 1974) and safflower seed (Latha and Prakash, 1986). On the other hand, in soybean, mustard seed, rapeseed and cottonseed, it is present to the extent of only 20–30% (Prakash and Narasinga Rao, 1986). However, in these seeds also, it forms a significant proportion of the total proteins.

In this report the similarity in amino acid composition, size and shape, M_r , secondary structure, subunit composition, association-dissociation at high and low pH, stability towards denaturants, hydrolysis by enzymes and quaternary structure of the high M_r proteins from various oilseeds will be discussed.

Amino acid composition

The amino acid composition of the high M_r proteins from 7 different seeds are given in table 2. All the high M_r proteins are rich in acidic amino acids, especially glutamic acid, and aromatic amino acids, and are low in lysine. Mosse (1973) and Pernollet and Mosse (1983) have reviewed the intraspecific variation in amino acid composition of legume seeds. In this review we have analyzed the amino acid data in terms of hydrophobicity and other related indices.

Hydrophobicity and related indices

The fundamental structural parameters of proteins which depend only on amino acid composition have been shown to be Waugh's NPS or frequency of non-polar side chains (Waugh, 1956; Bigelow, 1967), Fischer's P or the ratio of the volume occupied by polar residues to that occupied by non-polar residues (Bigelow, 1967; Fisher, 1964) and Bigelow's average hydrophobicity (Bigelow, 1967). These are based on Tanford's free energies of transfer of amino acid side chains from an organic

Table 2 Amino acid composition of the high M_r proteins from various oilseeds.

Amino acid	Glyc- nin ^d	Ara- chin ^b	α -glob- ulin ^c	Helia- nthin ^d	Bras- sin ^e (M)	Bras- sin ^f (R)	Carmin ^g	High M_r gliadin from wheat ^h	Rice globulin ⁱ
Aspartic acid	106	111	84	107	53	83	110	17	68
Threonine	44	19	41	36	28	38	27	22	29
Serine	74	57	59	51	30	56	62	60	63
Glutamic acid	169	171	155	197	134	162	181	324	91
Proline	50	40	21	52	ND	43	46	130	57
Glycine	64	59	90	85	75	81	92	42	103
Alanine	47	40	71	69	34	54	67	23	128
Valine	43	34	46	63	48	39	38	35	63
Methionine	9	1	20	19	22	14	10	11	18
Half cystine	7	8	7	11	ND	ND	10	11	0
Isoleucine	45	25	32	49	52	32	23	29	27
Leucine	56	60	63	68	9	66	57	64	58
Tyrosine	24	28	24	20	15	19	32	15	31
Phenylalanine	34	31	34	48	23	33	21	35	22
Lysine	33	26	16	19	17	23	17	9	20
Histidine	17	17	20	23	13	21	17	12	12
Arginine	45	92	91	66	ND	39	64	22	70
Tryptophan	7	11	11	10	9	9	11	ND	7

The amino acid compositions of gliadin from wheat and rice globulin are given for comparison.

Numbers are residues per molecule of protein. ND, Not determined.

^aShvarts and Vaintraub (1967); Tombs (1965); ^cPrakash and Nandi (1978); ^dSchwenke *et al.* (1974); ^eGururaj Rao and Narasinga Rao, (1981); ^fGill and Tung (1978); ^gLatha and Prakash (1986); ^hBietz and Wall (1973); ⁱTecson *et al.* (1971).

environment to an aqueous environment. In table 3 the average hydrophobicity, NPS and P values of high M_r proteins are listed. Also included in table 3 are data on

Table 3. Calculated values of hydrophobicity and related parameters of the high M_r proteins of various oilseeds.

Protein	HQ ^a	NPS ^b	P ^c
Ovalbumin	1110	0.34	0.92
Fibroin (silk)	440	0.02	0.07
Glycinin	782	0.30	1.28
Arachin	860	0.29	1.73
α -Globulin	872	0.26	1.36
Helianthin	832	0.26	1.25
Brassin (M)	962	0.31	1.03
Brassin (R)	900	0.30	1.00
Gossypin	804	0.24	1.00
Carmin	824	0.26	1.69
Poppyverin	878	0.28	1.49
Linin	881	0.27	1.22
Average \pm SD	860 \pm 50	0.28 \pm 0.02	1.31 \pm 0.25

Values for ovalbumin and fibroin are included for comparison.

^aaverage hydrophobicity.

^bFrequency of non-polar side chains.

^cRatio of the volume occupied by polar residues to that occupied by non-polar residues.

ovalbumin and fibroin for purposes of comparison. These are typical animal proteins, ovalbumin globular in shape and fibroin with an extended structure. The values were calculated by the methods described in the papers cited above.

Two interesting observations emerge from this analysis. Firstly, the values of these parameters are nearly the same for all the high M_r proteins except for gossypin whose values of average hydrophobicity and NPS are all lower than the average value. Except for the value of P which ranges from 1 to 1.73 the values of the other parameters are remarkably close to each other indicating a general trend. This may be related to the fact that these proteins have a low proportion of α -helix and a high content of β - and aperiodic structure. Further, the values for the high M_r proteins do not fit any portion of the Bigelow plot of average hydrophobicity vs M_r (Bigelow, 1967). Similarly the NPS values of high M_r proteins are higher than those reported for other proteins (Waugh, 1956; Bigelow, 1967).

Secondary structure

The circular dichroic spectra of the high M_r proteins are generally characterised by a minimum around 208–212 nm with a shoulder around 224–228 nm. This again suggests that the high M_r proteins have a low proportion of α -helix and are rich in β - and aperiodic structure. Table 4 summarises the secondary structure data for the high M_r proteins. They all have less than 10% α -helix and 20–30% β -structure, the rest being aperiodic structure. However, attempts have not been made to calculate from the circular dichroism (CD) data the number of β -turns. Blake and Johnson (1984) have classified proteins into 5 classes based on secondary structure. These are defined in terms of the proportion s and arrangement of α -helix and β -sheet. The various classes are (i) all- α proteins, (ii) all β -proteins, (iii) $\alpha + \beta$ proteins., (iv) α/β

Table 4. Intrinsic viscosity, secondary structure, M_r , number of subunits and carbohydrate content of the high M_r proteins from various oilseeds.

Protein	Viscosity, η (ml/g)	Secondary structure (%)			$M_r \times 10^{-5}$	No. of subunits	Carbo- hydrate (%)
		α -Helix	β -Structure	Aperiodic			
Glycinin ^a	4.9	5	20	75	3.0–3.5	6	—
Arachin ^a	4.7	5	20	75	3.0–3.3	6	0.3
α -Globulin ^{a, b}	3.0	5	25	70	2.3–2.7	6	0.8
Helianthin ^c	3.6	2	28	70	3.0–3.5	6	0.4
Brassin (M) ^d	3.6	9	28	63	2.3–2.4	6(8)	1.0
Brassin (R) ^e	3.7	9	28	63	2.9–3.0	6	1.0
Gossypin ^f	4.0	5	20	75	2.2–2.5	6(5)	0.5
Carmin ^g	3.7	3	15	82	2.4–2.9	6	1.0
Poppyverin ^h	3.5	5	20	75	2.0–2.3	6	1.2
Linin ⁱ	3.1	3	17	80	2.5–3.0	6	0.2
Ribonuclease ^{j, k}	3.3	40	13	24	—	—	—
Collagen ^l	1150	—	—	—	—	—	—
Elastase ^k	—	7	52	26	—	—	—

Data for ribonuclease, collagen and elastase included for comparison.

^aPrakash and Narasinga Rao (1986); ^bPrakash (1985); ^cSchwenke *et al.* (1974); ^dGururaj Rao and Narasinga Rao (1981); ^eGill and Tung (1978); ^fReddy and Narasinga Rao (1984) and Li *et al.* (1979); ^gLatha and Prakash (1986); ^hSrinivas (1984); ⁱMadhusudan (1984); ^jTanford (1961); ^kProvencher and Glockner (1981).

proteins and (v) 'coil proteins'. The high M_r proteins of oilseeds do not seem to fit into any of these classes since they are rich in β -sheet and aperiodic or coil structure. A sixth class termed β + coil proteins, may be appropriate to describe these proteins.

Intrinsic viscosity

The intrinsic viscosities of the high M_r proteins from various oilseeds are given in table 4. They all have an intrinsic viscosity value between 3 and 5 ml/g. Based on Tanford's criterion for globular proteins all the high M_r proteins appear to be globular in shape. For comparison, the values for ribonuclease, which is a globular protein, and collagen, a highly asymmetric protein, are also shown in table 4. These data also indicate a generality among the high M_r proteins.

M_r

It should be pointed out that different workers have used different techniques such as the sedimentation-diffusion method, sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE), approach-to-equilibrium method in the analytical ultracentrifuge etc. to determine the M_r 's of the high M_r proteins. The values obtained by the different methods are not strictly comparable. However, they are in the range 2×10^5 to 3.5×10^5 . The estimation of M_r by a thermodynamically sound method such as by sedimentation equilibrium in a single solvent is clearly indicated. Only in a few cases has this technique been used (Prakash, 1985; Latha and Prakash, 1986).

Fluorescence

The fluorescence emission maximum of the high M_r proteins is around 320–330 nm (Prakash and Narasinga Rao, 1984). The fluorescence emission spectrum suggests that in these proteins the tryptophan residues are embedded in the interior of the molecule. This is compatible with a highly compact globular shape for the protein. However, all the proteins contain a fair amount of tyrosine. There are 7–11 tryptophans and 15–32 tyrosines per molecule of each protein (table 2). But tyrosine emission is not observed. Shifrin *et al.* (1971) and Teale (1960) have shown that tryptophan fluorescence dominates over tyrosyl fluorescence. The results suggest that there is similarity in the location and microenvironment of non-polar groups in the interior of the molecules of these proteins in spite of the fact that these proteins possess very little α -helical structure. It has been shown that the aromatic amino acids of α -globulin are in the subunit contact areas stabilizing the quaternary structure (Prakash, 1985).

Hydrolysis

The *in vitro* hydrolysis of the high M_r proteins by proteolytic enzymes such as trypsin, chymotrypsin, pepsin and papain has been well documented (Prakash and Narasinga Rao, 1984, 1986). Compared to casein these proteins are hydrolysed to a smaller extent. Further, the susceptibility to proteolysis varies among the proteins.

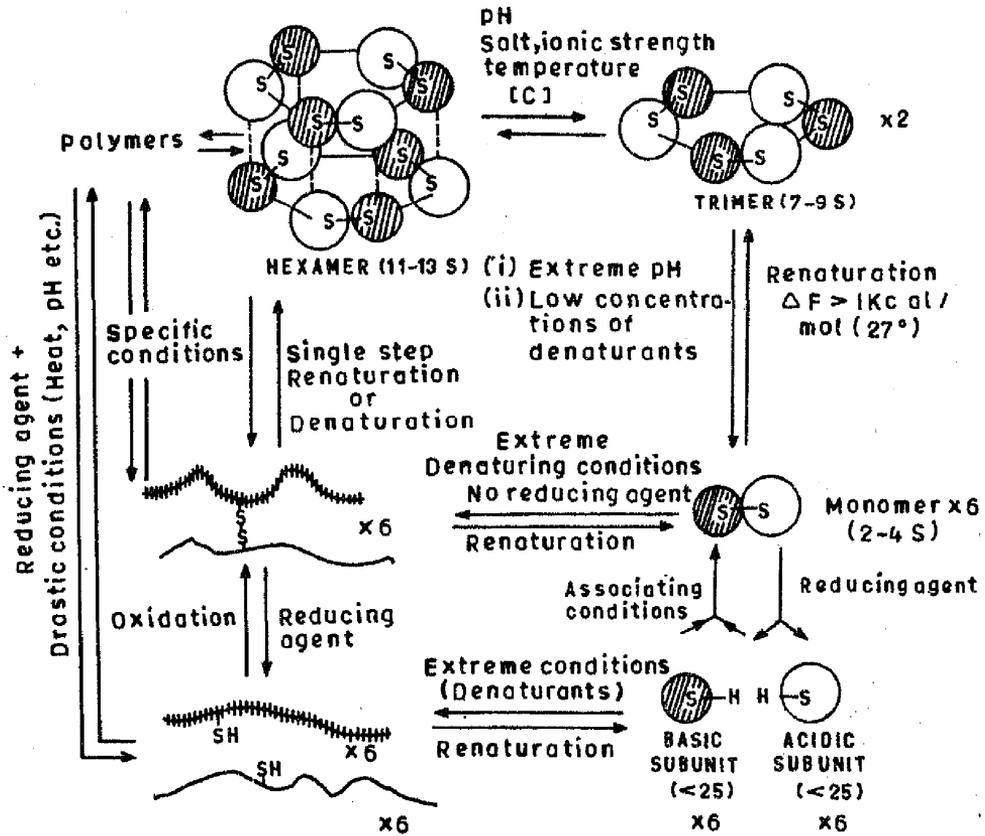
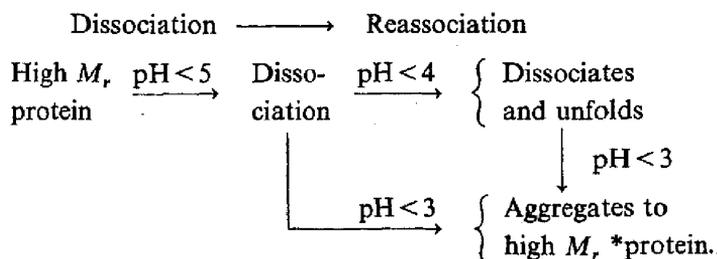


Figure 1. Proposed model for the association-dissociation, denaturation and reassociation of the high M_r protein fraction from various oilseeds. Part of the model is similar to the one proposed by Pernollet and Mosse (1983). Dashed lines indicate weak non-covalent interactions, solid lines indicate strong non-covalent interactions. S-S, Disulphide bridges.

Another anomalous but interesting property of these proteins is their unusual behaviour at low pH values. The effect of low pH in the range of 5 to 1 on the oligomeric structure, spectral properties and conformation of α -globulin (Prakash and Nandi, 1977), arachin (Prakash and Narasinga Rao, 1986), brassin (M) (Kishore Kumar Murthy and Narasinga Rao, 1984), poppyverin (Srinivas, 1984) and helianthin (Sripad, 1985) has been reported. In the pH range 5 to 3 the proteins dissociate and below pH 3, they reaggregate. The proteins have also been shown to unfold between pH 5 and 3 and refold below pH 3. These proteins are made up of acidic and basic subunits and it is possible that below pH 3, the acidic and basic subunits reassociate (possibly into a different molecule) because of charge effects facilitated by pH. The reassociation may also be due to entropically driven hydrophobic interaction, as has been shown for α -globulin (Prakash and Nandi, 1977). The dissociation and reassociation in acid pH can be represented schematically as follows:

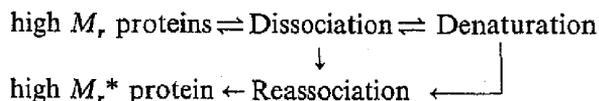


The reassociated molecule may be entirely different from the original molecule protein even though it has been shown to have a sedimentation coefficient of 11S (Prakash and Nandi, 1977). The results suggest that the forces which hold subunits together in α -globulin, arachin, brassin (M), poppyverin and helianthin may be the same. Also, the mechanism of reassociation must be very similar, since in all the cases, reassociation results in a species of nearly the same sedimentation coefficient *i.e.*, the 11S component.

Dissociation -denaturation

The high M_r proteins behave very similarly when they are treated with denaturants such as urea, guanidine hydrochloride (GuHCl) and SDS. Although various intermediates are produced in the course of dissociation, the end product appears to be a denatured 2S molecule (Prakash, 1985). The dissociation and denaturation involve more than a single step as suggested by the intermediates produced at different denaturant concentrations. All of them seem to have the same conformation namely, random coil, since intrinsic viscosity in 6 M GuHCl is close to that of random coil proteins and no ordered structures are present under these conditions. However, α -globulin appears to be more asymmetric than other high M_r proteins in 6 M GuHCl (Prakash, 1985).

These results indicate a certain similarity in the oligomeric structure of the high M_r proteins, and the following pattern:



Quaternary structure

These data on association-dissociation, dissociation-reassociation and dissociation-denaturation support the modified model of the hexamer sandwich, shown in figure 1. The heterohexamer model for the high M_r oilseed proteins can explain the physico chemical properties of the proteins.

Based on this model, some features of association-dissociation and denaturation can be explained. According to the model, the protein molecules are made up of two hexamers. In this structure the stabilizing force is non-covalent interaction, especially entropically-driven hydrophobic interaction similar to that in nucleic acid stacking. Minor variations in pH, ionic strength, nature of ions etc., can disrupt the stabilizing force, driving the reaction in the direction of the 7S trimer or

vice versa, or possibly to a 'polymer' state. Since temperature has a marked effect on non-covalent and in particular hydrophobic interactions, it may be expected that higher temperature (<50°C) favours association into the 11S form. Above 50°C dissociation may occur because of a totally different mechanism. The reaction may not proceed to unfolding of subunits because the basic and acidic subunits are held together by a much stronger covalent bond, *i.e.*, the S-S bond. The formation of the 7S trimer could also be facilitated by low concentration of urea, GuHCL or SDS, which destabilize the hydrophobic interactions between the monomers (acidic + basic subunit). This results in a situation where the non-covalent interactions between the trimers in the hexamer molecule are also destabilized. It could also mean that the nature of non-covalent interaction between the monomers in the trimer and the trimers in the hexamer could be different (figure 1, solid lines and dashed lines). At 8 M urea, 6 M GuHCL or 1×10^{-2} M SDS, the monomer or the 2S form is stabilized since all non-covalent interactions cease to exist at these concentrations of the denaturants. If a reducing agent is present, these monomers could further dissociate into acidic and basic subunits and would exist as completely unfolded polypeptide chains (figure 1). On the other hand if there is no reducing agent then the monomer would denature to a single molecule containing an S-S bond increasing the asymmetry of the uncoiled molecule (figure 1).

On the basis of the hexamer sandwich model, one can explain the unusual effect of acidic pH on the high M_r proteins. It is conceivable that as the pH is lowered from pH 7 to 4 hydrophobic interactions are weakened because of the increase in the concentration of H^+ . Hence the molecule dissociates to a 2S component. From spectral data it is known that at this pH 4 aromatic groups, probably the contact groups between the subunits, are exposed. The increased energy of interaction may lead to the reassociation of these polypeptide chains rather uniquely since the hexamer is stabilised by noncovalent interactions.

Summarising, the proposed model has 4 main features. They are (i) area of association between acidic and basic subunits originates probably from the same precursor polypeptide chain (*i.e.*, the disulphide linkage area) (Pernollet and Mosse, 1983; Prakash and Narasinga Rao, 1986), (ii) the monomers form trimers by noncovalent association where each acidic (or basic) subunit associates with two basic (or acidic) subunits (Pernollet and Mosse, 1983), (iii) the trimer non-covalently associates with another trimer forming the hexamer and (iv) the non-covalent interactions within and between the trimers appear to be different from one another. This probably is the deciding factor in the stabilization of the native structure of the high M_r protein.

Biological significance

It is believed that the main function of the high M_r proteins in oilseeds is to serve as storage proteins. Recently, they have also been considered as secretory proteins (Pernollet and Mosse, 1983). These proteins are hydrolysed to amino acids during germination to create a large amino acid pool for synthesizing new proteins and to serve as a readily available nitrogenous source to facilitate growth of the plant. It is conceivable that the subunits are held together by weak forces ($\Delta F < 1$ Kcal/mol at 27° C) which allow the breakdown of the oligomeric structure without much expenditure of energy.

In terms of the energetics, the globular shape of these proteins would tend to force dense packing of the proteins in cellular structures. Such a structure may be achieved by successive planes of polypeptides joined by glutamyl residues (Pernollet and Mosse, 1983). This may be one reason why the high M_r proteins are rich in glutamic acid.

References

- Bietz, J. A. and Wall, J. S. (1973) *Cereal Chem.*, **50**, 537.
- Bigelow, C. C. (1967) *J. Theor. Biol.*, **16**, 187.
- Blake, C. C. F. and Johnson, L. N. (1984) *Trends Biochem. Sci.*, **9**, 147.
- Fisher, H. F. (1964) *Proc. Natl. Acad. Sci USA*, **51**, 1285.
- Gill, T. A. and Tung, M. A. (1978) *Cereal Chem.*, **55**, 180.
- Gururaj Rao, A. and Narasinga Rao, M. S. (1981) *Int. J. Pept. Protein Res.*, **18**, 154.
- Kishore Kumar Murthy, N. V. and Narasinga Rao, M. S. (1984) *Int. J. Pept. Protein Res.*, **23**, 94.
- Latha, T. S. and Prakash, V. (1986) *Die Nahrung*, **30**, 833.
- Li, A. L., Piyakina, G. A., Yadgarov, E. G., Shadrina, T. Yu, Asatov, S. I., Yunusov, T. S. and Yuldashev, P. Kh. (1979) *Khim. Prir. Soedin.* 5, 680. *Chem. Abstr.* **94**, 27405, 1981.
- Madhusudan, K. T. (1984) *Studies on the linseed protein concentrate and isolate*, Ph.D. Thesis, University of Mysore, Mysore.
- Mosse, J. (1973) *Physiol. Veg.*, **11**, 361.
- Pernollet, J. C. and Mosse, J. (1983) in *Seed proteins* (eds J. Daussant, J. Mosse and J. G. Vaughan) (London: Academic Press) p. 155.
- Prakash, V. (1985) *J. Biosci.*, **9**, 165.
- Prakash, V. and Nandi, P. K. (1977) *Int. J. Pept. Protein Res.*, **9**, 97.
- Prakash, V. and Nandi, P. K. (1978) *J. Agric. Food. Chem.*, **26**, 320.
- Prakash, V. and Narasinga Rao, M. S. (1984) *Proc. Indian Acad. Sci. (Chem. Sci.)*, **93**, 1205.
- Prakash, V. and Narasinga Rao, M. S. (1986) *CRC Crit. Rev. Biochem.*, **20**, 265.
- Provenchar, S. W. and Glockner, J. (1981) *Biochemistry*, **20**, 33.
- Reddy, I. M. and Narasinga Rao, M. S. (1984) *Proc. 53rd Annual meeting Soc. Biol. Chem.*, p. 9.
- Schwenke, K. D., Schultz, M., Linow, H. J. Uhlig, J. and Franzke, C. L. (1974) *Die Nahrung*, **18**, 709.
- Sharon, N. and Lis, H. (1981) *Chem. Eng. News*, **59**, 21.
- Sharon, N. (1984) *Trends Biochem. Sci* **9**, 198.
- Shifrin, S., Luborsky, S. W. and Grochowski, B. J. (1971) *J. Biol. Chem.*, **246**, 7708.
- Shvarts, V. S. and Vaintraub, I. A. (1967) *Biochemistry (USSR)*, **32**, 135.
- Srinivas, H. (1984) *Studies on the proteins of poppy seed (Papaver somniferum L)*, Ph. D. Thesis, University of Bombay, Bombay.
- Sripad, G. (1985) *Effects of methods of polyphenol removal on the proteins of sunflower (Helianthus annuus L) seed*, Ph.D. Thesis, University of Mysore, Mysore.
- Tanford, C. (1961) *Physical Chemistry of macromolecules* (New York: John Wiley)
- Teale, F. W. J. (1960) *Biochem. J.*, **76**, 381.
- Tecson, E. M. E., Bernadotta, E., Lartok, L. P. and Juliano, B. O. (1971) *Cereal Chem.*, **48**, 168.
- Tombs, M. (1965) *Biochem. J.*, **96**, 119.
- Waugh, D. F. (1956) *Adv. Protein chem.*, **9**, 326.