Computational Analysis of the Amino Acid Residue Sequences of Amaranth and Some Other Proteins

Shela Gorinstein,¹† Marina Zemser,¹ Amit Fliess,¹ Inna Shnitman,² Octavio Paredez-Lopez,³ Kazutaka Yamamoto,⁴ Shoichi Kobayashi,⁴ and Hajime Taniguchi⁴

¹Department of Pharmaceutical Chemistry, School of Pharmacy, The Hebrew University of Jerusalem, Jerusalem 91120, Israel
²Computation Center, The Hebrew University of Jerusalem, Jerusalem, Israel
³Departamento de Biotecnología y Bioquímica, Centro de Investigacion y de Estudios Avanzados del IPN, Unidad Irapuato, Apdo. Postal 629, 36500, Irapuato, Gto, Mexico
⁴National Food Research Institute, Ministry of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305, Japan

Received March 6, 1998

Amaranth belongs to a nutritious class of pseudo-cereals. The well balanced amino acid composition of amaranth compared with those of major cereals would indicate that it deserves a quantitative study of its chemical properties. This work was undertaken to compare Amaranthus (A.) caudatus with a number of other plants on the basis of the sequences of various proteins and the composition of their alcohol-soluble protein mixture and glutelins. Alcohol-soluble proteins were extracted with 55% isopropanol (2-ProOH)+5% 2-mercaptoethanol (2-ME) and glutelin fractions were obtained with borate buffer +3% 2-ME+0.5% sodium dodecyl sulfate (SDS), pH 10. Protein fractions were then electrophoresed on sodium dodecyl sulfate polyacrylamide gels (SDS-PAGE). FASTA and TFASTA programmes were used for comparison of amino acid sequences. Dot matrix analysis and secondary structure predictions which were drawn by Plotstructure, were taken from the GCG package. Electrophoretic tests failed to indicate significant correlation between prolamins from cereals and other plants with the alcohol-soluble fractions from amaranth, proving that these proteins cannot represent the major fraction in amaranth. On the other hand, glutelins shared some common electrophoretic bands with other cereals and showed some identity by SDS-PAGE. Amino acid sequences of A. caudatus (100% identity) had degrees of similarity in the range of 71.4 to 52.2% with rice, garden pea, jobs' tears, maize, and yam. Rice glutelin had similarity in the range of 93.3% to 44.8% with oats, soybean, and pea. Secondary structures of A. caudatus (using conservative amino acid replacement), jobs' tears and rice glutelins, oat globulin, and pea legumin sequences were predicted. Some relationship was shown among electrophoretic patterns of alcohol-soluble proteins and glutelins of A. caudatus.

Key words: amino acids; computational analyses; characterization; electrophoresis

A number of wild and cultivated plants and cereals native to Brazil and other Latin American countries are distinguished by their high nutritional content including proteins. These plants and cereals have the potential to produce improved hybrids (cross-breeds) as well as being genetically altered to produce improved pure strains. Proteins, a dietary source of amino acids in food, have been studied extensively for more than three decades.¹-⁴ Alcohol-soluble proteins are the major storage proteins for the cereals such as wheat, barley, rye, maize, and sorghum.²,³,⁶ In some plants prolamins are not the storage proteins. For example, globulins and glutelins are accumulated in oats and rice, respectively, as their major storage proteins. It has been shown that oat globulin and rice glutelin are highly similar and are related to the legume 11S storage proteins.⁹

Amaranth (Amaranthus) belongs to pseudo-cereals and contains more protein and has a better balanced amino acid composition than the major cereals.²,³ There are three main species of amaranth that produce nutritive seeds with potential to become a cereal-like grain crop: Amaranthus (A.) caudatus, A. cruentus, and A. hypochondriacus. A. cruentus has the lowest crude protein content (13.2-18.2%) compared with A. caudatus (17.6-18.4%) and with A. hypochondriacus (17.9%). Their lysine contents are high (3.2 to 6.4%) compared with those found among the most common cereals (2.2 to 4.5%). The sulfur amino acid concentration (2.6 to 5.5%) is higher than that of the most important legumes (1.4%) such as pea, beans, and soybeans. This makes amaranth a promising crop as a food or source of dieta-

¹ To whom correspondence should be addressed: Shela Gorinstein, Department of Pharmaceutical Chemistry, School of Pharmacy, The Hebrew University of Jerusalem, P.O.B. 12065, Jerusalem, 91120, Israel; Phone: 972-2-6758690; Fax: 972-2-6410740; E-mail: gorin@cc.huji.ac.il

Shela Gorinstein is affiliated with the David R. Bloom Center for Pharmacy.

Abbreviations: aa, amino acids; Bis, N',N'-methylene-bis-AA; kDa, kilodalton; 2-ME, 2-mercaptoethanol; MW, molecular weight; PAAG, polyacrylamide gel; 2-ProOH, isopropanol; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis; TRIS, tris (hydroxy-methyl) aminomethane
ry proteins. According to our investigations and in the literature data, amaranth also has albumins and globulins as storage proteins. Our recent publications were focussed on the structure and functional properties of plant globulins. However, since there was no similar investigation about the identity of amino acid sequences in some plants and cereals, this work was undertaken. This paper reports the protein composition of alcohol-soluble and glutenin protein fractions in A. caudatus and other plants by electrophoresis. Proteins of A. caudatus are compared with several other plants and their amino acid sequence, identity, and structure are presented in this study.

Materials and Methods

Materials. Whole mature seeds of Amaranthus (A.) caudatus were donated by Dr. Alrindo Moreira Sales, Instituto De Tecnologia De Alimentos. Whole mature seeds of Coix lachryma-jobi L. (tribe Andropogoneae) plants were donated by Prof. Paulo Arruda, Centro de Biologia Molecular e Engenharia Genetica, Universidade Estadual de Campinas, Campinas, Brazil. High tannin and normal sorghum, maize, and rice were obtained from the Plant Breeding Laboratory, Sementes, Agroceres, Brazil. All seeds were ground in a mill with a 60-mesh screen; and the meal was defatted in a Soxhlet extractor with n-hexane for 10 hours. Some blends were prepared by mixing the meal: A. caudatus with rice (1:1) and (1.5:1); A. caudatus with maize (1:1) and (1.5:1); and maize with rice (1.5:1). The meal was stored at 5°C after removal of hexane.

Methods.

Protein extraction. Proteins were extracted stepwise following the method of Landry and Mourreau (1970). After extraction of albumins and globulins, the prolamins of all seeds were extracted with a solvent (55% 2-ProOH and 5% 2-ME w/v); sample ratio of 6:1 (optimum extraction conditions), which were found in previous studies. Then glutelins were extracted with sodium borate buffer (pH 10), containing 3% 2-ME and 0.5% (w/v) sodium dodecyl sulfate (SDS). The nitrogen content in each fraction was measured by the micro-Kjeldahl method, combined with a colorimetric measurement.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE was done by the method of Laemmli. Protein extracts were combined, lyophilized, and dissolved in a sample buffer that contained 10% glycerol, 5% 2-ME, and 2% SDS in 0.125 M Tris(hydroxy-methyl)aminomethane(TRIS-HCl), pH 6.8. Then the samples were boiled for 5 min before being put on the gel. The gels were 1.5 mm thick and consisted of a 2-cm stacking gel and a 10-cm running gel. The (5–20%) polyacrylamide gel (PAAG) gradients were made from stock solutions of 0% and 30% acrylamide in 0.8% N,N'-methylene-bis-AA (BIS) and 0.1% SDS in 0.375 M TRIS-HCl, pH 8.8. Fifty μg protein was put in each sample slot. Electrophoresis was done at 100 V for 4 hr. Gels were stained for 2 hr with 0.25% Coomassie Brilliant Blue R in methanol/water/acetate acid (5:1:1 v/v) and destained in the same solvent. Molecular mass (Mr) standards (phosphorylase b (94–90 kDa), hemoglobin, canine (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), trypsin inhibitor, soybean (20 kDa), and lactalbumin, bovine milk (14 kDa)) were used to estimate protein subunit molecular masses.

Computer analysis of amino acid sequences of proteins from plants and cereals. The amino acid sequence of proteins stored in data banks (Genbank, IR and SWISS-PROT), was compared with those of other proteins from the same source. Computer aided analyses were done by the FASTA and TFASTA programmes, based on the parameters and algorithm of known sequences. Dot matrix analysis was used for comparison of similar sequences. The computation of the secondary structure is based on the parameters and algorithm. The secondary structure predictions were drawn by the Plot Structure program from the GCG package. These programmes were written in the “C” programming language, originally on a VAX 11/780 with the UNIX operating system. It has since been moved to the VAX/VMS operating system, and to an IBM PC microcomputer and is available at the Computer Center at the Hebrew University.

Results and Discussion

The amount of total protein in A. caudatus was 16.6%. These seed proteins were fractionated as albumins and globulins, alcohol-soluble proteins, and glutelins. Their average values were 61.3, 1.4 and 24.1%, respectively. Albumins, globulins, and glutelins as the major fractions, located in the protein body and the alcohol-soluble proteins, probably, were derived from the perisperm. Extractability of the alcohol-soluble proteins from seeds was studied at 20°C using isopropanol (2-ProOH) mixtures with 2-mercaptoethanol (2-ME), varying extraction time, concentrations of reducing agent and proportion of solvent to solid. The optimum extraction conditions were based on the maximum yield of extracted proteins.

Figure 1 shows alcohol-soluble proteins of seeds extracted with 55% 2-ProOH containing 5% 2-ME (6:1 v/w). The measurement of the electrophoretic bands by densitometric analysis have shown that proteins from A. caudatus contain 80–85% polypeptides in the 10–14 kDa range, the rest being minor fractions. As can be seen from Figure 1, alcohol-soluble membrane proteins of amaranth have nonseparated subunits in the region of 8–14 kDa. These protein subunits differed completely from those of other plants and cereals such as normal and high tannin sorghum, rice, and Coix lachryma-jobi. These prolamins, mostly in the 20–30 kDa range, did not show any electrophoretic relationship with amaranth alcohol-soluble fractions. The data proved that these proteins cannot represent the main protein fraction in A. caudatus. Figure 2 characterizes glutelins, one of the most abundant storage proteins in A. caudatus (lane 2) and in rice (lane 4) in comparison with
maize (lane 3). These glutelin fractions shared electrophoretic bands at 14 kDa with the strongest one in A. caudatus and rice, at 20 kDa with the strongest one for A. caudatus, at 28 kDa for A. caudatus and maize; and at 60 kDa for all three samples.

Rice glutelin showed also some minor bands with high MW, around 90 kDa. Electrophoretic patterns of the glutelin fraction extracted from the meal of a mixture of amaranth and rice such as 1:1 (lane 5) and 1.5:1 (lane 6) were intermediate between those of A. caudatus and rice. The abundant components at 14 kDa (lane 5) and between 14 and 20 kDa, 30 kDa and 65 kDa (lane 6) are shown on Fig. 2. Glutelins were also extracted from a mixture of A. caudatus and maize such as 1:1 (lane 7) and 1.5:1 (lane 9). Still the most abundant bands were the region of 14 kDa (lane 7), 30 and 64 kDa in lane 9. Lane 8 shows a mixture of glutelin fraction extracted from maize and rice (1.5:1). The dominant band at 30 kDa (lane 3) was not the strongest one in the shown mixture. As can be seen from the electrophoretic patterns on Fig. 2 all investigated samples and their mixtures shared some common electrophoretic bands (14, 20, 30, and 65 kDa) and showed identity between all glutelin fractions. In future it will be possible to use A. caudatus in blends with other cereals for food nutrients. The search for parts of the polypeptide chains that form the structural regions was done by comparison of the sequences between various plants to detect identity. Using the FASTA and TFASTA computer programs, similarity was detected between a short region of 29–30 residues of legumes, cereals, and other plants. Similar affinities between legume and other plants have been noted previously. Amino acid sequences of plants, which are related in their nutritional values and amino acid composition, of amaranth were compared with other plants. Amino acid similarity of proteins in 30 amino acid (aa) overlaps, using Amaranthus caudatus (amaranth) as 100%, showed some interesting data (Table 1): Oriza sativa (rice)-71.4; Pisum sativum (garden pea)-66.7; Phaseolus vulgaris (kidney bean)-66.7; Coix lachryma-jobi (jobs' tears)-64.3; Triticum aestivum (wheat)-63.6; Hordeum vulgare (barley)-63.6; Solanum tuberosum (potato)-61.9; Zea mays (maize)-57.1. Other sequences are also shown in Table 1. The

<table>
<thead>
<tr>
<th>Plants</th>
<th>Sequence</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>Similarity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amaranth</td>
<td>V G E C</td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
<tr>
<td>Rice</td>
<td>G K Q N D</td>
<td></td>
<td></td>
<td></td>
<td>71.4</td>
</tr>
<tr>
<td>Common tobacco</td>
<td>G K Q A G G A R C</td>
<td></td>
<td></td>
<td></td>
<td>69.6</td>
</tr>
<tr>
<td>Garden pea</td>
<td>G R Q A G G A T C P N N L C S Q G Y G</td>
<td></td>
<td></td>
<td></td>
<td>66.7</td>
</tr>
<tr>
<td>Jobs' tears</td>
<td>C C S K F G Y G C L T D A Y F Y</td>
<td></td>
<td></td>
<td></td>
<td>64.3</td>
</tr>
<tr>
<td>Wheat</td>
<td>G E Q G S G M E C P N N L C S Q Y G C M G G D G Y G C K</td>
<td></td>
<td></td>
<td></td>
<td>63.6</td>
</tr>
<tr>
<td>Barley</td>
<td>G E Q G S N M E C P N N L C S Q Y G C M G G D G Y G C K</td>
<td></td>
<td></td>
<td></td>
<td>63.6</td>
</tr>
<tr>
<td>Potato</td>
<td>G S Q G G G K A C A S G Q C S C S K F G W G N T D Y C G S</td>
<td></td>
<td></td>
<td></td>
<td>61.9</td>
</tr>
<tr>
<td>Western balsam</td>
<td>G Q Q A G D A L C P G G L C S S Y G W G T A D Y C G D</td>
<td></td>
<td></td>
<td></td>
<td>57.1</td>
</tr>
<tr>
<td>Maize</td>
<td>P A A A Q N C G C Q P N V C S C S K F G Y C G T T D E Y C G D</td>
<td></td>
<td></td>
<td></td>
<td>57.1</td>
</tr>
<tr>
<td>Fission</td>
<td>T D V E T F V A T E G M Y T Q Q F Y V Y C G K N A L T Y V G</td>
<td></td>
<td></td>
<td></td>
<td>57.1</td>
</tr>
<tr>
<td>Tomato</td>
<td>K H Q K E L F V A A E G M Y T Q Q F Y V Y C G K A T L M V G</td>
<td></td>
<td></td>
<td></td>
<td>57.1</td>
</tr>
<tr>
<td>Great nettle</td>
<td>G S Q G G G G T C P A L W C C S I W G W G D S E P Y C G R</td>
<td></td>
<td></td>
<td></td>
<td>54.2</td>
</tr>
<tr>
<td>Para rubber</td>
<td>G R Q A G G K L C P N N L C S Q W G W G C T D E Y C S P</td>
<td></td>
<td></td>
<td></td>
<td>52.4</td>
</tr>
<tr>
<td>Yam</td>
<td>Z N C Q C D T T I Y C S Q H G Y C G N S Y D Y C G P</td>
<td></td>
<td></td>
<td></td>
<td>52.2</td>
</tr>
</tbody>
</table>
The predicted secondary structures of amaranth and jobs' tears protein sequences are shown on Fig. 4. As might be expected, the two plants did not show similarity in the pattern of predicted Chou-Fasman structure. Fig. 4(B) shows the changes in the secondary structure of amaranth, when some of the amino acid residues were replaced on the basis of their properties of amino acids. The highest similarity of amaranth with rice (71.4%) is shown in Table 1. The conservative amino acid replacement was done between these two protein sequences, C-N; M-L; P-R; and K-D. Amaranth:

\[ \text{VGENVRGCRPSGLCCSFGYCGKGRDYCGR} \]

Comparison of these two secondary structures [Fig. 4(A and B)] showed that the conservative amino acid replacement is tolerated and has little effect on secondary structure. This makes possible, in the future, replacement of some residues with lysine, without drastic changes in the structure of a, making the plant more nutritional and digestible. Based on the sequence within the *Oryza sativa* (rice) glutelin subunit, we searched for similarity within other plants.17,18

Assigning rice glutelin type 1 precursor 100.0% identi-

---

**Fig. 4.** Plot Structure by Chou-Fasman Prediction of Proteins. (A) *A. caudatus*; (B) *A. caudatus* with the substitution of amino acid residues; and (C) *Coix lacryma-jobi*.

---

**Fig. 3.** Amino Acid Sequence Alignment for Similarity between Proteins. a: Chib-Maize [*Zea mays* (maize)]; Chit-Dioja [*Dioscorea japonica* (yam)]; Amp-Amaca [*Amaranthus caudatus* (amaranth)]; Chib-Pea [*Pisum sativum* (garden pea)]; b, c, d: Amp-Amaca [*Amaranthus caudatus* (amaranth)]; Agi-Oryza [*Oryza sativa* (rice)]; Chit-Soltu [*Solano num tuberosum* (potato)]; and Iamy-Coila [Coix lacryma-jobi] [jobs' tears].
ty in 29 amino acid overlaps, it was found that rice (oryzalin) showed 93.3%; oats 73.3%; fava bean β 63.3%; field bean β 62.1%; cucurbit β 60.0%; soybean Gly(2) 56.7%; pea (legJ) 56.7%; pea (legK) 56.7%; arabidopsis 56.0%;; rape 56.0%; pea (A) 53.3%; soybean (A3-B4) 53.3%; cotton 53.3%; sunflower 52.2%; and pea (β) 44.8%, where * 28 overlaps; ** 25 overlaps, and *** 23 overlaps.

The sequences of legumin from garden pea showed the highest identity of 44.8% in β-chain, 4.0, 53.3% and 56.7% in others. Soybean sequences showed identity of 53.3%; 50.0 and 56.7% in chains of A3-B4, Gly(3) and Gly(2). Fava bean sequences of β-chains of fava and field beans were of 62.1 and 63.3%. It was also observed that globulin of Arabidopsis thaliana, and rape, soybean Gly(3) and pea J have the same identity of 56.0% and 56.7%. The storage protein of Cucurbita maxima (pumpkin) and globulin of Gossypium hirsutum (upland cotton) had in 29 amino acid overlaps, 60.0% and 53.3% sequence similarity in comparison with rice globulin. Globulins of Helianthus annuus (common sunflower) revealed 52.2% identity to globulins from Avena sativa (oats)-73.3%. Earlier work by others have shown that the rice glutelin basic subunit showed similarity to pea legumin and other 11S legumin-like storage proteins. Conservation of peptide sequences among the globulin-type storage proteins from several legumes and cereals has been reported. Despite their different solubility properties, rice glutelin shares many biochemical and cellular properties with 11S legume storage proteins such as pea legumin. Both types of proteins are synthesized as larger precursors on rough endoplasmic reticulum membranes transported and packaged into protein bodies via the Golgi complex and subjected to post-translational proteolysis resulting in the formation of acidic and basic subunits. Each of the listed sequences covers the well-known cutting points (ASN-GLY) to form the acidic and the basic subunits. These sequences, each having a β-sheet structure, can possibly be recognized specifically by processing enzymes to produce the two subunits.

Comparison and dotplot were done using rice glutelin on horizontal and oats (A); fava bean (B); pea J (C); rape (D); cotton (E); sunflower (F) and pea β (G) on vertical dimensions (Fig. 5). Since the diagonal line of similarity is much more evident (Fig. 5), in this comparison it is used throughout to compare the peptide and nucleotide sequences of the identical regions for investigated storage proteins. Dot matrix analysis corresponds with the data of sequence identity (%) of 73.3; 63.3; 56.7; 56.0; 53.3; 52.2 and 44.8. The highest identity is shown in a diagonal across all points of comparison of storage proteins, and a diagonal line of similarity throughout is discernible. Some matching can be seen in the position of Fig. 5 (F and G), and tandem repeated sequences occurring in both proteins appear as blocks of dots. The predicted secondary structures of rice glutelin, oat globulin, and pea legumin sequences are shown on Fig. 6. As might be expected, two cereals; rice glutelin [Fig. 6(A)] and oat globulin [Fig. 6(B)], which have close identity of 100% and 73.3% show also the greatest similarity in the pattern of predicted Chou-Fasman structure. Plot structure of pea legumin [Fig. 6(C)] with 44.8% identity showed a completely different pattern. Secondary structure prediction corresponds with the regions of high identity within the chains of amino acid residues. It was found through the sequence analyses presented here that rice glutelin, the sequence of which is available has vestigial sequence similarity with legume storage proteins.

In conclusion, some relationship was shown between electrophoretic patterns and morphology of alcohol-soluble, glutelin and chitin-binding proteins of A. caudatus and other plants. This relation can be useful for comparison of pseudo-cereals, legumes, and cereals, but analysis of additional protein properties is needed to clarify it. The range of similarity within the chitin-binding proteins and among this particular family of proteins was calculated to be in the range of 90 to 40%.
But between the families there is no sequence relationship. The biochemical and phytophysiological properties of these investigated proteins depend on the similarity in their amino acid sequences. Maybe some substitutions of amino acids can predict the nutritional requirements of amaranth, and the secondary structures found by computer methods will be useful to predict functional changes in protein systems in response to processing conditions which may be encountered in foods.

Acknowledgments

We acknowledge the Research Development Corporation of Japan (JRDC) for STA Fellowship for Shela Gorinstein and the financial support of the David R. Bloom Center for Pharmacy.

References

19) Gatehouse, J. A., Bown, D., Gilroy J., Levasseur Castleton, J.,


