SEQUENCE DATABASES FOR ASSESSING THE POTENTIAL ALLERGENICITY OF PROTEINS USED IN TRANSGENIC FOODS

STEVEN M. GENDEL

Biotechnology Studies Branch
Food and Drug Administration
National Center for Food Safety and Technology
Summit-Argo, Illinois 60501

I. INTRODUCTION

The development of transgenic food plants has progressed to the point that a significant number of these plants are in agricultural use, and many more will be introduced in the next several years. In 1992, the FDA issued a Statement of Policy to clarify the regulatory status of foods derived from new plant varieties, including transgenic plants (FDA, 1992). Part of this Statement of Policy was a Guidance to Industry outlining scientific considerations for evaluating the safety and nutritional aspects of foods from new plant varieties. One of the major issues discussed in the Guidance was allergenicity.

Food allergies occur in approximately 2–10% of the population (Sampson and Metcalfe, 1991; Chandra *et al.*, 1995). Allergic reactions to foods can range from mild itching and redness to lethal anaphylactic shock; sensitive individuals can experience severe reactions when exposed to extremely small amounts of an allergen. Because the only reliable way to deal with food allergy is to avoid the offending food, it is important that allergic

individuals be aware of the composition of all foods consumed (Sampson, 1992; Hingley, 1993). An allergic individual must avoid both whole foods that cause reactions (for example milk) and composite foods that contain components of the allergenic food (such as casein).

The production of transgenic foods raises two major concerns regarding allergenicity (FDA, 1992). The first is the possibility that an allergenic protein could be transferred to a host where the sensitive consumer does not expect it. Coupled with this is the question of whether the protein will remain allergenic in the new host. If the transferred protein is derived from a commonly allergenic donor, it may be possible to obtain some measure of the allergenicity of the protein in the new host by testing with sera from allergic individuals. Recently, this type of testing was carried out for a transgenic soybean containing a protein derived from Brazil nut (Nordlee *et al.,* 1996). The transferred protein retained immunologic activity when tested with sera from Brazil nut-allergic individuals. Although there is no standard protocol for conducting such testing, this example provides a model for developing procedures for assessing the allergenicity of proteins derived from allergenic foods when allergic sera are available. However, if the transferred protein is derived from an allergenic food for which well-characterized sera are not readily available, this approach to safety assessment may be very difficult to carry out.

The second concern is the possibility that a protein not previously part of the food supply will become an allergen (FDA, 1992; Fuchs and Astwood, 1996). As with proteins derived from allergenic foods for which sera are not available, there are no appropriate immunologic tests for potential allergenicity that can be used in this case.

Food allergens, and allergens in general, are a diverse group of proteins. Food allergen proteins are often described as being between 10 and 70 kDa, highly expressed, possibly glycosolated, and resistant to degradation (Taylor, 1992; Fuchs and Astwood, 1996; Hefle, 1996). However, there are no data to show that any of these properties are either necessary or sufficient to cause either sensitization or an allergic reaction in a previously sensitized individual. Therefore, in the absence of reactive human sera, the assessment of potential allergenicity for transferred proteins requires consideration of a number of properties, including the original source of the protein, stability to digestion, stability to processing and/or cooking (which may not be relevant for all allergens, such as the heat-labile allergens associated with oral allergy syndrome), level of expression in the host, and similarity to known allergens.

Because the amino acid sequences of a number of protein allergens have been determined, it has been suggested that sequence comparison can be used as a tool for assessing potential allergenicity (Astwood and Fuchs,

1996; Fuchs and Astwood, 1996; Metcalfe *et al.,* 1996). For example, a recent publication by the International Food Biotechnology Council suggests criteria for using sequence comparison as one component of an allergenicity assessment for foods derived from transgenic crops (Metcalfe *et al.,* 1996). Although several papers report that such comparisons have, in fact, been used in the safety assessment process for transgenic foods, little information has been published on how these comparisons were performed, or on the allergen data sets used (Fuchs *et al.,* 1995; Astwood and Fuchs, 1996; Fuchs and Astwood, 1996).

Sequence comparisons are frequently used to identify functional motifs or domains within proteins. Specific functional domains, such as protease digestion sites or DNA-binding sites, can be located by comparing a test sequence to a previously defined motif or consensus sequence. Unfortunately, too few allergenic epitopes have been identified to permit recognition of a common motif or consensus sequence (if one exists), particularly for food allergens. Consequently, the use of sequence information for assessing potential allergenicity requires that the sequence of each test protein be compared to the sequences of all known allergens. Because minor sequence variations might have major effects on allergenicity, the value of such comparisons depends on using the most complete set of allergen sequences possible.

Therefore, two databases of allergen sequences (food allergens and nonfood allergens) were constructed using information from three large reference protein sequence databases. Allergen sequences were identified in each of the reference databases and compared to homologous sequences in each reference database to identify equivalent sequences and allelic variants. This information was used to construct nonredundant allergen sequence databases that contain all currently available sequence variants for both food and nonfood allergens. In addition, beause of possible immunologic involvement in celiac disease, also known as gluten-associated enteropathy, a third database of wheat gluten protein sequences was also constructed. These databases are available for use in assessing the potential allergenicity of proteins introduced into transgenic foods.

II. METHODS

All of the sequence analysis programs used were part of Version 8 of the GCG Wisconsin sequence analysis package (Genetics Computer Group, Inc., Madison, WI) running on a Digital Equipment Corp. (Maynard, MA) AXP 2100 computer under the Open VMS 6.1 operating system.

All known food allergens are proteins. Therefore, amino acid sequence comparisons should be used for assessing potential allergenicity. The direct comparison of amino acid sequences avoids three problems that could occur with nucleic acid sequence comparisons. First, because the genetic code is degenerate, different nucleic acid coding sequences can specify proteins with identical amino acid sequences. Second, because all known food allergen proteins originate from eukaryotes, the genomic sequences that code for these proteins contain introns. Although it may be possible to identify and use only the coding regions of these sequences, this can be much more complex than simply using the translated amino acid sequence. Third, although most allergen sequences have been obtained by nucleic acid sequencing of cDNA or genomic clones, some have been obtained by direct amino acid sequencing. Therefore, the only way to access the complete set of allergen sequences is by using amino acid sequences.

The amino acid sequences for all proteins used in this study were obtained from the following reference databases: GenPept, release 94; Protein Identification Resource (PIR), release 48; and SwissProt, release 33. The PIR and SwissProt databases were supplied by GCG; the GenPept database was obtained from the National Center for Biotechnology Information via FTP. The PIR is compiled by the National Biomedical Research Foundation (Washington, D.C.) (George *et al.,* 1996), the SwissProt database by Amos Bairoch in collaboration with the European Molecular Biology Laboratory (Bairoch and Apweiler, 1996). Both contain amino acid sequences obtained by peptide sequencing and by translation of nucleic acid sequences as well as extensive annotation. The GenPept database is produced by translation of protein-coding sequences in the GenBank database, and all GenPept accession numbers match the corresponding GenBank accession (Benson *et al.,* 1996). Not all coding sequences in GenBank are included in GenPept due to insufficient information in the annotation. In addition, GenPept does not include any separate annotation, so GenBank was used for all activities that required access to sequence annotation.

All sequences were identified and accessed by accession numbers. Accession numbers were used rather than sequence names because related sequences that are listed independently in one release of a database may be merged into a single entry in subsequent releases. Although the original entry names may be altered or lost, all accession numbers are retained in the new entry.

Searches of database annotation were carried out using the GCG LOOKUP and STRINGSEARCH functions. Sequence comparisons were carried out using the BESTFIT implementation of the Smith and Waterman algorithm for pairwise alignments and the PILEUP implementation of the

Feng and Doolittle progressive alignment method for alignment of multiple sequences (Smith and Waterman, 1981; Feng and Doolittle, 1987).

Updated versions of the databases described here will be made available on-line at http://www.iit.edu/~sgendel.

III. RESULTS

The use of sequence comparisons for food safety assessment is justified only if the database of allergen sequences is as complete as possible. Keyword searching of the annotation in the reference databases did not adequately identify most food allergen sequences. Table I shows the number of accessions that were found in each database by using keyword searching. Most food allergens, and some nonfood allergens, have been sequenced because they are of nutritional, enzymatic, structural, or evolutionary interest. In many cases, the sequence annotation does not indicate that they are also allergens. For example, the keyword searches failed to find known allergens present in milk and eggs. Therefore, in addition to keyword searches, allergenic proteins were identified from several literature sources (Yunginger, 1991; Taylor, 1992; Matsuda and Nakamura, 1993; King *et al.,* 1994; Bush and Hefle, 1996; Metcalfe *et al.,* 1996).

It is important to note that not all allergenic proteins have been characterized with the same degree of precision. In some cases, such as Ara h1 from peanuts or Gad c1 from codfish, the allergenic proteins in a particular food have been studied in detail. In other cases, such as casein, it is not clear whether all components of a protein family are allergenic. Further, allergenic proteins differ in clinical significance, both in terms of the number of sensitive individuals and the severity of reaction. Because the etiology of food allergy is so poorly understood, all available accessions for proteins that have been identified as food allergens were included in the database (Yunginger, 1991; Taylor, 1992; Matsuda and Nakamura, 1993; King *et al.,* 1994; Bush and Hefle, 1996; Metcalfe *et al.,* 1996). Therefore, any use of

TABLE I

NUMBER OF ALLERGEN ACCESSIONS FOUND IN EACH REFERENCE DATABASE BY KEYWORD SEARCHING

Database	Food allergens	Nonfood allergens
GenBank	28	160
PIR	32	169
SwissProt	14	110

these databases should include consideration of the clinical significance and the degree of characterization for each allergen.

Each reference database was searched to locate all accessions containing sequences for each allergen protein. All the accessions for each protein within each reference database were compared to determine whether any were redundant. Redundant sequences occur within a database for several reasons, including deposit of partial or preliminary sequences and sequencing of both cDNA and genomic clones of the same gene. Only accessions that represent unique sequences within each reference database were used to construct the allergen databases. Further, all sequences for each protein were compared between databases, and the allergen databases were constructed to show which accessions in each database contain identical sequences. In some cases, sequences for known food allergens (such as Pen a1) were not included because these sequences had not yet been deposited in the reference databases, so no appropriate accession number was available. These sequences will be included in future updates as they become available.

The results of these searches and comparisons were used to construct two allergen databases, food allergen sequences (Table II) and nonfood allergen sequences (Table III). A third database of wheat gluten sequences was also constructed (Table IV). The wheat sequences were compiled separately because the relationship between gluten-sensitive enteropathy (celiac disease) and food allergy is not clear (O'Mahony and Ferguson, 1991; Metcalfe, 1992; Hefle, 1996). All three databases are available online (see Methods).

The overall content of the two allergen databases is summarized in Table V. The food allergen database contains 138 unique sequences and the nonfood allergen database contains 218 unique sequences. No single reference database contains more than about 60% of the unique food allergen sequences or 75% of the nonfood allergens. In addition, no combination of two of the reference databases contains all of the sequences in either allergen database. Therefore, a complete search of all allergen sequences requires the use of accessions from all three databases.

All accessions listed on the same line in all three databases have the same amino acid sequence; accessions for the same gene with differing sequences are listed on separate lines. For example, in Table II, the GenPept accession J00922, the SwissProt accession P01014, and the PIR accession A01244 contain identical sequences for chicken ovalbumin and can be used interchangeably. However, SwissProt accession P01012 does not exactly match any other ovalbumin sequence.

In some cases, it was necessary to combine two or more accessions from one reference database to completely match a single accession in another

TABLE II

FOOD ALLERGEN SEQUENCES

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes[b]
Animals							
Cod							
	Parvalbumin	Gad c1		P02622	A94236	1,3,4,5,6	
Egg (chicken)							
	Ovomucoid	Gal d1		P01005	A92754	2,3,4,5,6,10	coding?
	Ovalbumin	Gal d2	J00922	P01014	A01244	1,2,3,4,5,6	
	Ovalbumin	Gal d2	V00438		A90455	2,3,4,5,6,10	
	Ovalbumin	Gal d2	V00383			2,3,4,5,6,10	coding?
	Ovalbumin	Gal d2		P01012		2,3,4,5,6,10	
	Ovalbumin	Gal d2	V00385+ V00386+ V00387	P01013	A01243	2,3,4,5,6,10	
	Ovalbumin	Gal d2	V00382			2,3,4,5,6,10	coding?
	Ovotransferrin	Gal d3	Y00407		A03262	1,2,3,4,6	coding?
	Ovotransferrin	Gal d3	X02009			2,3,4,6,10	coding?
	Ovotransferrin	Gal d3		P02789		2,3,4,6,10	
	Lysozyme	Gal d4	J00885	P00698	A00853	1,3,4,6	
	Lysozyme	Gal d4	M10640			3,4,6,10	
	Lysozyme	Gal d4	X61002	P27042	S18463	1,3,4,6	
	Vitellogenin		K02113+	P02845	A92941	1,6	
	Vitellogenin		X00204			6,10	
	Vitellogenin		M18060			6,10	
	Apovitellenin		J00810	P02659	A91484	1,3,6	coding?
Milk (cow)							
	BSA		M73993			1,3,6	
	BSA			P02769	A38885	3,6,10	coding?

(*continues*)

TABLE II (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes[b]
	β-Lactoglobulin		X14712			1,2,3,5,6	
	β-Lactoglobulin		Z48305	P02754	A03218	2,3,5,6,10	
	β-Lactoglobulin		K01086			2,3,5,6,10	
	β-Lactoglobulin		M19088			2,3,5,6,10	
	α-Lactalbumin		J05147	P00711	A34188	1,3,6	coding?
	α-Lactalbumin		X06366			3,6,10	
	α-S1 Casein		M33123	P02662	S22575	1,2,3,5,6	coding?
	α-S1 Casein		M38641			1,2,3,5,6	
	α-S1 Casein		M38658			2,3,5,6,10	
	α-S1 Casein		K01084			2,3,5,6,10	
	α-S2 Casein		M16644	P02663	A29087	1,2,3,5,6	coding?
	β Casein		M15132			1,2,3,5,6	coding?
	β Casein		M55158			2,3,5,6,10	
	β Casein		M16645	P02666	A03110	2,3,5,6,10	coding?
	κ Casein		M36641			1,2,3,5,6	
	κ Casein				S23202	2,3,5,6,10	
	κ Casein			P02668	A03112	2,3,5,6,10	coding?
	κ Casein		K01085			2,3,5,6,10	
Shrimp							
	Tropomyosin	Met e1	U08008			1,3,6	
Plants							
Apple							
	Profilin	Mal d 1	X83672	P43211	S51119+	1,6	
	Profilin	Mal d 1	Z48969		S57625	1,6	
Barley							
	α-Amylase/trypsin inhib.	Hor v 1	X63517		S26197	1,5,6	
	α-Amylase/trypsin inhib.	Hor v 1		P16968		5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1	X69937	P28041			coding?
	α-Amylase/trypsin inhib.	Hor v 1			A24536	5,6,10	

	α-Amylase/trypsin inhib.	Hor v 1	X69938	P32936	B24536+	5,6,10	coding?
	α-Amylase/trypsin inhib.	Hor v 1		P34951		5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1			C24536	5,610	
	α-Amylase/trypsin inhib.	Hor v 1	X69939	P11643	JA0071+	5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1		P01086	A01325+	5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1		P13691	S00332	5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1	X13443	P16969	S01655	5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1	M15207		A25859+	5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1	X59264		S15573	5,6,10	
	α-Amylase/trypsin inhib.	Hor v 1			JQ0342	5,6,10	coding?
Brazil nut							
	2S Albumin	Ber e 1	X54490	P04403	S06252	1,3,6	coding?
	2S Albumin	Ber e 1	X54491			3,6,10	
	2S Albumin	Ber e 1			A25802	3,6,10	
	2S Albumin	Ber e 1			B25802	3,6,10	
Celery							
		Api g 1	Z48967			11	
Kidney bean							
	PR Protein		X61365	P25985+	S11929	1	
	PR Protein		X61364	P25986	S11930	1	
Mustard (leaf)							
	2S Albumin	Bra j 1		P80215	S35591+	1,4,6	coding?
		L			S35592		
Mustard (white)							
	Amylase inhibitor	Sin a 1	S54101			1,2,4,6	coding?
	Amylase inhibitor	Sin a 1		P15322	S01791+	2,4,6,10	coding?
					S01792		
	Amylase inhibitor	Sin a 1			PC1247	2,4,6,10	
	Amylase inhibitor	Sin a 1			PC1246	2,4,6,10	
	Amylase inhibitor	Sin a 1	X91798			11	
	Amylase inhibitor	Sin a 1	X91799			11	

(*continues*)

TABLE II (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes[b]
	Amylase inhibitor	Sin a 1	X91800			11	
	Amylase inhibitor	Sin a 1	X91801			11	
	Amylase inhibitor	Sin a 1	X91802			11	
Papaya							
	Papain		M15203	P00784	A26466	1,3,6	coding?
Peanut							
	Vicilin	Ara h 1	L34402	P43238		1,3,5,6	
	Vicilin	Ara h 1	L38853	P43237		1,3,5,6	
	Agglutinin				A03364	10	
	Agglutinin		S42352	P02872+	S24044	1	coding?
	Agglutinin		U22472			10	
	Agglutinin		U22473			10	
	Arachin			P04149	A03350	2,3,5,6	
	Arachin			P20780	JK0226	2,3,5,6	
Rice							
	α-Amylase/trypsin inhib.	RA 1	D11433	Q01884	S31081	1,2,3,5,6	
	α-Amylase/trypsin inhib.	RA 2	D11434	Q01885	S31082	1,2,3,5,6	
	α-Amylase/trypsin inhib.	RA 5	D11430	Q01881	S31078	1,2,3,5,6	
	α-Amylase/trypsin inhib.	RA 5 b			S59925	11	
	α-Amylase/trypsin inhib.	RA 14	D11432	Q01882	S31080	1,2,3,5,6	
	α-Amylase/trypsin inhib.	RA 14b			S59922	11	
	α-Amylase/trypsin inhib.	RA 14c			S59923	11	
	α-Amylase/trypsin inhib.	RA 16			S59924	11	
	α-Amylase/trypsin inhib.	RA 17	D11431	Q01883	S31079	1,2,3,5,6	
Soybean							
	Glycinin A1aBx		M36686	P04776	A23497	1,3,5,7	coding?
	Glycinin A1aBx		X02985				
	Glycinin A2B1 a		Y00398	P04405	S04604	1,3,5,7	coding?
	Glycinin A2B1 a		X02806			3,5,7,10	

Glycinin A3B4	M10962	P04347	A22615	1,3,5,7	coding?
Glycinin A3B4	M35671			3,5,7,10	
Glycinin A3B4			PQ0200	3,5,7,10	
Glycinin A3B4			PQ0808	3,5,7,20	
Glycinin A3B4	X79467		PQ0809	3,5,7,10	soja Gy5
Glycinin A5A4B3	X02626	P02858	A25207	1,3,5,7	coding?
Glycinin Gy3	X15123	P11828	S04605	1,3,5,7	
Glycinin Gy4	X52863		S20946		
Glycinin A1aB1b	X53404A+		PS0009	3,5,7,10	
Glycinin A7			JA0152	3,5,7,10	
Glycinin			PQ0199	3,5,7,10	
Glycinin A5A4B3	X86970		S54802	3,5,7,10	soja
Glycinin A5A4B3			A91145	3,5,7,10	
Glycinin			S10851	3,5,7,10	
Glycinin			S11003	3,5,7,10	
Glycinin			S11004	3,5,7,10	
β-Conglycinin α-subunit	X17698	P13916	S14681	1,3,6	
β-Conglycinin β-subunit	S44893	P25974	JQ0969	1,3,6	
β-Conglycinin α′-subunit	M13759	P11827	B24810	3,6,10	coding?
β-Conglycinin α-subunit	M26128		S16334	3,6,10	coding?
β-Conglycinin α′-subunit			S16335	3,6,10	
β-Conglycinin β-subunit			S16336	3,6,10	
β-Conglycinin α-subunit			S20007	3,6,10	

(*continues*)

TABLE II (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes[b]
	Lectin		K00821	P05046	S27365	1	
	Trypsin inhibitor		X80039		S49196	1,2,3,5,6	
	Trypsin inhibitor		X64447		S19189	1,2,3,5,6	
	Trypsin inhibitor		X64448		S19190	1,2,3,5,6	
	Trypsin inhibitor				A91998	2,3,5,6,10	
	Trypsin inhibitor			P01071	A01310	2,3,5,6,10	
	Trypsin inhibitor		S45035A		JQ1091	2,3,5,6,10	
	Trypsin inhibitor		S45035B		JQ1092	2,3,5,6,10	
	Trypsin inhibitor		S45092			2,3,5,6,10	
	Oil-body associated	Gly m 1	J05560			5,6	
	Oil-body associated	Gly m 1		P22895	A37126	5,6	

[a] References: 1. Metcalfe *et al.* (1996) (used for accessions listed in Table 8.1 of the reference); 2. Yunginger (1991); 3. Taylor (1992); 4. King *et al.* (1994); 5. Matsuda and Nakamura (1993); 6. Bush and Hefle (1996); 10. Metcalfe *et al.* (1996) (used for accessions homologous to accessions of genes listed in Table 8.1 of the reference); 11. Used for accessions located by keyword searching of the reference databases for which there is no other published reference at this time.

[b] Notes: Coding? = Used for those genes in which one or more database entries indicate that the amino acid sequence reported by originator does not match the sequence obtained by translation of the corresponding nucleic acid, or in cases where annotation indicates that genetic variants exist, but are not reported in the sequence. See text for details.

TABLE III

NONFOOD ALLERGEN SEQUENCES

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes
Alder							
		Aln g 1	S50892	P38948	A53288+	1,2,3	
Alternaria alternata							
	Aldehyde dehydrogenase	Alt a 2	X78227	P42041	S43108	1,2,3	
	Ribosomal protein	Alt a 6	X78222	P42037	S43109	1,3	
		Alt a 7	X78225	P42058	S43111	1,3	
	Ribosomal protein	Alt a 2	X84216	P49148		1	
Ant (jumper)							
		Myr p 1	X70256	Q07932		1,3	
		Myr p 1			S28180	1	
Aspergillus							
	Mitogillin	Asp f 1	X56176	P04389	S16479	1,2,3	coding?
	Mitogillin	Asp f 1	M83781		A46497	1,2,3	
Barley							
		Hor v 9	U06640			1,3	
Bee (honey bee)							
	Phospholipase	Api m 1	X16709	P00630	S05650	1,2,3	coding?
	Hyaluronidase	Api m 2	L10710	Q08169	A47477	1,2	coding?
	Melittin	Api m 4	X02007	P01501	A01761	1,2,3	nomencl?
Bent grass							
		Agr a 1			E37396	3	
Bermuda grass							
		Cyn d 1			A61226	1,2,3	
Birch							
		Bet v 1			A45786	1,2	

(*continues*)

TABLE III (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes
		Bet v 1a	X15877	P15494	S05376	1,2,3	
		Bet v 1b	X77200	P45431	A55699	1,2	
		Bet v 1c	X77265	P43176	B55699	1,2	
		Bet v 1d	X77266	P43177	C55699	1,2	
		Bet v 1e	X77267	P43178	D55699	1,2	
		Bet v 1f	X77268	P43179	E55699	1,2	
		Bet v 1g	X77269	P43180	F55699	1,2	
		Bet v 1j	X77271	P43183	G55699	1,2	
		Bet v 1k	X77272	P43184	H55699	1,2	
		Bet v 1L	X77273	P43185	I55699	1,2,3	
		Bet v 1m	X81972	P43186	A57427	1,2	
	Profilin	Bet v 2			B45786	1,2,3	
	Profilin	Bet v 2	M65179	P25816	JC2082	1,2	
		Bet v 3	X79267	P43187	S45011	1,3	
Blue grass							
		Poa p 1			F37396	1,2,3	
		Poa p 1			A60372	1,2,3	
		Poa p 9	M38342	P22284	C39098	1,2,3	
		Poa p 9	M38343	P22285	A39098	1,2,3	
		Poa p 9	M38344	P22286		1,2,3	
		Pao p 9			B39098	1,2	
		Poa p 9			A60373	1,2	
Candida							
	Alcohol dehydrogenase	Cand a	X81694	P43067		1,2	coding?
	Alcohol dehydrogenase	Cand a	U15924			1,2	
	Alcohol dehydrogenase	Cand a			A61504	1,2	

Cat							
		Fel d 1	M77341	P30440	C56413	1,2,3	coding?
		Fel d 1	X62478		JC1127	1,2	
		Fel d 1	M74952			1,2,3	
		Fel d 1		P30438		1,2	coding?
		Fel d 1			JC1136	1,2	
		Fel d 1	M74953			1,2,3	
		Fel d 1		P30439		1,2	coding?
		Fel d 1			JC1126	1,2	
		Fel d 1			A56413	1,2	
		Fel d 1			B56413	1,2	
Cladosporium							
	Enolase	Cla h 2	X78226	P42040		1,3	
	Alcohol dehydrogenase	Cla h 3	X78228			1,3	
	Alcohol dehydrogenase	Cla h 3			S43114	1	
	Alcohol dehydrogenase	Cla h 3		P40108		1	
	Ribosomal	Cla h 4	X78223	P42039		1,3	
	Ribosomal	Cla h 4	X77253	P42038	S41866	1	
	HSP	Cla h ?	X81860	P40918	S49303	1	nomencl?
		Cla h 5	X78224	P42059	S43116	1,3	
Cockroach							
	Protease	Bla g 2	U28863		A57164	1,2	
		Bla g 4	U40767			1	
Cow (dander)							
			L39834			1	
	Lipocalin		L42867			1	
European hornet							
		Ves c 5.01		P35781	G44522	1,2,3	
		Ves c 5.02		P35782	H44522	1,2,3	
European chestnut							
		Cas s 1			PC2001	1,3	
Filarial worm							
			U03103			3	

(*continues*)

TABLE III (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes
Fire ant (*S. invicta*) red							
	Phospholipase	Sol i 2		P35775	A37330	1,2,3	
		Sol i 3		P35778	B37330	1,2,3	
		Sol i 4		P35777	C37330	1,2,3	coding?
Fire ant (*S. richteri*) black							
	Phospholipase	Sol r 2		P35776	E60727	1,3	
		Sol r 3		P35779	D60727	1,3	
Hazel							
		Cor a 1-5	X70999	P43216	S30053	1,2,3	
		Cor a 1-6	X71000		S30054	1,3	
		Cor a 1-11			S30055	1,3	
		Cor a 1-11	X70997			1	
		Cor a 1-16	X70998		S30056	1,3	
Hornbeam tree							
		Car b 1			C53288	1,2,3	
		Car b 1	X66932	P38949		1,2,3	coding?
		Car b 1	X66918			1,2	
		Car b 1	X66933	P38950		1,2	
Hornet (*D. arenaria*)							
		Dol a 5	M98859	Q05108		1,2,3	
Hornet (*D. maculata*)							
	Phospholipase	Dol m 1	X66869	Q06478	S32406	1,2	
	Phospholipase	Dol m 1			A44563	1,2,3	
	Hyaluronidase	Dol m 2	L34548	P49371	A56090	1,2,3	
		Dol m 5	J03601	P10736	A31085	1,2,3	
		Dol m 5	J03602	P10737		1,2,3	
		Dol m 5			B31085	1,2	

Lilac							
		Syr v 1			S43242	1	
		Syr v 1			S43243	1	
		Syr v 1			S43244	1	
Maize							
		Zea m 1	L14271	Q07154	JC1524	1,3	coding?
			S44171	P33050+	JQ1107+	3	homolog
Meadow velvet							
		Hol L 1	Z27084	P43216	S38581	1,3	
		Hol L 1			S38291	1	
		Hol L 1	Z68893			1	
Mite (*Blomia*)							
			U27479			1	
			U27702			1	
Mite (*D. farinae*)							
		Der f 1		P16311	A61500	1,2	coding?
		Der f 1	X65196			2,3	
		Der f 2	D10447		A61241	1,2,3	
		Der f 2	D10448	Q00855		1,2,3	
		Der f 2	D10449		B61241	1,2,3	
		Der f 2			A61501	1,2	
	Trypsin	Der f 3		P49275		1,3	
	Chymotrypsin	Der f 6		P49276		1	
		Der f mag	D13961	P36973		1	
		Der f mag29	D17676	P39674	JX0313	1	
Mite (*D. pteronyssinus*)							
		Der p 1	U11695	P08176		1,2,3	
		Der p 1	M24794+		JQ0337	1,2,3	
		Der p 1			A31657	1,2	
		Der p 1			S03380	1,2	
		Der p 2		P49278	A60381	1,2,3	

(*continues*)

TABLE III (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes
	Trypsin	Der p 3	U11719	P39675		1,2,3	
	Trypsin	Der p 3			A39997	1,2	
	Amylase	Der p 4		P49274	A61242	1,2,3	
		Der p 5		P14004	S06734	1,2,3	
		Der p 5	X17699			1,2,3	
		Der p 5	S76340			1,2	
	Chymotrypsin	Der p 6		P49277		1,2	
		Der p 7	U37044	P49273		1,2	
		Der p 15	S75286	P46419	S50146	1	
Mite (*D. microceras*)							
		Der m 1		P16312	B27634	1,2,3	
Mite (*Euroglyphus*)							
	Proteinase	Eur m 1		P25780	S21864	1,3	X60073(GB)
Mite (*Lepidoglyphus*)							
		Lep d 1	X83876	P80384+	S56034+	1,2,3	coding?
		Lep d 1	X83875			1,2	
Mugwort							
		Art v 2			A38642	1,2,3	
Oak							
		Que a 1			D53288	1,2,3	
Olive tree							
		Ole e 1		P19963	S36872	1,2,3	coding?
		Ole e 1			A36153	1,2,3	
		Ole e 1			A38968	1,2,3	
		Ole e 1			A53806	1,2,3	
		Ole e 1			B53806	1,2,3	
		Ole e 1			C53806	1,2,3	
		Ole e 1			D53806	1,2,3	

	Ole e 1			E53806	1,2,3	
	Ole e 1			F53806	1,2,3	
	Ole e 1			G53806	1,2,3	
	Ole e 1			H53806	1,2,3	
	Ole e 1			I53806	1,2,3	
Orchard grass						
	Dac g 2	S45354			2,3	
	Dac g 3			A60359	1,2,3	
Pareiteria (*P. judaica*)						
	Par j 1	X77414			1,3	
	Par j 1		P43217	S43682	1	
	Par j 1	X85012		S52933	1	
Parietaria (*P. officinalis*)						
	Par o 1			A53252	1,3	
Pea						
		X85187		S53082	1	
Penicillium notatum						
		S77837			1	
Ragweed (*A. artemisiifloria*)						
	Amb a 1.1	M80558	P27759	A39099	1,2,3	coding?
	Amb a 1.2	M80559	P27760	B39099	1,2,3	coding?
	Amb a 1.2			B53240	1,2	
	Amb a 1.3	M62961	P27761	C39099	1,2,3	coding?
	Amb a 1.3	M80560			1,2	
	Amb a 1.3			C53240	1,2	
	Amb a 1.4	M80562	P28744	D53240	1,2,3	coding?
	Amb a 2	M80561	P27762	A46469	1,2	coding?
	Amb a 2			E53240	1,2,3	
	Amb a 3		P00304	A00313	1,2,3	
	Amb a 5		P02878	A03371	1,2,3	coding?
Ragweed (*A. psilostachya*)						
	Amb p 5	L24465	P43174		1,3	coding?

(*continues*)

TABLE III (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes
		Amb p 5	L24466			1,3	
		Amb p 5	L24467	P43175		1,3	coding?
		Amb p 5	L24468			1,3	
		Amb p 5	L24469			1,3	
Ragweed (*A. trifida*)							
		Amb t 5	S39336	P10414	JQ1001	1,2,3	
Reed fescue							
		Fes e 1a			C37396	1,3	
		Fes e 1b			D37396	1,3	
Roundworm (*Ascaris*)							
		Asc l 1	L03211		A48576	1,2,3	species
		Asc s 1			A49139	1,2	confused
Roundworm (*Toxicara*)							
					B49139	1	
Rye							
		Sec c			S38292	1,3	
Ryegrass							
		Lol p 1	M57474		B37881	1,2,3	
		Lol p 1		P14964		1,2	
		Lol p 1	M57476+		S13614	1,2	
		Lol p 1			A23341	1,2	
		Lol p 1b	M59163			1,2,3	
		Lol p 1b			A38582	1,2	
		Lol p 2	X73363			1,2	
		Lol p 2a		P14947	A34291	1,2,3	
		Lol p 2b			A48595	1,2,3	
		Lol p 3		P14948	A33422	1,2,3	
		Lol p 4			A60737	1,3	

		Lol p 5			S38288	1,2,3	
		Lol p 5			S38289	1,2,3	
		Lol p 5			S38290	1,2,3	
		Lol p 9	L13083		JT0756	1,2,3	
		Lol p 11			A54002	1	
Soybean							
		Gly m cim1	U03860		S48032	1,3	
Sugi (*Japanese cedar*)							
		Cry j 1 a	D26544	P18632	JC2123	1,2,3	
		Cry j 1 b	D26545		JC2124	1,2,3	
		Cry j 2	D29772		JC2498	1,2,3	
		Cry j 2	D37765	P43212	S48730	1,2	coding?
Sweet vernal grass							
		Ant o 1			G37396	1,3	
Timothy grass							
		Phl p 1	X78813	P43213	S44182	1,2,3	
		Phl p 1	Z27090		S38620	1,2,3	
		Phl p 2	X75925	P43214	S39457	1,3	
		Phl p 5	Z27083		S38584	1,2,3	
		Phl p 5			A61505	1,2	
		Phl p 5			S37400	1,2	
		Phl p 5a			S32101	1,2	
		Phl p 6	Z27082	P43215	S38585	1,3	
	Profilin	Phl p 11	X77583	P35079	S42023	1,3	
		Phl p 32K			S38294	1,3	
		Phl p 38K			S38293	1,3	
Tomato							
			X15855	P13447	S04765	3	analog?
Wasp (*P. annularis*)							
		Pol a 5	M98857	Q05109		1,2,3	

(*continues*)

TABLE III (*Continued*)

Species	Protein	Allergen name	GP accession	SP accession	PIR accession	References[a]	Notes
Wasp (*P. exclamans*)		Pole e 5		P35759	A37329	1,2,3	
Wasp (*P. fascatus*)		Pol f 5		P35780	F44583	1,2,3	
Wheat		Tre a 3	Z50867			1	
Yellow jacket (*V. flavopilosa*)		Ves f 5		P35783	B44522	1,2,3	
Yellow jacket (*V. germanica*)		Ves g 5		P35784	A44522	1,2,3	
Yellow jacket (*v. maculifrons*)	Phospholipase	Ves m 1			A44564	1,2,3	
		Ves m 5		P35760	B37329	1,2	
Yellow jacket (*V. pensylvanica*)		Ves p 5		P35785	C44522	1,2,3	
Yellow jacket (*V. squamosa*)		Ves s 5		P35786	D44522	1,2,3	
Yellow jacket (*V. vidua*)		Ves vi 5		P35787	E44522	1,2,3	
Yellow jacket (*V. vulgaris*)	Phospholipase	Ves v 1	L43561	P49369		1,2	
	Hyaluronidase	Ves v 2	L43562	P49370		1,2	
		Ves v 5	M98858	Q05110		1,2,3	

[a] References: 1. Used for accessions identified by keyword searching of the reference databases; 2. King *et al.* (1994); 3. Metcalfe *et al.* (1996) (used for accessions listed in Table 8.2 of the reference).

TABLE IV

WHEAT GLUTEN SEQUENCES

Species	GP accession	SP accession	PIR accession	References[a]	Notes
Gliadin	U08287			1,2,3,6	
Gliadin	K02068			1,2,3,6	
Gliadin	K02069	P04728		2,3,6,10	
Gliadin	X02538	P04726	S07361	1,2,3,6	
Gliadin	X02539			1,2,3,6	
Gliadin	X01130			2,3,6,10	
Gliadin	X02540			1,2,3,6	
Gliadin	K03075	P04727		1,2,3,6	
Gliadin	K03076			1,2,3,6	
Gliadin	M11074	P04721		1,2,3,6	
Gliadin			B22364	2,3,6,10	
Gliadin	M10092	P04722		1,2,3,6	
Gliadin			C22364	2,3,6,10	
Gliadin	M11076	P04723		1,2,3,6	
Gliadin			E22364	2,3,6,10	
Gliadin	M11075	P04724		1,2,3,6	
Gliadin			D22364	2,3,6,10	
Gliadin	M11073	P04725	A22364	1,2,3,6	
Gliadin	X17361	P18573	S10015	1,2,3,6	
Gliadin	X00627	P02863	A03354	1,2,3,6	coding?
Gliadin	M36999	P21292	JA0153	1,2,3,6	
Gliadin	M16064	P08453	JS0402	1,2,3,6	
Gliadin	M13713	P06659	A25632	1,2,3,6	
Gliadin	M11077	P04729		1,2,3,6	
Gliadin	M11336			1,2,3,6	
Gliadin	M11335	P04730	S07398	1,2,3,6	

(*continues*)

TABLE IV (*Continued*)

Species	GP accession	SP accession	PIR accession	References[a]	Notes
Gliadin	M16496		A27319	1,2,3,6	
Gliadin	M16060	P08079	PS0094	2,3,6,10	
Gliadin		P02865	A03356	2,3,6,10	
Gliadin	X04532			2,3,6,10	
Gliadin			S52126	2,3,6,10	
Glutenin	X03041	P08488	A24266	2,3,6	
Glutenin	X13306	P10386	S04325	2,3,6	
Glutenin	X12929	P10387	S04832	2,3,6	
Glutenin	X00054	P02861	A03352	2,3,6	
Glutenin	X00055	P02862	A03353	2,3,6	
Glutenin	X03346	P08489	A24107	2,3,6	
Glutenin	X12928	P10388	S02262	2,3,6	
Glutenin	X07747	P10385	S01992	2,3,6	
Glutenin	X51759	P16315	S08683	2,3,6	
Glutenin	M22209		A30843	2,3,6	
Glutenin	X61009		S15720	2,3,6	
Glutenin	X62588		S20853	2,3,6	
Glutenin	X61026		S18733	2,3,6	
Glutenin			S29176	2,3,6	
Glutenin			S29177	2,3,6	
Glutenin			S29178	2,3,6	
Glutenin			S29179	2,3,6	
Glutenin			S06645	2,3,6	
Glutenin	M22208			2,3,6	
Glutenin			JN0689	2,3,6	
Glutenin			JC2099	2,3,6	
Glutenin	X13928			2,3,6	
Glutenin	X84887			2,3,6	
Glutenin	X84959			2,3,6	

Glutenin	X84960			2,3,6	
Glutenin	X84961			2,3,6	
Agglutinin	M25536	P10968	S09623	1,2	coding?
Agglutinin	M25537	P02876	S09624	1,2	coding?
Agglutinin	J02961	P10969	A28401	1,2	
α-Amylase/trypsin inhib.		P01084	A01323	2,5,10	
α-Amylase/trypsin inhib.		P10846	S05017	2,5,10	
α-Amylase/trypsin inhib.		P01083		2,5,10	coding?
α-Amylase/trypsin inhib.			A01322	2,5,10	
α-Amylase/trypsin inhib.		P01085	A01324	2,5,10	
α-Amylase/trypsin inhib.		P16852	D25310	2,5,10	
α-Amylase/trypsin inhib.	X16733	P16159	S08466	2,5,10	
α-Amylase/trypsin inhib.	X17574	P17314	S10029	2,5,10	
α-Amylase/trypsin inhib.	X55454	P16851	S13376	2,5,10	
α-Amylase/trypsin inhib.	X17575	P16850	S10027	2,5,10	
α-Amylase/trypsin inhib.	X59791		S18241	2,5,10	
α-Amylase/trypsin inhib.			S16920	2,5,10	
α-Amylase/trypsin inhib.			S38955	2,5,10	
α-Amylase/trypsin inhib.			S10849	2,5,10	
α-Amylase/trypsin inhib.			S10850	2,5,10	

[a] References: Same as in Table II.

TABLE V

SUMMARY OF ALLERGEN DATABASES

Food allergens	
Unique sequences	138
GenPept accessions	89
SwissProt accessions	53
PIR accessions	90
Species	15
Proteins	44
Nonfood allergens	
Unique sequences	218
GenPept accessions	118
SwissProt accessions	105
PIR accessions	162
Species	65
Proteins	142

reference database. Figure 1 shows an example in which two PIR accessions must be combined to completely match a single SwissProt accession. In these cases, the two (or more) accessions necessary are both listed in the database, separated by a +. In addition, in some cases an entry in one database was an exact match to part of an accession in another database.

```
        1                                                          50
P15322  PAGPFRIPKC RKEFQQAQHL RACQQWLHKQ AMQSGSGPSP QGPQQRPPLL
S01791  PAGPFRIPKC RKEFQQAQHL RACQQWLHKQ AMQSGSGPS. ..........
S01792  .......... .......... .......... .........P QGPQQRPPLL

        51                                                         100
P15322  QQCCNELHQE EPLCVCPTLK GASKAVKQQV RQQLEQQGQQ GPHVISRIYQ
S01791  .......... .......... .......... .......... ..........
S01792  QQCCNELHQE EPLCVCPTLK GASKAVKQQV RQQLEQQGQQ GPHVISRIYQ

        101                         127
P15322  TATHLPKVCN IPQVSVCPFK KTMPGPS
S01791  .......... .......... .......
S01792  TATHLPKVCN IPQVSVCPFK KTMPGPS
```

FIG. 1. An example of the combination of two accessions from one reference database (PIR S01791 and S01792) to match a single accession from another reference database (SP P15322). These code for the mustard allergen Sin a1, an amylase inhibitor.

For example, SwissProt accession P02872 for peanut agglutinin contains 236 amino acids that exactly match the middle of the 273 amino acids in PIR accession S24044 and GenPept accession S42352. In these cases, the entry for the short accession is followed by a + in the database.

Multiple nonidentical accessions were found for many proteins during construction of these databases. Most of these sequence differences probably reflect the presence of multiple homologous genes within a single genome and multiple alleles within a population. Because the significance of multiple alleles for food allergy is not known, it is important to include all sequences in the databases used for assessing potential allergenicity. For example, at least six accessions are necessary to include all the variant sequences of chicken ovalbumin that were found in the three reference databases (SwissProt P01014, P01012, P01013; PIR A90455; and GenPept V00383, V00382).

The annotation of the accessions used in the allergen databases revealed three other common problems. First, the sequence present in a single accession may be a consensus sequence derived from multiple sources. The conflicts between the original sequences and the consensus are indicated in the annotation, but the alternative sequences are not available for searching. This is the case for the ovalbumin sequence in GenBank accession V00383 (Fig. 2).

Second, as discussed above, sequencing of multiple clones or cDNAs may reveal the presence of sequence variants present in a single genome or population. These variants may be reported in the annotation, but not in the actual sequence data. For example, the annotation for SwissProt accession P02872, for peanut agglutinin, indicates sequence differences for a minor variant but these differences are not available for sequence searching (Fig. 3). The annotation for SwissProt accession P04405, for soybean glycinin, indicates that both conflicts and sequence variants are present.

Third, although the annotation for some entries cross-reference accessions in the other databases, these cross-references do not always point to identical sequences. For example, SwissProt accession P02666, for β-casein, cites GenBank/GenPept accessions M15132 and M55158, neither of which contain exactly matching sequences. However, GenBank/GenPept accession X16645 does match exactly.

Therefore, in constructing the allergen databases, all apparently identical or homologous sequences were directly compared to ensure that all sequence variants present in the three reference databases were identified and duplicate sequences were eliminated. In addition, the annotation for each accession was scanned and the presence of any problem or conflict was indicated in the allergen database.

```
LOCUS     GGALB2     1873 bp  RNA        VRT     17-MAY-1995
DEFINITION  Chicken messenger RNA for ovalbumin.
ACCESSION   V00383
          .
          .
          .
FEATURES          Location/Qualifiers

   conflict       replace(35,"g")
                 /citation=[2]
                 /citation=[3]
   conflict       replace(44,"g")
                 /citation=[2]
                 /citation=[3]
   conflict       replace(80,"t")
                 /citation=[2]
                 /citation=[3]
   conflict       replace(224,"g")
                 /citation=[2]
                 /citation=[3]
   conflict       replace(627,"g")
                 /note="A is G in [2]"
                 /citation=[2]
   conflict       replace(1308,"a")
                 /note="C is A in [2]"
                 /citation=[2]
   conflict       replace(1459. .1474,"gg")
                 /citation=[2]
```

FIG. 2. An example of part of the GenBank annotation for an accession noting conflicts between sequences from different sources (citations). Only the consensus sequence is available in GenPept.

IV. DISCUSSION

Databases of allergen-related amino acid sequences have been constructed by using accessions derived from three large reference databases. Each allergen database was designed to allow identification of a set of unique sequences that includes all accessible alleles and variants of allergenic proteins. These databases will be updated periodically as new allergens sequences become available and as new proteins are identified as allergens and the updated databases will be made available on-line (see Methods).

Sequence matching between the accessions in these databases and proteins used in the production of transgenic foods can be used as part of the safety assessment process for these new food varieties (Metcalfe *et al.*,

```
ID   LECG_ARAHY     STANDARD;      PRT;   236 AA.
AC   P02872;
      .
      .
DE   GALACTOSE-BINDING LECTIN (AGGLUTININ) (PNA).
OS   ARACHIS HYPOGAEA (PEANUT).
OC   EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; FABALES;
OC   FABACEAE.
      .
      .
FT   METAL        137     137        MAGNESIUM (BY SIMILARITY).
FT   VARIANT       92      92        E -> V (IN MINOR FORM).
FT   VARIANT      149     149        K -> A (IN MINOR FORM).
FT   VARIANT      162     162        K -> I (IN MINOR FORM).
FT   VARIANT      212     213        LG -> RA (IN MINOR FORM).
SQ   SEQUENCE   236 AA;  25189 MW;  1B4552A8 CRC32;
```

FIG. 3. An example of part of the SwissProt annotation for an accession noting the presence of variant sequences. The varient sequences are not available for sequence comparisons.

1996). However, it is clear that these databases do not map all of the relevant sequence space. In addition, because little is known about the etiology of food allergies, there are no generally accepted criteria for defining significant matches (Fuchs and Astwood, 1996; Metcalfe *et al.,* 1996). Therefore, sequence matching should be combined with other considerations such as the source of the protein, stability to digestion, and stability to processing in an overall safety assessment, possibly using a scheme similar to that described in Metcalfe *et al.* (1996).

The problems identified during the construction of these databases highlight the importance of a thorough understanding of the structure and content of any molecular data sources used in public health-related safety assessments. For example, the degree of sequence heterogeneity between the reference databases was unexpected.

These databases are currently being used to test methods of sequence matching to determine the optimal procedure for using sequence information in allergenicity assessment. Further, as more epitopes are identified within allergenic proteins, these databases will be useful for determining whether common structural or sequence features exist in food allergen proteins, and (if they do) for deriving consensus sequences or motifs.

ACKNOWLEDGMENTS

I thank Dr. James Maryanski and Dr. Nega Beru for helpful discussion and comments, and Ted Chambers and Charles Baynard for assistance with the computing. This work was partially supported by Cooperative Agreement no. FD000431 between the FDA and the National Center for Food Safety and Technology.

REFERENCES

Astwood, J., and Fuchs, R. 1996. Allergenicity of food derived from transgenic plants. *Monogr. Allergy* **32,** 105–120.

Bairoch, A., and Apweiler, R. 1996. The SWISS-PROT protein sequence data bank and its supplement TREMBL. *Nucleic Acids Res.* **24,** 21–25.

Benson, D., Boguski, M., Lipman, D., and Ostell, J. 1996. GenBank. *Nucleic Acids Res.* **24,** 1–5.

Bush, R., and Hefle, S. 1996. Food allergens. *Crit. Rev. Food Sci. Nutr.* **36**(S), S119–S163.

Chandra, R., Gill, B., and Kumari, S. 1995. Food allergy and atopic disease. *Clin. Rev. Allergy Immunol.* **13,** 293–314.

FDA. 1992. Statement of policy: Foods derived from new plant varieties. *Fed. Reg.* **57,** 22,984–23,005.

Feng, D., and Doolittle, R. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25,** 351–360.

Fuchs, R., and Astwood, J. 1996. Allergenicity assessment of foods derived from genetically modified foods. *Food Technol.* **50**(2), 83–88.

Fuchs, R., Re, D., Rogers, S., Hammond, B., and Padgette, S. 1995. Safety evaluation of glyphosate-tolerant soybeans. *Proc. OECD Workshop Food Safety Eval.,* 47–55.

George, D., Barker, W., Mewes, H., Pfeiffer, F., and Tsugita, A. 1996. The PIR-international protein sequence database. *Nucleic Acids Res.* **24,** 17–20.

Hefle, S. 1996. The chemistry of food allergens. *Food Technol.* **50**(3), 86–92.

Hingley, A. 1993. Food allergies: When eating is risky. *FDA Consumer* **27**(10), 27–31.

King, T., Hoffman, D., Lowenstein, H., March, D., Platts-Mills, T., and Thomas, W. 1994. Allergen nomenclature. *Int. Arch. Allergy Immunol.* **10,** 224–233.

Matsuda, T., and Nakamura, R. 1993. Molecular structure and immunologic properties of food allergens. *Trends Food Sci. Technol.* **4,** 289–293.

Metcalfe, D. 1992. The nature and mechanisms of food allergies and related diseases. *Food Technol.* **46**(5), 136–139.

Metcalfe, D., Astwood, J., Townsend, R., Sampson, H., Taylor, S., and Fuchs, R. 1996. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. *Crit. Rev. Food Sci. Nutr.* **36**(S), S165–S186.

Nordlee, J., Taylor, S., Townsend, J., Thomas, L., and Bush, R. 1996. Identification of a Brazil-nut allergen in transgenic soybeans. *N. Engl. J. Med.* **334,** 688–692.

O'Mahony, S., and Ferguson, A. 1991. Gluten-sensitive enteropathy (celiac disease) *In* "Food Allergy: Adverse Reactions to Foods and Food Additives" (D. Metcalfe, H. Sampson, and R. Simon, eds.), pp. 186–197. Blackwell Publications, Boston.

Sampson, H. 1992. Food hypersensitivity: Manifestations, diagnosis, and natural history. *Food Technol.* **46**(5), 141–144.

Sampson, H., and Metcalfe, D. 1991. Immediate reactions to foods. *In* "Food Allergy: Adverse Reactions to Foods and Food Additives" (D. Metcalfe, H. Sampson, and R. Simon, eds.), pp. 99–112. Blackwell Publications, Boston.

Smith, T., and Waterman, M. 1981. Comparison of biosequences. *Adv. Appl. Math.* **2,** 482–489.

Taylor, S. 1992. Chemistry and detection of food allergens. *Food Technol.* **46**(5), 146–152.

Yunginger, J. 1991. Food antigens. *In* "Food Allergy: Adverse Reactions to Foods and Food Additives" (D. Metcalfe, H. Sampson, and R. Simon, eds.), pp. 36–51. Blackwell Publications, Boston.