Phylogenetic relationships of chitinases

Anastassis Perrakis[1], Keith S.Wilson[1], Ilan Chet[2], Amos B. Oppenheim[3] and Constantin E. Vorgias[1]*

[1]European Molecular Biology Laboratory, c/o DESY Notkestraße 85, 2000 Hamburg 52, GERMANY. Tel: ++49-40-89902118, Fax: ++49-40-89902149.[2]The Otto Warburg Center for Biotechnology, Faculty of Agriculture, Hebrew University, Rehovot, Israel. [3]Department of Molecular Genetics, Hebrew University-Hadassah Medical School, Jerusalem, Israel.

Summary

Recent interest in enzymes capable of degrading chitin has lead to accumulation of sequences for about 56 chitinases and genes coding for chitinases. Sequence alignments allow us to arrange all known chitinases into two distinct classes I and II which correspond to families 19 and 18 of glycosyl hydrolases. Enzymes belonging to class I are rather homogeneous in their size, with an average of about 300 amino acid residues, and have homologous primary structures. This class was found to be restricted to the plant kingdom. Enzymes of class II were found to be present in plant, fungi, gram positive and negative bacteria including streptomyces. Although the proteins of this class vary in size (290-820 residues) they all have a central region containing several highly conserved domains. A multiple sequence alignment and evolutionary analysis of both chitinase classes are presented.

chitinases, classification, alignment, evolution.

*Corresponding author

Introduction

The hydrolysis of chitin to disaccharides and larger oligomeric saccharides usually occurs extracellularly. Depending on the specific properties of their hydrolytic activity, chitinases have been classified as endochitinases and exochitinases. Endochitinases randomly hydrolyze chitin and produce a population of disaccharides (chitobiose), and oligosaccharides which are further degraded by chitobiase. Exochitinases hydrolyze chitin from its non-reducing end and produce disaccharides. Chitinases and/or their encoding genes have been isolated and studied from various prokaryotic and eukaryotic organisms.

Chitinases (E. C. 3. 2. 1. 14) include all the enzymes with the ability to cleave chitin. Chitin is a fibrous polysaccharide of β-1,4-linked N-acetyl-*D*-glucosamine. It serves as the major structural component in many fungi and arthropods. In yeast, chitin maintains the structure of the mother-bud junction, whereas in filamentous fungi it is very often the major component of the cell wall. It is also the main component of the exoskeleton of arthropods. The major role of chitinases in fungi and arthropods is the modification of the chitin which acts as a structural component of the cell wall. Bacteria produce chitinases primarily to utilize chitin as a carbon and energy source.

In plants, the response to microbial attack involves *de novo* synthesis of an array of proteins designed to restrict the growth of the pathogen (elicitor). Until now several such pathogenic related proteins have been identified and characterized. Among them, hydroxyproline-rich glycoproteins, proteinase inhibitors, enzymes for the synthesis of phytoalexins, enzymes contributing to the reinforcement of cell walls and certain hydrolytic enzymes such as chitinases and glucanases are the most studied. For recent reviews and further references see (1, 2). Purified chitinases are able to inhibit fungal growth *in vitro* by causing lysis of hyphal tips in combination with the activity of b-1,3-glucanases (6). These properties led to the proposal that chitinases are mainly, if not exclusively, involved in the defence mechanism primarily against fungal elicitors (3-10). Therefore chitinases become an attractive subject of intensive investigation. Recently it has been shown that enhanced chitinase levels in transgenic plants can indeed reduce the damage caused by pathogens (6).

Methods

The chitinase sequences used were extracted mainly from the Swissprot Data Bank (41 entries). Additional searches identified more entries in PIR (6 entries), EMBL (6 entries) and GenBank (2 entries). One chitinase sequence was extracted from the original paper (30). They have been identified using a three step procedure: (a) all four Data Banks were searched for the text pattern "chitin" or the EC 3. 2. 1. 14 pointer with the GCG program StringSearch; (b) usable sequences were defined as those with at least 66% sequence completeness; (c) the usable nucleotide sequences were translated to protein according to the data originally submitted by the authors. The selected sequences gave a library of chitinase sequences data files.

The use of both the EC 3. 2. 1. 14 pointer and the text pattern "chitin" ensures the identification of all sequences with possible chitinolytic activity. The results of these searches as well as usefull data are summarized in Table I. From the numerous incomplete sequences found in the databases only one was finally used: Chi1_Horvu (nr 24) see Table I, note b.

The Swissprot chitinase entries have been already classified into two families as described in (71). These families are named family 18 and 19 of glycosyl hydrolases (71, 72). Chitinase entries from the other three data banks were classified in the above families according to sequence similarities. Multiple sequence alignments were carried out utilizing the CLUSTALV program (66). Final alignment for each family was carried out in several steps and in some cases manual "fine tuning" was performed using a text editor. In all cases the Dayhoff's Pam 250 weight matrix was used. Pairwise percent divergence matrices were calculated considering the identities in any position that is not a gap in each pair. No correction was applied for multiple substitutions.

Phylogenetic trees were constructed from the percent divergence matrices with the Neighbour Joining method (68) supplied with CLUSTALV. The trees constructed with that method are unrooted. Therefore the root was defined as the midpoint between the two most distant sequences using the RETREE program, included in the phylogeny inference package PHYLIP 3.5c (67). Finally the rooted trees were plotted with the DRAWGRAM program of the same package.

Alignment profiles for both families were constructed with the GCG program ProfileMake and used to search the Swissprot Data Bank.

Table I. Synoptic presentation of chitinases with full primary structure informaion. The names proposed and used throughout the text and alignments, is based on the Swissprot nomeclature. a:The pI was calculated from amino acid composition using the GCG/Isoelectric, b: fragment with more than 66.6% of the total length of the protein, * GenBank entries, not in EMBL, c: not in 1.12.92 release, sequence was copied from publication

Nr	Ref	Species	Proposed Name	Swissprot	PIR	EMBL	Class	aa	pI[a]
1	11	*Alteromonas sp*	Chit_Alter			ASCHIA *	II	820	4.31
2	12	*Bacillus circulans*	ChiA_Bacci	Chi1_Bacci	A38368	BCCHIA3	II	699	6.27
3	13	*Bacillus circulans*	ChiD_Bacci	ChiD_Bacci		BCCHIDA	II	488	7.75
4	14	*Saccharopolyspora erythrea*	Chit_Sacer	Chit_Sacer	JX0076		II	290	4.27
5	15	*Serratia marcesens*	ChiB_Serma	ChiB_Serma	S04856	SMCHIB	II	499	6.32
6	16	*Serratia marcesens*	ChiA_Serma	ChiA_Serma	A25090	SCCHIA	II	561	8.00
7	17-19	*Streptomyces plicatus*	Chit_Strpl	Chit_Strpl	JH0573	SPCHIA	II	610	4.84
8	20-22	*Aphanocladium album*	Chit_Apalb			X64104[c]	II(a)	423	5.98
9	23	*Saccharomyces cerev*	Chi2_Yeast	Chi2_Yeast	B41035	SCCTS11B	II(a)	562	4.25
10	23	*Saccharomyces cerev*	Chi[illegible]ast	Chi1_Yeast	A41035	SCCTS11A	II(a)	552	4.25
11	24-27	*Kluyveromyces lactis*	Ktxa_Klula	Ktxa_Klula	S07915	KLK1P	-	1149	5.98
12	28	*Brugia malayi*	Chit_Bruma	Chit_Bruma		BMCHIT	II	504	4.47
13	29	*Allium sativum*	Chi1_Alsat			ASCHINTIA	Ia	302	5.17
14	29	*Allium sativum*	Chit_Alsat			ASCHITIN	Ia	318	6.76
15	30	*Arabidopsis thaliana*	Chia_Arath	Chia_Arath		ATCHIA	II(a)	302	9.29
16	30	*Arabidopsis thaliana*	Chit_Arath	Chit_Arath		ATCHIB	Ia	322	6.97
17	31	*Brassica napus*	Chit_Brn25			BNCH25A	Ia	322	6.61
18	32	*Brassica napus*	Chit_Brnap			BNCHITIN	Ia	268	7.79
19	33	*Cicer arietinum*	Chit_Carie			CACHIT*	II(a)	293	4.55
20	34	*Cucumber sativus*	Chia_Cucsa	ChiA_Cucsa	A31455	M24365	II(a)	292	4.31
21	35	*Dioscorea japonica*	Chit_Dioja	Chit_Dioja	A40173		Ia	250	4.60
22	36	*Hevea brasiliensis*	Chly_Hevbr	Chly_Hevbr	A33179		II	273	8.20
23	37	*Hordeum vulgare*	Chi1_Horvu	Chi1_Horvu	S04131	HVENDCHT	Ib	178[b]	8.22
24	38	*Hordeum vulgare*	Chi2_Horvu	Chi2_Horvu	A38664	HVCHI	Ib	266	8.55
25	39	*Lycopersicon esculentum*	Chi1_Lyces		S25634	Z15138	Ib	246	8.12
26	39	*Lycopersicon esculentum*	Chi2_Lyces		S25635	Z15139	Ib	247	4.53
27	39	*Lycopersicon esculentum*	Chi3_Lyces		S25636	Z15141	Ib	253	6.25
28	39	*Lycopersicon esculentum*	Chit_Lyces		S25637	Z15140	Ia	322	6.58
29	40	*Nicotiana tabacum*	Chi1_Tobac	Chi1_Tobac	S08627	NTECHITR	Ia	329	7.88
30	41	*Nicotiana tabacum*	Chi2_Tobac	Chi2_Tobac	S13322	NTECH	Ia	324	7.79
31	42	*Nicotiana tabacum*	Chi3_Tobac	Chi3_Tobac	S20982	NTCHNI4G	Ia	334	8.22
32	43	*Nicotiana tabacum*	ChiA_Tobac	ChiA_Tobac	S19733	NTACIDL3	II(a)	291	4.75
33	43	*Nicotiana tabacum*	ChiB_Tobac	ChiB_Tobac	S19734	NTBASICL3*	II(a)	294	9.01
34	44	*Nicotiana tabacum*	Chip_Tobac	Chip_Tobac	A34801	NTPRP	Ib	253	4.75
35	44	*Nicotiana tabacum*	Chiq_Tobac	Chiq_Tobac	B34801	NTPRQ	Ib	253	4.93
36	45	*Nicotiana tabacum*	Chit_Tobac	Chit_Tobac	JQ0993		Ia	328	7.96
37	46	*Oryza sativa*	Chi1_Orysa	Chi1_Orysa	S14948	OSCHIT	Ia	318	5.13
38	47	*Oryza sativa*	Chi2_Orysa	Chi2_Orysa	S15997		Ia	336	6.46
39	44	*Petunia hybrida*	Chit_Pethy	Chit_Pethy	S20741	PHACHITIN	Ib	254	5.83
40	48	*Phaseolus vulgaris*	Chi1_Phavu		JQ0965		Ia	327	7.91
41	49	*Phaseolus vulgaris*	Chi4_Phavu	Chi4_Phavu	S16579	PVCHITIN	Ia	270	4.47
42	50-52	*Phase[illegible]aris*	[illegible] Phavu	Chit_Phavu	A25898	PVCHM	Ia	328	7.94
43	53	*Pisu[illegible]um*	[illegible] Pisat			PSCHITIN	Ia	320	7.53
44	54	*Populus trichocarpa*	[illegible] Poptr	Chib_Poptr		PTGWIN62B	Ia	305	4.10
45	55	*Populus trichocarpa*	Chi8_Poptr	Chi8_Poptr	A33985	PSCHIB	Ia	316	4.24
46	56	*Rhizopus niveus*	Chi1_Rhini	Chi1_Rhini		RNCHI1	II(a)	493	6.48
47	56	*Rhizopus oligosporus*	Chi1_Rhiol	Chi1_Rhiol		ROCHIT1	II(a)	540	7.56
48	56	*Rhizopus oligosporus*	Chi2_Rhiol	Chi2_Rhiol		ROCHIT2	II(a)	542	7.89
49	57	*Solanum tuberosum*	Chi1_Soltu		S06161	STENCHIT	Ia	315	6.93
50	58	*Solanum tuberosum*	Chit_Soltu	Chit_Soltu	S05426	STCHITIN	Ia	328	8.13
51	59	*Solanum tuberosum*	Chi2_Soltu			STMREN	Ib	264	8.49
52	60-61	*Urtica dioica*	Agl_Urtdi	Agl_Urtdi	B23616	[illegible]ECCHIT	Ia	372	7.25
53	62	*Vigna angularis*	ChiA_Phaan	ChiA_Phaan		[illegible]T	II(a)	298	5.26
54	63	*Zea mays*	ChiA_Maize	ChiA_Maize		ZMCHITA	Ia	269	7.83
55	63	*Zea mays*	ChiB_Maize	ChiB_Maize		Z[illegible]HITB	Ia	280	8.47
56	64	*Zea mays*	ChiC_Maize	ChiC_Maize		Z[illegible]HITC	Ia	318	7.74

Table I

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1 Ag_Urtdi		36	36	32	35	37	41	42	43	48	47	46	47	44	44	47	47	47	47	45	48	43	47	41	40	44	43	43	46	46	50	48	47	46
2 Cha_Maize	36		86	61	56	57	44	45	44	47	51	46	46	44	45	47	48	47	48	48	49	46	43	40	44	40	40	40	43	38	50	48	47	46
3 Chb_Maize	36	86		60	58	54	42	46	45	46	49	47	46	45	46	49	50	49	49	49	50	47	45	39	45	41	42	42	46	38	52	46	48	47
4 Cht_Brnas	32	61	60		60	55	43	41	42	47	47	46	43	41	44	47	48	47	46	45	46	43	44	40	46	43	42	41	44	40	50	44	47	46
5 Ch4_Phavu	35	56	58	60		58	40	40	41	47	43	44	44	41	42	44	44	44	43	44	45	43	39	37	41	39	38	40	41	37	48	42	44	43
6 Cht_Diosp	37	57	54	55	58		40	43	45	45	45	44	44	42	43	46	46	46	45	45	47	44	41	36	41	42	42	41	41	42	49	44	44	46
7 Chb_Poptr	41	44	42	43	40	40		50	61	58	61	55	61	53	53	52	63	51	62	61	60	57	59	54	52	54	56	58	57	53	61	57	57	56
8 Chi1_Phavu	42	45	45	41	40	43	60		93	65	65	60	65	68	59	59	69	59	68	67	66	66	64	59	49	56	58	58	61	54	74	68	68	64
9 Cht_Phavu	43	44	45	42	41	45	61	93		67	66	63	68	69	69	71	72	71	70	70	67	69	64	60	50	55	58	60	61	56	77	70	69	66
10 Chi1_Alnst	48	47	46	47	47	45	58	65	67		91	70	73	69	68	69	69	58	69	68	56	69	66	65	53	62	62	65	65	60	75	72	68	70
11 Cht_Alnst	47	51	49	47	43	45	51	65	66	91		69	71	58	69	68	69	58	71	59	67	66	66	65	55	62	63	65	66	61	75	72	67	68
12 Chi2_Orysa	46	46	47	46	44	44	55	60	63	70	69		78	62	63	63	63	62	64	60	59	62	64	66	48	56	57	60	60	58	77	74	64	64
13 Chi1_Orysa	47	46	46	43	44	44	61	65	68	73	71	78		65	65	65	65	64	66	63	62	65	65	66	53	57	59	61	60	57	72	70	66	66
14 Cht_Arath	44	44	45	41	41	42	63	68	69	69	68	62	55		90	74	74	73	73	70	69	71	68	59	52	60	62	66	66	59	74	71	69	69
15 Cht_Brn25	44	45	46	44	42	43	63	69	69	68	69	63	65	90		72	72	71	71	68	68	68	68	59	52	61	62	65	64	61	76	70	68	68
16 Chi1_Tobac	47	47	49	47	44	46	62	69	71	69	68	63	65	74	72		90	90	89	86	86	82	67	62	56	61	62	65	63	60	74	68	71	71
17 Chi2_Tobac	47	48	50	48	44	46	63	69	72	69	69	63	65	74	72	90		97	90	87	87	82	67	62	56	61	62	65	63	60	74	68	72	71
18 Cht_Tobac	47	47	49	47	44	46	61	59	71	58	68	62	64	73	71	90	97		89	86	85	81	66	61	56	61	62	65	63	60	74	67	71	71
19 Cht_Lycos	47	48	49	46	43	45	62	68	70	69	71	64	66	73	71	89	90	89		93	83	79	68	61	55	61	63	64	63	60	76	70	72	73
20 Chi1_Soltu	45	48	49	45	44	45	61	67	70	68	69	60	63	70	68	86	87	86	93		83	76	66	58	52	63	64	66	63	61	75	70	72	73
21 Cht_Soltu	48	49	50	46	45	47	60	66	67	56	67	59	62	69	68	86	87	85	83	83		74	65	57	53	60	60	63	62	59	74	67	70	69
22 Chi3_Tobac	43	46	47	43	43	44	57	66	69	69	66	62	65	71	68	82	82	81	79	76	74		66	59	53	57	59	65	64	56	71	65	71	69
23 Cht_Pissa	47	43	45	44	39	41	59	64	64	66	66	64	65	68	68	67	67	66	68	66	65	66		57	51	61	61	67	67	57	74	66	66	66
24 Chc_Maize	41	40	39	40	37	36	54	59	60	65	65	66	66	59	59	62	62	61	61	58	57	59	57		45	48	49	50	51	47	66	61	56	57
25 Chi6_Poptr	40	44	45	46	41	41	52	49	50	53	55	48	53	52	52	56	56	56	55	52	53	53	51	45		49	49	48	52	50	51	48	51	51
26 Chip_Tobac	44	40	41	43	39	42	54	56	55	62	62	56	57	60	61	61	61	61	61	63	60	57	61	48	49		93	78	82	75	63	59	59	59
27 Chiq_Tobac	43	40	42	42	38	42	55	58	58	62	63	57	59	62	62	62	62	62	63	64	60	59	61	49	49	93		78	81	76	63	60	62	61
28 Cht_Pethy	43	40	42	41	40	41	55	58	60	65	66	60	61	65	65	65	65	65	64	66	63	65	67	50	48	78	78		85	70	66	62	62	64
29 Chi2_Lycos	46	43	45	44	41	41	57	51	51	65	65	60	60	65	64	63	63	63	63	63	62	64	67	51	52	82	81	85		73	69	66	61	63
30 Chi3_Lycos	46	38	38	40	37	42	53	54	56	60	61	55	57	59	61	60	60	60	60	61	59	56	57	47	50	75	76	70	73		62	57	58	58
31 Chi1_Horvu	50	50	52	50	48	49	61	74	77	75	75	77	72	74	76	74	74	74	76	75	74	71	74	65	51	63	63	66	69	62		94	71	75
32 Chi2_Horvu	48	45	46	44	42	44	57	68	70	72	72	74	70	71	70	68	68	67	70	70	67	65	66	61	48	59	60	62	65	57	94		67	68
33 Chi1_Lycos	47	47	48	47	44	44	57	68	69	68	67	64	65	69	68	71	72	71	72	72	70	71	65	55	51	59	62	62	64	58	71	67		90
34 Chi2_Soltu	46	46	47	45	43	45	56	64	66	70	68	64	65	69	68	71	71	71	73	73	69	69	65	57	51	59	61	64	63	58	75	68	90	

(a)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1 Chi1_Yeast		99	32	32	35	32	35	34	36	35	32	32	12	[illegible]	[illegible]	[illegible]	[illegible]	12	15	13	14
2 Chi2_Yeast	99		32	32	34	32	35	34	36	35	32	32	12	[illegible]	[illegible]	[illegible]	[illegible]	12	15	12	13
3 Chi1_Rhi[illegible]	32	32		95	61	35	39	38	38	36	34	34	11	[illegible]	14	14	10	11	15	12	13
4 Chi2_Rhi[illegible]	32	32	95		60	35	39	38	38	36	34	33	12	[illegible]	14	14	11	11	15	12	12
5 Chi1_Rhi[illegible]	35	34	61	60		34	38	37	37	39	35	33	10	[illegible]	[illegible]	[illegible]	[illegible]	13	16	15	14
6 Cha_Arath	32	32	35	35	34		75	52	59	57	55	56	14	[illegible]	15	15	13	17	14	13	12
7 Chv_Hevbr	35	35	39	39	38	75		72	65	65	62	62	15	[illegible]	[illegible]	[illegible]	13	18	19	16	17
8 Cha_Tobac	34	34	38	38	37	52	72		55	55	52	62	13	[illegible]	[illegible]	[illegible]	13	15	17	14	15
9 Cha_Phavu	36	36	38	38	37	59	65	52		72	54	58	16	[illegible]	[illegible]	[illegible]	15	15	19	16	17
10 Cha_Carpa	35	35	36	36	39	57	65	55	72		52	56	13	[illegible]	[illegible]	[illegible]	13	14	18	15	16
11 Cha_Cucsa	32	32	34	34	35	55	52	52	54	52		55	13	[illegible]	[illegible]	[illegible]	15	14	18	16	17
12 Chb_Tobac	32	32	34	33	33	56	62	52	58	56	55		[illegible]	[illegible]	[illegible]	[illegible]	15	17	13	15	15
13 Cha_Serma	12	12	11	12	10	14	15	13	16	13	13	17		51	30	29	27	27	10	19	15
14 Cht_Alter	12	12	11	11	10	13	15	12	16	15	13	16	51		23	24	25	20	10	19	12
15 Cht_Acab	11	11	14	14	12	15	18	15	15	17	17	17	30	23		33	30	24	14	26	15
16 Chi1_Bacci	10	10	14	14	12	15	17	14	17	16	16	16	29	24	33		33	24	16	26	14
17 Chb_Serma	10	10	10	11	10	13	13	13	15	13	15	15	27	25	30	33		19	14	21	13
18 Cht_Bruma	12	12	11	11	13	17	18	15	15	14	14	17	27	20	24	24	19		15	21	14
19 Chd_Bacci	15	15	15	15	16	14	19	17	19	18	18	13	10	10	14	16	14	15		27	20
20 Cht_Str[illegible]	13	12	12	12	15	13	16	14	16	15	16	15	19	19	26	26	21	21	27		17
21 Cht_Sacer	14	13	13	12	14	12	17	15	[illegible]	16	17	15	15	12	15	14	13	14	20	17	

(b)

Table II.

Amino acid sequence similarity scores among the chitinases belonging to class I (a) and class II (b). The number were derived from the pairwise percentage divergence matrices (percentage similarity = 100 - percentage divergence)

Results and Discussion

We define class I and class II chitinases, those belonging to family 19 and 18 of glycosyl hydrolases, respectively (71). The nomenclature class Ia and class Ib is the same as found in older versions of Swissprot and is used to distinguish the two subgroups of family 19.

Classification of chitinases

Until recently there was a major problem in the classification of chitinases. In previous releases of Swissprot (until November 1992) chitinases were classified in classes I (a and b), II, III and IV. In another classification, three groups of chitinases (I*, II*, III*) were proposed (40). Class Ia and Ib of Swissprot correspond to I* and II* of (40) and classes II, III and IV to III*.

A recent study on the classification of all glycosyl hydrolases by Henrissat (72) proposed the classification of chitinases in two families, named 18 and 19, of glycosyl hydrolases. The results presented here are in agreement with this classification and new unclassified sequences are included. Several sequence fragments included in the classification of Henrissat (71, 72) are not used here.

Class I contains 34 chitinases of higher plants and is separated into two distinct subgroups. Group Ia (25 proteins) has a roughly 40 amino acid Cys-rich N-terminal domain connected to the main structure via a Gly/Pro-rich hinge region. Group Ib (9 proteins) lacks both the Cys-rich domain and the hinge region and perfectly aligns with the Ia group. This class is very well conserved and pairwise percent similarities in aligned sequences are greater than 40% (Table IIa).

In constrast, enzymes belonging to class II (21 proteins), which are found in bacteria, fungi and plants, are not well conserved. Within class II, the eukaryotic chitinases form a distinct subgroup of 13 proteins with an average pairwise similarity from 32-99% between its members: this subgroup is here proposed to be named IIa to underline this close relationship (Table IIb). Chit_Apalb exibits 10-30% similarity to other members of class IIa and 14-38% to bacteria chitinases of class II. It has been included into class IIa because it is eukaryotic. The other 8 members exhibit a similarity to each other ranging from 10-38%. However, the similarity of these 8 sequences to subclass IIa chitinases is only 10 - 19% (Table IIb). Although the sequence similarity within class II chitinases is weak, there are several central sequence segments which are highly conserved as shown in Figure 2. The relatively low similarity scores are due to the different lengths of the members of class II. A detailed representation of similarity scores is shown in Table II a and b. Given the sequence similarity it is likely that the members of this class function using an identical enzymatic mechanism.

Multiple sequence alignments

Figure 1 shows the sequence conservation among the chitinases in class I. The presence of the Cys-rich domain within group Ia and the Gly/Pro-rich hinge region are indicated. In Agi_Urtdi (see nr 52, Table I) the duplication of the Cys-rich domain causes a large gap in the overall alignment.

In contrast, chitinases belonging to class II show only limited regions of homology (Figure 2). Enzymes of this class greatly vary in their size. It is possible that these conserved regions, about 200-300 amino acid long, embody the catalytic chitinase domain. The N and C terminal regions confer additional properties to the enzymes. For example, it was shown that the C-terminus of the yeast chitinase is responsible for the high affinity of the enzyme to chitin (23).

Sequence	Block 1 end	Block 2 end
1.(Ag) Urtdi)	94	193
1.(Chia Maize)	66	144
1.(Chib Maize)	61	133
1.(Chit Brnap)	60	132
1.(Chi4 Phavu)	56	132
1.(Chit Dioja)	42	110
1.(Chib Poptr)	68	150
1.(Chi1 Phavu)	71	140
1.(Chit Phavu)	72	141
1.(Chi1 Alnat)	59	127
1.(Chit Alnat)	64	143
1.(Chi2 Oryza)	65	143
1.(Chi1 Oryza)	62	141
1.(Chit Arath)	68	138
1.(Chit Brn25)	66	136
1.(Chi1 Tobac)	68	145
1.(Chi2 Tobac)	68	145
1.(Chit Tobac)	67	144
1.(Chit Lycea)	60	130
1.(Chi1 Soltu)	72	131
1.(Chit Soltu)	70	144
1.(Chi3 Tobac)	65	150
1.(Chit Pisat)	61	145
1.(Chic Maize)	61	139
1.(Chi8 Poptr)	19	128
1.(Chip Tobac)	19	92
1.(Chiq Tobac)	18	92
1.(Chit Pethy)	11	91
1.(Chi2 Lycea)	19	84
1.(Chi3 Lycea)	0	92
1.(Chi1 Horvu)	16	2
1.(Chi2 Horvu)	0	90
1.(Chi1 Lycea)	14	70
1.(Chi2 Soltu)	[illegible]	81

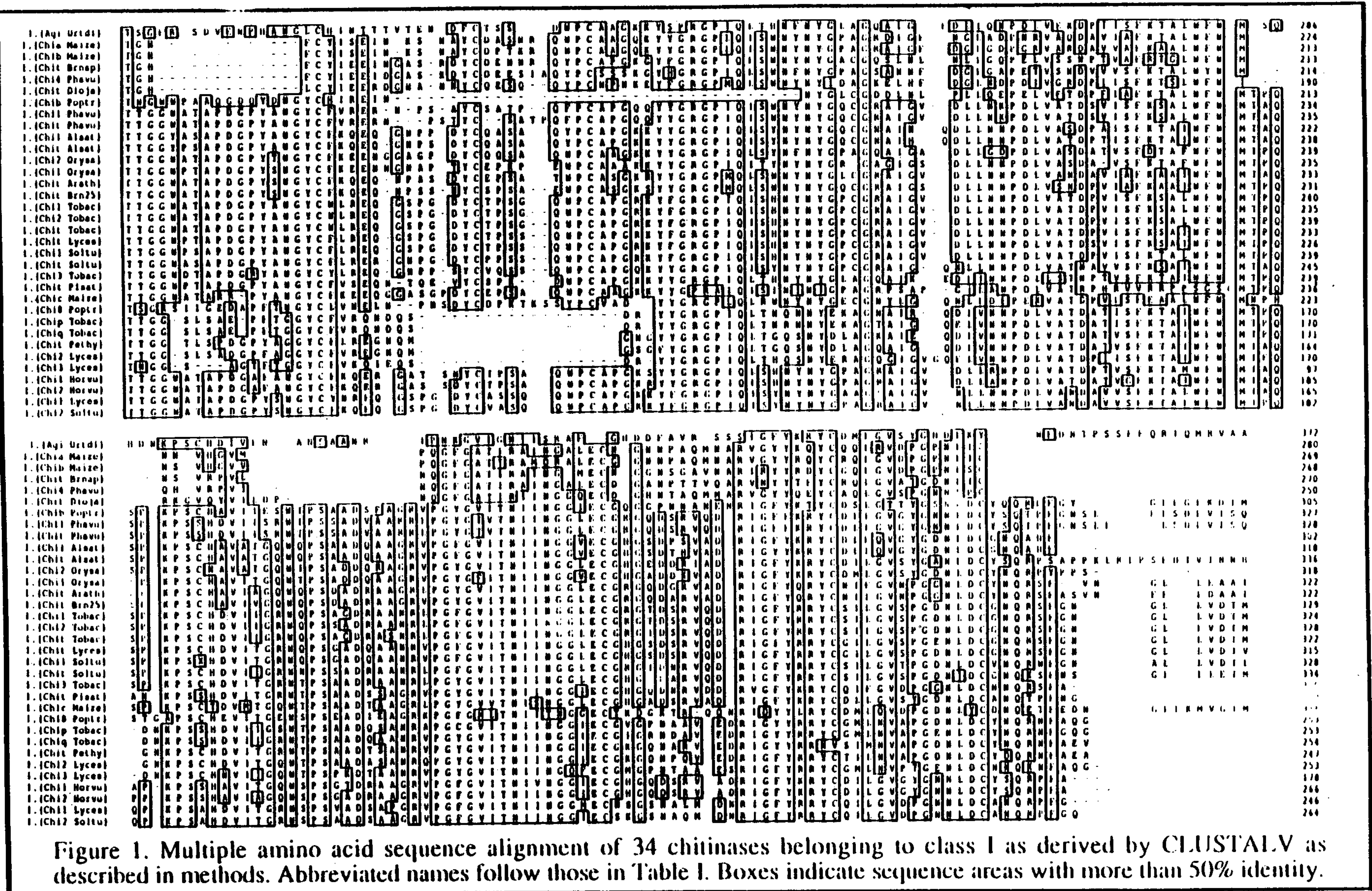

Figure 1. Multiple amino acid sequence alignment of 34 chitinases belonging to class I as derived by CLUSTALV as described in methods. Abbreviated names follow those in Table I. Boxes indicate sequence areas with more than 50% identity.

```
II.(Chi1_Yeast)
II.(Chi2_Yeast)
II.(Chi1_Rhiol)
II.(Chi2_Rhiol)
II.(Chi1_Rhini)
II.(Chia_Arath)
II.(Chiy_Hevbr)
II.(Chia_Tobac)
II.(Chia_Phaan)
II.(Chit_Carle)
II.(Chia_Cucsa)
II.(Chib_Tobac)
II.(Chia_Serma)
II.(Chit_Alter)
II.(Chit_Apalb)
II.(Chi1_Bacci)
II.(Chib_Serma)
II.(Chit_Bruma)
II.(Chid_Bacci)
II.(Chit_Strpl)
II.(Chit_Sacer)
```

```
II.(Chi1_Yeast)
II.(Chi2_Yeast)
II.(Chi1_Rhiol)
II.(Chi2_Rhiol)
II.(Chi1_Rhini)
II.(Chia_Arath)
II.(Chiy_Hevbr)
II.(Chia_Tobac)
II.(Chia_Phaan)
II.(Chit_Carle)
II.(Chia_Cucsa)
II.(Chib_Tobac)
II.(Chia_Serma)
II.(Chit_Alter)
II.(Chit_Apalb)
II.(Chi1_Bacci)
II.(Chib_Serma)
II.(Chit_Bruma)
II.(Chid_Bacci)
II.(Chit_Strpl)
II.(Chit_Sacer)
```

```
II.(Chi1_Yeast)   96
II.(Chi2_Yeast)   96
II.(Chi1_Rhiol)   108
II.(Chi2_Rhiol)   108
II.(Chi1_Rhini)   106
II.(Chia_Arath)   96
II.(Chiy_Hevbr)   69
II.(Chia_Tobac)   91
II.(Chia_Phaan)   90
II.(Chit_Carle)   92
II.(Chia_Cucsa)   94
II.(Chib_Tobac)   93
II.(Chia_Serma)   236
II.(Chit_Alter)   235
II.(Chit_Apalb)   90
II.(Chi1_Bacci)   126
II.(Chib_Serma)   61
II.(Chit_Bruma)   66
II.(Chid_Bacci)   239
II.(Chit_Strpl)   296
II.(Chit_Sacer)   46
```

Sequence	Block 1 end	Block 2 end	Block 3 end
II (Chi1 Yeast)	165	247	351
II (Chi2 Yeast)	165	247	356
II (Chi1 Rhiol)	174	254	361
II (Chi2 Rhiol)	174	254	363
II (Chi3 Rhiol)	172	251	353
II (Chia Arath)	164	246	302
II (Chly Hevbr)	135	217	273
II (Chia Tobac)	157	237	291
II (Chia Phaan)	164	242	290
II (Chit Carte)	157	236	293
II (Chia Cucsa)	160	236	292
II (Chib Tobac)	158	218	294
II (Chia Serma)	327	415	491
II (Chit Alter)	325	420	522
II (Chit Apalb)	177	265	332
II (Chi1 Bacci)	216	302	387
II (Chit Serma)	153	253	357
II (Chit Bruma)	153	230	319
II (Chio Bacci)	316	381	433
II (Chit Strpl)	390	476	547
II (Chit Sacer)	130	211	265

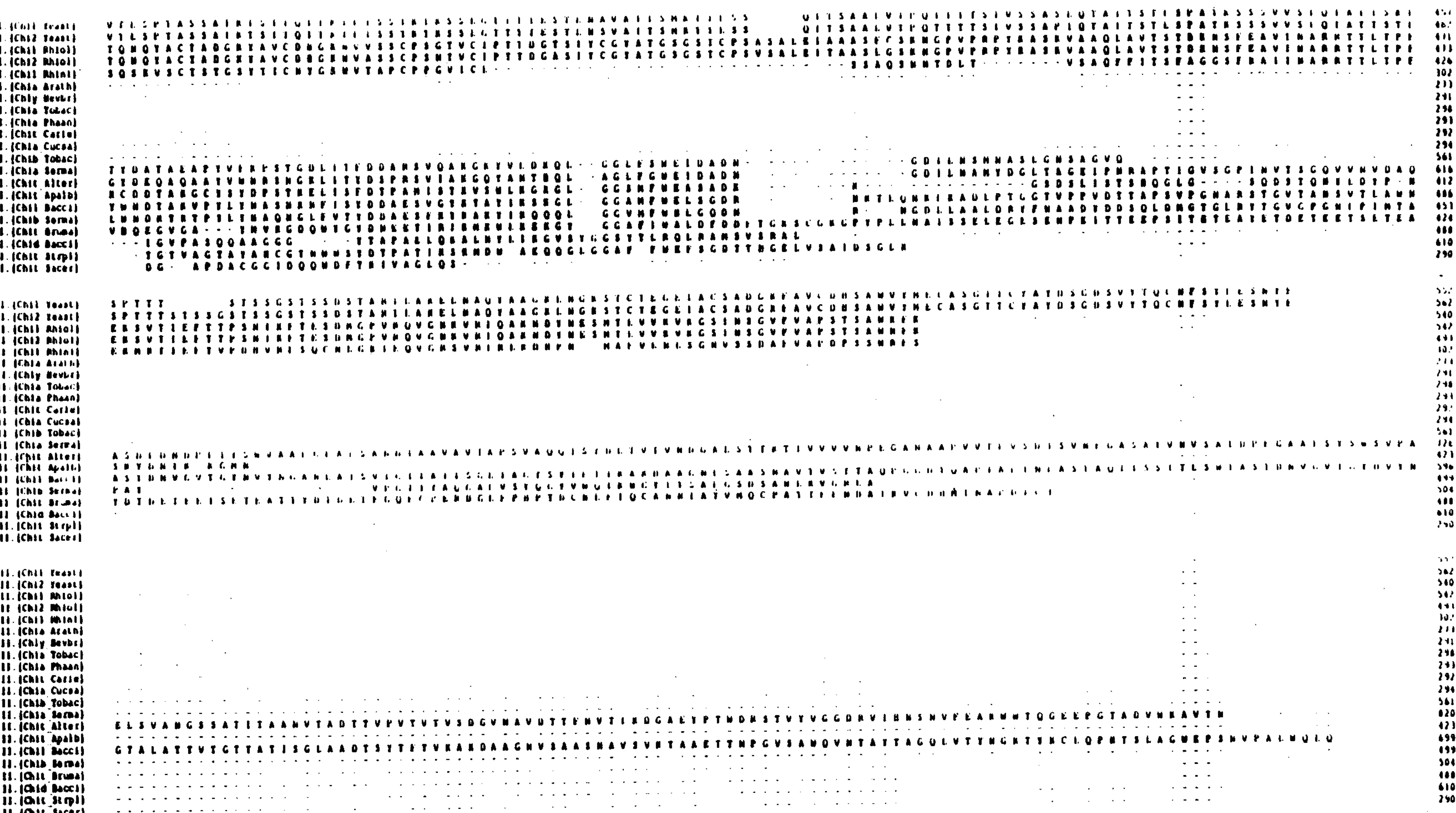

Figure 2. Multiple amino acid sequence alignment of 21 chitinases belonging to class II as derived by CLUSTAL V as described in methods. Abbreviated names follow those in Table I. Boxes indicate sequence areas with more than 50% identity.

Sequence profile searching

Alignment profiles were constructed for both aligned classes. The Swissprot databank was searched for the identification of possible similarities. The profile of class I failed to identify any other sequences except those that were used to construct it. In contrast, the profile of class II identifies a similarity with the α, β, γ chains of hemoglobin from several species.

The results of the sequence profile searches also support this classification scheme and the absence of detectable homology between the two chitinase classes.

Evolutionary aspects

Several approaches were used to construct phylogenetic trees. These included several distance methods as well as a parsimony method, implemented in ClustalV and Phylip 3.5c (66 - 69). The constructed trees show no significant differences in topology. However, different branch lengths can be calculated using different formulas to produce the distance matrix from the multiple alignments.

The trees presented in Figures 3 and 4 were derived from ClustalV. Trees were rooted with the midpoint rooting method, which simply places the origin of the tree in the midpoint between the two most distant sequences. This approach does not require an additional hypothesis for the root of the tree. We decided to present a rooted tree because it is easier to follow, particularly when a large number of sequences is included.

According to this classification there are two major questions concerning the evolution of chitinase genes. The evolution of the Cys-rich N-terminal domain in class I and the evolution of the two chitinase families.

Concerning the evolution of the Cys-rich N-terminal domain two different evolutionary events may have occured. Either there was an ancestral "proto-chitinase" gene bearing this domain which was then excised to give rise to class Ib, or this domain was introduced to a "proto-chitinase" gene by a transposition event and gave rise to class Ia.

Shinshi *et. al* (40) describes the presence of direct sequence repeats at the edges of the DNA sequence encoding for the Cys-rich domain. These are characteristic of plant transposons and indicate the potential process of one or more transposition events. These were also found in other chitinases lacking the Cys-rich domain. That supports the hypothesis of pre-existence of this domain in the "proto-chitinase" gene which was then excised to give rise to class Ib.

As shown in Figure 3, class Ia and Ib did not diverge near the root of the tree and the members of the two subclasses are not clustered but evenly distributed along the tree. For example: Chi1_Horvu (class Ib) is closer to Chi1_Orysa (class Ia) than to Chiq_Tobac (class Ib), although both belong to the same subclass (Figure 3). This indicates that Chiq_Tobac diverged from Chi1_Horvu prior to the divergence of Chi1_Orysa from Chi1_Horvu. Thus we assume that the absence of the Cys-rich N-terminal domain from Chi1_Horvu and Chiq_Tobac (class Ib) seems to be due to an independent excision event. Otherwise Chi1_Horvu and Chiq_Tobac should be expected to be evolutionary closer. Alternatively, the presence of this domain in Chi1_Orysa (class Ia) could have appeared via a recent transposition event in its ancestral gene. Combined with the observation that the two subclasses seem not to diverge near the root, the process of several transposition events could be assumed. These can be responsible for this evolutionary inconsistency between these two subclasses.

The Cys-rich domain has been found in several other proteins such as wheat germ agglutinin, wound-induced proteins, either isolated or in tandem (Table I, Agi_Urdti). This support the excision/insertion hypothesis. It should be pointed out that the division of class I in a and b subclasses has a purely structural meaning.

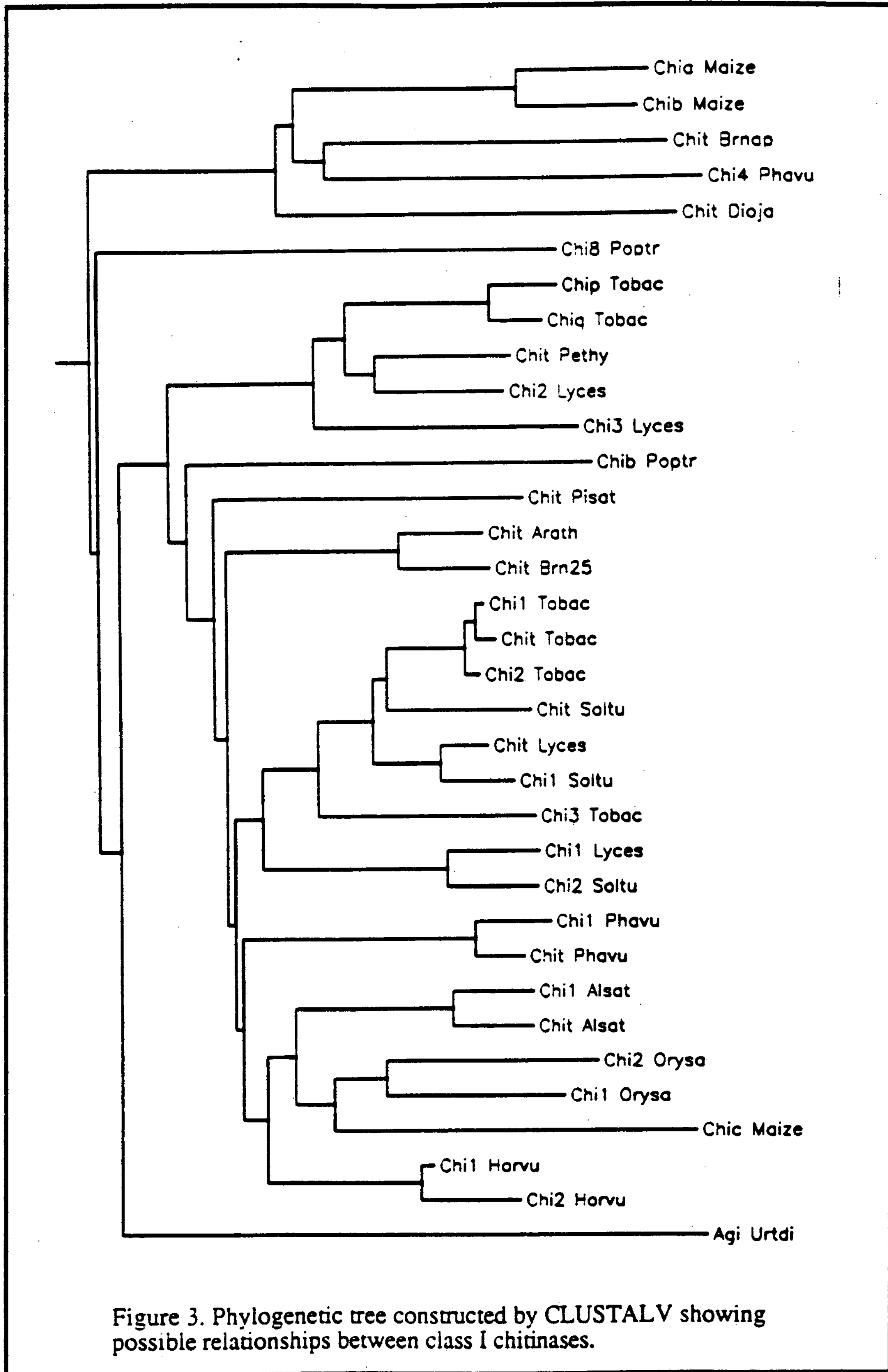

Figure 3. Phylogenetic tree constructed by CLUSTALV showing possible relationships between class I chitinases.

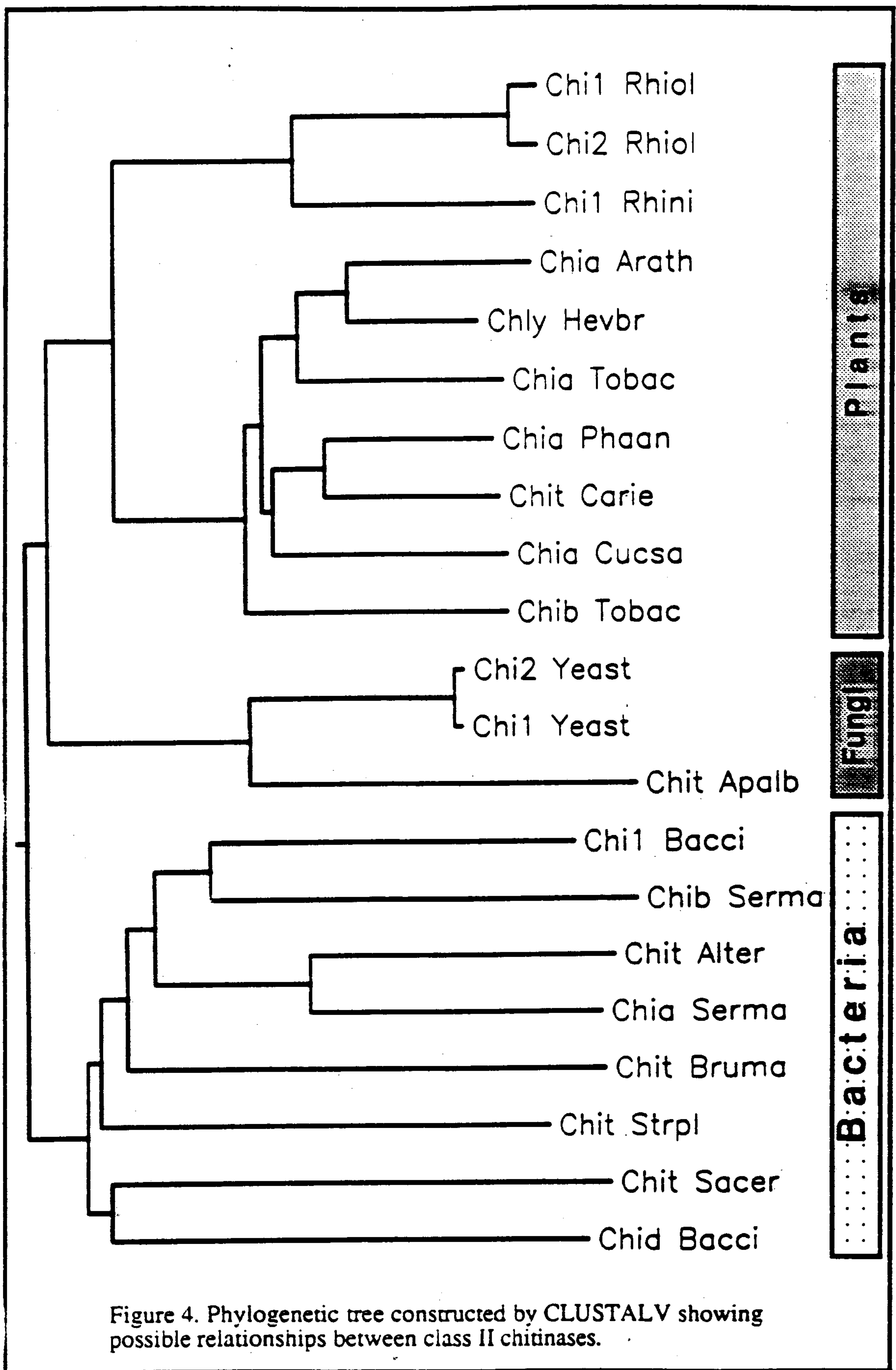

Figure 4. Phylogenetic tree constructed by CLUSTALV showing possible relationships between class II chitinases.

The second question concerns the evolution of the two chitinase families. The absence of any similarity between the two families could be due to two different evolutionary pathways. Either the chitinase classes evolved independently, probably from two different ancestral genes that acquired chitinolytic activity or they evolved from the same gene diverged very early in the evolutionary process. Present data strongly suggest that the two enzyme families represent two independent evolutionary solutions for an identical or similar activity.

The two classes are different without common structural domains. It is very likely that they represent different enzymatic chitin degrading mechanism. The plant chitinases function mainly as a plant defense mechanism attacking the cell wall of the invading pathogen. In contrast, the bacterial and fungal enzymes main function is to degrade chitin to smaller chitobiose dimers to be used as energy, carbon and nitrogen source. It is suggested that class I chitinase activity which is found exclusively in plant kingdom is an endochitinase. The class II enzymes act as exochitinases. Conclusive evidence for this hypothesis is as yet lacking.

We are currently working on the structure of Chitinase A from *Serratia marcescens* (73) which belongs to class II. Recently the crystal structure of an endochitinase from *Hordeum vulgare* (class I) was published (74). We hope that structural information will support the elucidation of the structure-function relatioships between these two classes.

Acknowledgments

We are greatful to D. Higgins for helpful discussions and advice and B. Henrissat for constructive criticism of the manuscript.

References

1. Collinge, D. B., Slusarenko, A. J. (1987) Plant Mol. Biol. **9**, 389-410.
2. Collinge, D. B., Ktagh, K. M., Mikkelsen, J. D., Nielsen, K. K., Rasmussen, U. and Vad, K. (1993) Plant J. **3**, 31-40.
3 De Jong, A., Cordewener, J., Lo Schiavo, F., Terzi, M., Vandekerchove, J., Van Kammen, A. and De Vries, S. C. (1992) Plant Cell **4**, 425-433.
4. Lotan, T., Ori, N. and Fluhr, R. (1989) Plant Cell **1**, 881-887.
5. Linthorst, H. J. M. (1991) Crit. Rev. Plant Sci. **10**, 123-150.
6. Broglie, K., Chet, I., Holliday, M., Cressman, R., Biddle, P., Knowlton, S., Mauvais, C. J. and Broglie, R. (1991) Science **254**, 1194-1197.
7. Boller, T. (1987) In *Plant-Microbe Interactions, Molecular and Genetic Perspectives,* Volume 3 (oshuge, T. and Nester, E. W., eds). New Yotk: Macmillan, pp. 385-411.
8. Boller, T. (1988) In *Surveys of Plant Molecular and Cell Biology,* Volume 5 (Mifflin, B. J.,. ed.). Oxford: Oxford University Press, pp. 145-174.
9. Oppenheim A. B. and Chet, I. (1992) TIBTEC **10**, 392-394.
10. Bowels, D. J. (1990) Ann. Rev. Biochem. **59**, 873-907.
11. Tsujibo, H. (Submitted in EMBL Data Bank, 25. 11. 1992).
12. Watanabe, T., Suzuki, K., Oyanagi, W., Ohnishi, K. and Tanaka, H. (1990) J. Biol. Chem. he **265**, 15659-15665.
13. Watanabe, T., Oyanagi, W., Suzuki, K., Ohnishi, K. and Tanaka, H. (1992) J. Bacteriol. **174**, 408-414.
14. Kamei, K., Yamamura, Y., Hara, S. and Ikenama, T. (1989) J. Biol. Chem. **105**, 979-985.
15. Harpster, M.. (1989) Nucleic Acid. Res. **17**, 5395-5395.

16. Jones, J. D.G., Grady, K. L., Suslow, T. V. and Bedbrook, J. R. (1986) EMBO J. **5**, 467-473.
17. Miyashita, K;, Fujii, T. and Sawada, Y. (1991) J. Gen. Microbiology **137**, 2065-2072.
18. Robbins, P. W., Overbye, K., Albright, C., Benfield, B., Pero, J. (1992) Gene **111**, 69-76.
19. Robbins, P. W., Albright, C. and Benfield, B. (1988) J. Biol. Chem. **263**, 443-447.
20. Blaiseau, P-L. and Lafay, J-F.. (1992) Gene **120**, 243-248.
21. Blaiseau, P. L., Kunz, C., Grison, R., Berthean, Y. and Brygoo, Y. (1992) Curr. Genet. **21**, 61-66.
22. Vasseur, V., Arigoni, F., Andersen, H., Defaco, G., Bompeix, G. and Seng, J-M. (1990) J. Gen. Microb. **136**, 2561-2567.
23. Kuranda, M. and Robbins, P. W. (1991) J. Biol. Chem. **266**, 19758-19767.
24. Sor, F. and Fukuhara, H. (1985) Curr. Genet. **9**, 147-155.
25. Stark, M. J. R., Mileham, A. J., Romanos, M. A. and Boyd, A. (1984) Nucleic Acid Res. **12**, 7581-7595.
26. Tokunaga, M., Kawamura, A. and Hishinuma, F. (1984) Nucleic Acid Res. **12**, 7581-7597.
27. Bradshaw, H. D. Jr. (1990) Nature **345**, 299-299 (1990)
28. Fuhrman, J. A, Lane, S. W., Smith, R. F., Piessens, W. F. and Perler, F. B. (1992) Proc. Natl. Acad. Sci. U. S. A. **89**, 1548-15525.
29. van Damme E. J., Willems, R. and Peumans, W. (Submitted in EMBL Data Bank, 1992)
30. Samac, D. A., Hironaka, C. M., Yallaly, P. E., and Shah, D. M. (1990) Plant Physiol. **93**, 907-914.
31. Hamel, F. and Bellemare, G. (Submitted in EMBL Data Bank, 1992)
32. Rasmussen, U., Bojsen, K. and Collinge, D. B. (1992) Plant Mol. Biol. **20**, 277-287.
33. Vogelsang, R. and Barz, W. (Submitted in EMBL Data Bank, 1993)
34. Metraux, J. P. Burkhart, W., Moyer, M., Dincher, S., Middlesteadt, W., Williams, S., Payne, G., Carnes, M. and Ryals, J. (1989) Proc. Natl. Acad. Sci. U. S. A. **86**, 896-900.
35. Araki, T., Funatsu, J., Kuramoto, M., Konno, H. and Torikata, T. (1992) J. Biol. Chem. **267**, 19944-19947.
36. Jekel, P. A., Hartmann, J. B. H. and Beintema, J. J. (1991) Eur. J. Biochem. **200**, 123-130.
37. Swegle, M., Huang, J. K., Lee, G., and Muthukrishnan, S. (1989) Plant. Mol. Biol. **12**, 403-412.
38. Leah, R., Mikkelsen, J. D., Mundy, J. and Svendsen, I. (1987) Carlsberg. Res. Comm. **52**, 31-37.
39. Danhash, N., Wagemakers, C. A. M., van Kam, J. A. L. and de Wit, P. J. G. (Submitted in EMBL Data Bank, Sept. 1992)
40. Shinshi, H., Neuhaus, J-M., Ryals, J. and Meins, F. Jr. (1990) Plant Mol. Biol. **14**, 357-368.
41. Shinshi, H., Mohnen, D. and Meins, F. Jr. (1987) Proc. Natl. Acad. Sci. U. S. A. **84**, 89-93.
42. van Buuren, M., Neuhaus, J. M., Shinshi, H., Ryals, J. and Meins, F. (1992) Mol. Gen. Genet. **232**, 460-469.
43. Lawton, K., Ward, E., Payne, G., Moyer, M. and Ryals, J. (1992) Plant Mol. Biol. **19**, 735-743.
44. Linthorst, H. J. M., Van Loon, L. C., van Rossum, C. M. A., Mayer, A., Bol. J. F., Van Roekel, J., Meulenhof, J., Conelissen, B. J. C. (1990) Mol. Plant Microbe. Interactions. **3**, 252-258.

45. Neale, A. D., Wahleithner, J. A, Lund, M., Bonnet, H. T., Kelly, A., Meeks-Wagner, D. R., Peacock, W. J. and Dennis, E. S. (1990) Plant Cell **2**, 673-684.
46. Huang, J-K., Wen, L., Swegle, M., Tran, H-C., Thin, T. H., Naylor, H. M., Muthukrishnan, S. and Reeck, G. R. (1991) Plant Mol. Biol. **16**, 479-480.
47. Zhu, Q. and Lamb, C. J. (1991) Mol. Gen. Genet. **226**, 289-296.
48. Broglie, K. E., Biddle, P., Cressman, R. and Broglie, R. (1989) Plant Cell **1**, 599-607.
49. Margis-Pinheiro, M., Metz-Boutigue, M. H., Awade, A., De Tapia, M., Le Ret, M. and Burkard, G. (1991) Plant Mol. Biol. **17**, 243-253.
50. Broglie, K. E., Gaynor, J. J. and Broglie, R. M. (1986) Proc. Natl. Acad. Sci. U. S. A **83**, 6820-6824.
51. Lucas, J., Henschen, A., Lottspeich, F., Voegelt, U. and Boller, T. (1985) FEBS Lett. **193**, 208-210.
52. Hedrick, S. A., Bell, J. N., Boller, T and Lamb, C. J. (1988) Plant Physiol. **86**, 182-186.
53. Vad, K., Mikkelsen, J. D. and Collinge, D. B. (Submitted in EMBL Data Bank, 1992)
54. Davis, J. M., Clarke, H. R. G., Bradshaw, H. D. and Gordon, M. P. (1991) Plant Mol. Biol. **17**, 631-6394.
55. Parson, T. J., Bradshaw, H. D. Jr. and Gordon, M. P. (1989) Proc. Natl. Acad. Sci. U. S. A. **86**, 7895-7899.
56. Yanai, K., Takaya, N., Kojima, N., Horiuchi, H., Ohta, A. and Takagi, M. (1992) J. Bacteriol. **174**, 7398-7406.
57. Laflamme, D. and Roxby, R. (1989) Plant Mol. Biol. **13**, 249-250.
58. Gaynor, J. J. (19889 Nucleic Acids Res. **16**, 5210-5210.
59. Wemmer, T., Kaufmann, H., Lottspeich, F., Thompson, R. D., Kirch, H. H. and Schneider, K. (Submitted in EMBL Data Bank, 1992)
60. Lerner, D. R. and Raikhel, N. V. (1992) J. Biol. Chem. **267**, 11085-11091.
61. Beintema, J. J. and Peumans, W. J. (1992) FEBS Lett. **299**, 131-134..
62. Ishige, F., Mori, H., Yamazaki, K. and Imaseki, K. (Submitted in EMBL Data Bank, 1992)
63. Huynh, Q. K., Hironaka, C. M., Levine, E. B., Smith, C. E. Borgmeyer, J. R. and Shah, D. M. (1992) J. Biol. Chem. **267**, 6635-6640.
64. Kriz, A. L. and Wu, S. (Submitted in EMBL Data Bank, 1992)
65. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. Microbiol. Rev. **55**, 303-315.
66. Higgins, G. D., Bleasby, J. A. and Fuchs, R. (1992) CABIOS **8**, 189-191.
67. Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c Distributed by the author, Dep. of Genetics, University of Washington, Seattle.
68. Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. **4**, 406-425.
69. Felsenstein, J. (1985) Evolution **39**, 783-791.
70. Henrissat, B. (1990) Protein Seq. Data. Anal. **3**, 523-526.
71. Henrissat, B. (1991) Biochem. J. **280**, 309-316.
72. Henrissat, B. and Bairoch, A. (1993) Biochem. J. (in press).
73. Vorgias, C. E., Kingswell, A. J., Dauter, Z. and Oppenheim, A. B. (1992) J. Mol. Biol. **226**, 897-898.
74. Hart, P. J., Monzingo, A. F., Ready, M. P., Ernst, S. R. and Robertus, J. D. (1993) J. Mol. Biol. **229**, 189-193.