FEBS Letters 352 (1994) 131–136

FEBS 14548

Comparative modelling of barley-grain aspartic proteinase: a structural rationale for observed hydrolytic specificity

Kunchur Guruprasad[a], Kirsi Törmäkangas[b], Jukka Kervinen[b], Tom L. Blundell[a,*]

[a]*Laboratory of Molecular Biology, Department of Crystallography, Malet Street, Birkbeck College, University of London, London, WCIE 7HX, UK*
[b]*Institute of Biotechnology, PO Box 45, Karvaamokuja 3 A, University of Helsinki, FIN-00014, Finland*

Received 1 July 1994; revised version received 1 August 1994

Abstract A model of the barley-grain aspartic proteinase (HvAP; *Hordeum vulgare* aspartic proteinase) has been constructed using the rule-based comparative modelling approach encoded in the COMPOSER suite of computer programs. The model was based on the high resolution crystal structures of six highly homologous aspartic proteinases. Results suggest that the overall three-dimensional structure of HvAP (excluding the plant-specific insert; 104 residues in HvAP) is closer to human cathepsin D than other aspartic proteinases of known three-dimensional structure. Comparisons of the complexes with the substrate modelled in the active site of HvAP with those of the same substrate modelled in the active site of other aspartic proteinases of known three-dimensional structure and specificity, define residues that may influence hydrolytic specificity of the barley enzyme. We have identified residues in the S_4 (Ala^{12}), S_3 (Gln^{13}, Thr^{111}), S_2 (Ala^{222}, Thr^{287}, Met^{289}), S_1' and S_3' (Ile^{291}), S_2' and S_3' (Gln^{74}), S_2' (Arg^{295}), and S_3' (Pro^{292}) pockets, that may account for the observed trends in the kinetic behaviour and specificity when compared to other aspartic proteinases. The plant-specific inserted sequence, which may play a role in the transport of HvAP to plant vacuoles (lysosomes), is similar to the saposins and is predicted to be a mixed α-helical and β-strand domain.

Key words: Hordeum vulgare; Barley grain; Aspartic proteinase; Hydrolytic specificity; Protein structure prediction; Specificity pocket

1. Introduction

Aspartic proteinases [1,2] (EC 3.4.23) are widely dispersed in the plant kingdom ([3] and references therein) but their structures and biological functions are less well characterized than those of animal, microbial and viral aspartic proteinases. Barley aspartic proteinase, HvAP, exists as two enzymatically active, two-chain forms in vivo: 48kDa (32 + 16 kDa) and 40 kDa (29 + 11 kDa) [4], probably due to sequential processing of the same proenzyme. Sequence alignment of HvAP with animal and microbial aspartic proteinases [5] shows that major parts of HvAP are similar to other aspartic proteinases, especially to mammalian lysosomal cathepsin D and yeast proteinase A. However, HvAP contains a sequence of 104 amino acid residues bearing no similarity to animal or microbial aspartic proteinases. This 'large insertion' region is partly removed during processing of the 48 kDa enzyme to the 40 kDa form.

The structural similarity of HvAP to other aspartic proteinases extends to its catalytic properties: HvAP preferentially cleaves peptide bonds between amino acid residues with large hydrophobic side chains [6]. The specificity constants (k_{cat}/K_m) for HvAP, yeast proteinase A [7] and human cathepsin D [7] were very similar but about 10-fold smaller than those of porcine pepsin [7] and human cathepsin E [8]. Previous studies have demonstrated the value of rule-based structural modelling in exploring substrate interactions of related aspartic proteinases ([9] and references therein). Our earlier prediction of the hydrolytic specificity based on the three-dimensional model for human cathepsin D [9] has been recently supported by site-directed mutagenesis experiments [10].

In this work, we have constructed a three-dimensional model for HvAP (corresponding to the 48 kDa enzyme excluding the barley insert) and modelled the substrate Lys-Pro-Ile-Glu-Phe-Phe(4-NO_2)-Arg-Leu [7], into the active site of HvAP and several enzymes, in order to make a comparative study of the structural basis for the observed hydrolytic specificity in the barley enzyme. In addition, we have predicted the secondary structure for the barley insert.

2. Materials and methods

2.1. Modelling

A rule-based automated approach to comparative protein modelling [11–14] encoded in the COMPOSER suite of computer programs and based on the crystal structures of six aspartic proteinases (porcine pepsin [15], bovine chymosin [16], human renin [17], mouse renin [18], human cathepsin D [19], yeast proteinase A [20]) which have greater than 40% sequence identity to the sequence of HvAP, was used to model the barley-grain aspartic proteinase. The model was further refined using energy minimization techniques (SYBYL; Tripos Co., USA).

The sequences of these six aspartic proteinases were aligned together with the sequence of the barley-grain aspartic proteinase using the multiple sequence alignment program MALIGN [21]. The 'large insertion' and three additional residues, Ser^{240}, Pro^{241} and Met^{242}, were not modelled. This insertion is unique to the plant aspartic proteinases. We do not yet have a crystal structure for any of the plant aspartic proteinases and therefore it is not possible to model this region. The structures of the six aspartic proteinases chosen for modelling were superimposed as rigid bodies by first specifying the active site residues (Asp^{32}–Gly^{34} and Asp^{215}–Gly^{217}) as topological equivalences. An iterative re-weighted least-squares procedure was then used to define the new set of topologically equivalent residues using a distance cut-off value of 3.0 Å. This procedure determined the structurally conserved regions (SCRs) that make up the 'framework'. The sequence of HvAP was then aligned to a template representing the sequences of the known structures in the 'framework' region. Fragments from these structures with the closest similarity to HvAP were least-squares fitted to the 'framework' in order to give the model real protein geometry. The contributions of the known structures are weighted by the square of the sequence similarity [22], to model the SCRs in HvAP.

In order to model the structurally variable regions (SVRs), protein fragments, which are of the same length as the loop to be modelled and are compatible with the geometry of a three-residue overlap with the SCR on either side, were first identified by searching a local distance matrix database of all structures, including members of the aspartic

*Corresponding author. Fax: (44) (71) 631 6805.

SSDI 0014-5793(94)00935-X

proteinase family. Several filters were then successively applied in the selection of the most appropriate fragment to be annealed to the model core [23–25].

The side chains were modelled using a set of environment-dependent rules [12] which use knowledge derived from both the side chain conformations of residues at equivalent positions within the homologous family, and from the most probable side chain conformations in the secondary structural environment. The COMPOSER built model was further refined on an Evans and Sutherland PS390 interactive graphics system using Hermans and McQueen's options [26] in FRODO [27], and energy minimization techniques on a Silicon Graphics workstation using SYBYL software supplied by TRIPOS Associates, USA.

2.2. Substrate modelling

The three-dimensional structure of endothiapepsin complexed with the pepsin inhibitor H-256 [28] was used as a guide to model the transition state of the substrate Lys-Pro-Ile-Glu-Phe-Phe(4-NO_2)-Arg-Leu, into the active site of these enzymes. H-256 was chosen because it contains a Pro at P_4, Glu at P_2, Phe at P_1 and P_1', and Arg at P_2', as in the substrate used. The substrate modelling was achieved by superposing each of these enzymes onto the enzyme–inhibitor complex of endothiapepsin and then transferring the coordinates of the inhibitor on to the model structures. These models of the inhibitor complexes of the different enzymes were then used to model the sequence of the substrate into the active site. Specificity pockets for the models complexed with the substrate are listed in Table 1. These pockets were compared to identify significant interactions that might affect the observed substrate specificities.

2.3. Coordinates

The coordinates of the model have been deposited in the Brookhaven Protein Data Bank [29,30] under the file name (1) HVAP-SUBS.PDB (substrate model) (also see footnote in Table 1).

2.4. Structure prediction of the 'large insert' between Gly^{239} and Gly^{243} in HvAP

In order to check for proteins of known three-dimensional structure that might be useful for modelling the 107 residue insert in HvAP, we used the computer programs QSLAVE and PSLAVE [32], which compare templates (either derived from structure or sequence) against a single sequence or a set of aligned sequences using the computer program MALIGN [21]. Structural templates have been constructed for the database of aligned three-dimensional structures of related proteins as well as individual structures [33]. The sequences corresponding to the plant-specific insert were aligned using MALIGN and searched against all of the structural templates using QSLAVE. Likewise, the aligned set of plant-specific insert sequences were searched against all sequences in the protein sequence data bank.

We have predicted the secondary structure corresponding to the insert using two recently developed methods which require several aligned homologous sequences: SAPIENS [34,35] and PHD [36,37]. SAPIENS uses amino acid environment-dependent substitution tables and conformational propensities. This method requires as its input a set of homologous sequences that are aligned using the multiple sequence alignment program MALIGN [21]. SAPIENS [34,35] has a relatively higher mean accuracy of prediction than the GOR method [38] and gives fewer poor predictions. The confidence in secondary structure prediction using this method increases if there is a large number of aligned sequences with a high percentage of sequence identity. The PHD method is based on neural networks and also requires multiple related sequences as input for obtaining greater than 70% three-state (helix, strand, loop) prediction accuracy for globular proteins. The gene sequences for two other plant aspartic proteinases from rice [39] and cardoon (referred to as 'cyprosin') [40], which have been recently cloned, were used in the secondary structure prediction of the plant-specific barley insert.

3. Results and discussion

3.1. Model validity, selection of SCRs and SVRs

The PROCHECK suite of programs [41] was used to assess the stereo-chemical quality of the model. The overall rating for the model is 'good'; 87% of the residues are found to occur in the most favoured regions (A,B,L) of the Ramachandran plot. Six residues (V95, D98, E161, F205, A278d and A280) fall in the generously allowed regions of the map and two residues (A279, S202) fall in the disallowed regions. Most of these residues are either close to or within the structurally variable regions (SVRs) in the model; however, none is close to the active site and therefore the conclusions drawn regarding the specificity are not affected.

The model is closest to human cathepsin D (Table 2); 93% of the Cα residues are topologically equivalent between the two enzymes, with an rmsd of 0.67 Å. A schematic representation, showing the arrangement of β-strands, α-helices together with the side chains of the substrate from P_4 to P_3' and the site of the 'large insertion' is shown in Fig. 1. The presence of serine in the second active site triad (Asp Thr/Ser Gly) of the plant aspartic proteinases, in contrast to the more common Thr^{216} in mammalian and microbial aspartic proteinases, is unlikely to alter the conformation at the active site and affect overall specificity. The model indicates that the insert is in the C-domain near the surface, and is present after the long helix (Thr^{225}– Ile^{235}) between a short β-strand (Ala^{238}–Gly^{239}) and a slightly longer β-strand Ser^{245}–Asp^{248} (see Fig. 1).

3.2. Specificity pockets

The specificity constants (k_{cat}/K_m) measured at pH 3.1 and at 37°C using 0.1 M sodium formate buffer for cleavage of the substrate Lys-Pro-Ile-Glu-Phe-Phe(-4-NO_2)-Arg-Leu by the barley-grain aspartic proteinase (HvAP) [6], human cathepsin D [7] and yeast proteinase A (YPRA) [7] are very similar and differ significantly from the specificity constants for cleavage of the same substrate by either porcine pepsin (5PEP) [7] or human cathepsin E [8] (CATE) (see Table 1). The differences in the specificities of these enzymes may be due to the differences in interactions between residues in the enzyme binding

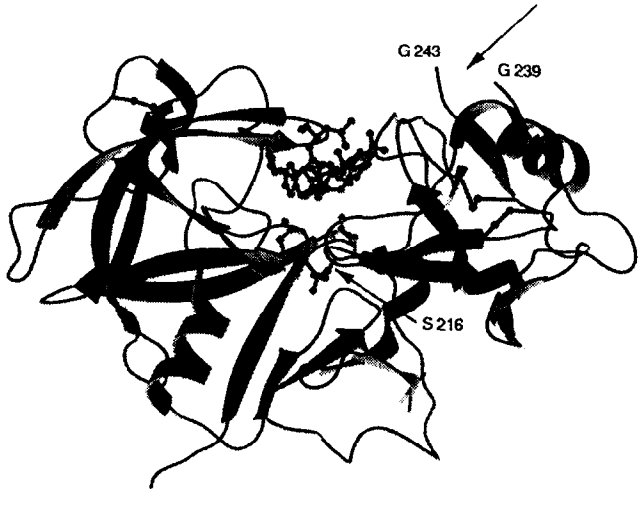

Fig. 1. Schematic representation showing the overall 'fold' of HvAP, generated using the program SETOR [45]. Helices are shown as cylinders and sheets as flat ribbons. The remaining polypeptide chain is shown as a rope which represents the turn/coil regions in the molecule. The location of the 'large insert' is indicated by an arrow. The substrate modelled into the active site, the disulphide bridges, the catalytic aspartate side chains (Asp^{32} and Asp^{215}) and Ser^{216}, unique to the plant aspartic proteinases as opposed to the more common Thr^{216} in the mammalian and fungal aspartic proteinases are also shown. The side chain of Thr^{33} in the active site is also shown.

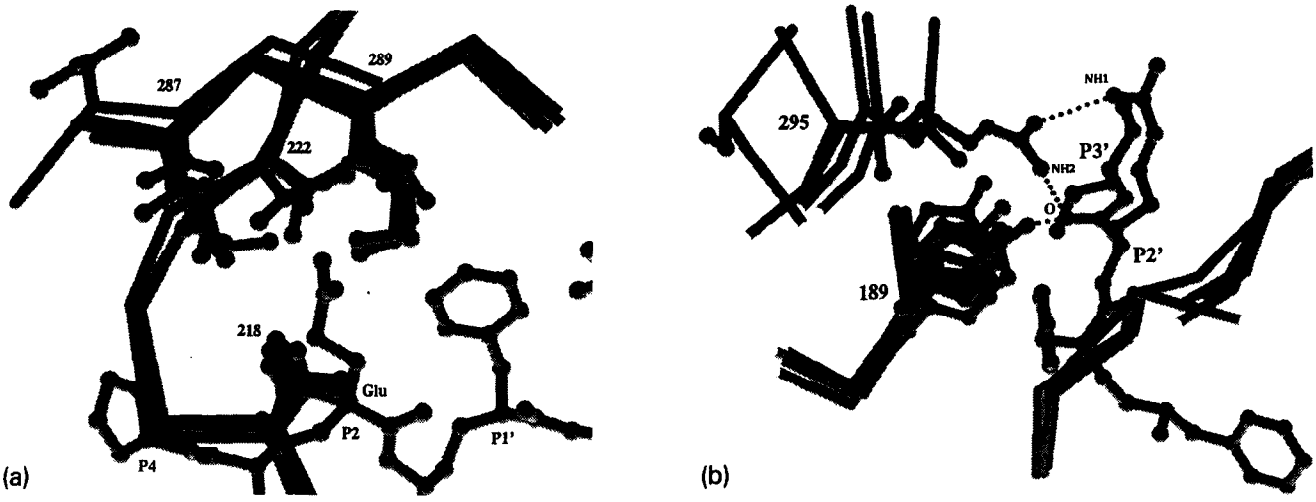

Fig. 2. (A) S_2 pocket in the HvAP model and in the different enzymes showing interactions with P_2 (Glu) of the substrate. Residue numbering in the figure corresponds to pepsin. Note the differences in size of this pocket due to differences in the type of residues at positions 287, 222 and 289. P_4, P_2 and P_1' correspond to residues in the substrate. (B) S_2' pocket in the model showing amino acid residues at position 295 in the different enzymes. Side chain interactions of Arg^{295} via hydrogen bonds (filled spheres) in HvAP with the substrate can be seen. Porcine pepsin is relatively displaced with respect to the other enzymes and does not interact with the substrate. This may be due to lack of the *cis*-prolines Pro^{294} and Pro^{297} common to the other enzymes.

pockets and side chains of the substrate. However, individual enzymes display different pH dependencies and this may also contribute to the observed kinetic variability. Table 1 indicates the residues contributing to the specificity pockets in the differ-

Table 1
Comparison of the specificity pockets in the models of the barley-grain aspartic proteinase (HvAP), human cathepsin D (1LYB), yeast proteinase A (YPRA), porcine pepsin (PDB Code: 5PEP), and human cathepsin E (CATE), with substrate[a] modelled in the active site[b]

Protein resolution	HvAP (model)	1LYB (2.5 Å) [19]	YPRA (2.5 Å) [20]	5PEP (2.3 Å) [15]	CATE (model) [31]
k_{cat}/K_m ($mm^{-1} \cdot s^{-1}$)	149 [6]	195 [7]	185 [7]	2120 [7]	2500 [8]
Pocket					
S_5	–	–	–	–	–
S_4	S219	E244	S219, L220	**F111**, S219	**M12**, S219
S_3	I30, F117, G217, T218, S219	**Q13**, S77, F117, G217, T218, S219	T77, G217, T218, S219,	**E13**, I30, **F111**, F117, G217, T218, S219	**E13**, I30, T77, G217, T218, S219
S_2	Y75, G76, T77, T218, **T287** **M289**	G76, S77, G217, T218, **V222**, **M287, M289**	Y75, G76, T77, T218, **T222** **T287, M289**	G76, T77, G217, T218, **T222, E287, M289**	T77, G217, T218, **Q287, L289**
S_1	D32, Y75, T77 F117, D215, G217, T218	D32, Y75, S77, F112, F117, I120, D215, G217, T218	D32, Y75, T77, I120 D215, G217, T218	I30, D32, Y75, T77, **F111**, F117, D215, G217, T218	I30, D32, T77, F117, D215, G217, T218
S_1'	Y75, G76, Y189, I213, D215, **I291**, I300,	G34, Y75, G76, I213, D215, T218, **M289, I291**, I300	G34, Y75, G76, Y189, D215, **F291**, I300	G34, G76, Y189, I213, D215, **M289** **V291**, I300	G34, D215, T218, **L289**, **I291**, I300
S_2'	**Q74**, Y75, I128, V130, Y189, **R295**	G34, I73, **H74**, I128, Y189	G34, **Q74** I128, Y189	G34, **T74**, I128, Y189	G34, L128, V130, Y189
S_3'	**Q74**, Y75, G76, **I291, P292**	**H74**	**Q74**, Y189, **F291**	**T74**, Y189, **V291**	**Q74, I291**

Coordinates: the models with substrate in the active site of porcine pepsin, yeast proteinase A, human cathepsin D and the human cathepsin E, will be available upon request from K.G. Due to paucity of space, sequence alignments are not presented but can also be provided upon request from K.G.

[a]Substrate: Lys-Pro-Ile-Glu-Phe-Phe(4-NO_2)-Arg-Leu.

[b]The distance cutoff was 4.0 Å; residues are numbered according to porcine pepsin (Protein Data Bank [29,30] code 5PEP) and residue numbering after 230 is corrected in pepsin; bold lettering indicates variant residues in the aspartic proteinases that are likely to affect specificity pockets in each of these enzyme models.

ent enzymes. Significant differences in the types of interactions are observed in the S_4, S_3, S_2, S_1', S_2' and S_3' pockets which may be responsible for the differences in specificities.

From visualisation of the models on graphics, we could infer that the side chain of isoleucine (P_3) interacts with the side chains of Glu^{13} and Phe^{111} in pepsin. In all the other enzymes shown in Table 1, interactions in the S_3 pocket are less favourable due to the replacement of Phe^{111} by a smaller polar residue, threonine. In human cathepsin E the weaker interactions in the S_3 pocket appear to be compensated for by Met^{12} in the S_4 pocket which interacts with proline (P_4). In the barley-grain proteinase, yeast proteinase A and human cathepsin D, a smaller hydrophobic residue, Ala^{12}, at an equivalent position possibly leads to weaker binding. Fig. 2A and Table 1 show that Glu P_2 interacts with Thr^{287} in the S_2 pocket which is large enough to accommodate a residue with a long side chain. Glu^{287} (protonated at pH 3.1) in porcine pepsin or Gln^{287} in human cathepsin E may form hydrogen bonds with P_2 Glu, and this may enhance the catalytic efficiency. However, in yeast proteinase A (YPRA) and barley proteinase (HvAP), residue 287 is the smaller polar threonine. On the other hand, in human cathepsin D (CATD), residue 287 is the bulky and more hydrophobic methionine. Therefore, the corresponding S_2 pocket is either relatively larger in the case of HvAP and YPRA, or relatively smaller in the case of CATD. The size of the S_2 pocket in HvAP is comparable to that in the yeast proteinase A but is larger compared with human cathepsin D due to smaller residues at postions 222 (Ala) and 287 (Thr) in the barley enzyme. This would account for the observation that porcine pepsin and human cathepsin E are more similar to each other in terms of their specificity and differ from barley proteinase, human cathepsin D or the yeast proteinase A.

Pro^{294} and Pro^{297} are highly conserved in mouse and human renins, human cathepsin D, yeast proteinase A, human cathepsin E and in the plant aspartic proteinases, but not in pepsin and chymosin. These prolines are in the *cis*-conformation as observed in the crystal structures of renins, cathepsin D and yeast proteinase A. There is a positively charged residue within this region at position 295 (Arg^{295} in the barley and rice enzymes and His^{295} in cardoon) that appears to be unique to the plant aspartic proteinases. Modelling suggests that Arg^{295} in HvAP may be involved in hydrogen bond interactions with the substrate (Fig. 2B), and therefore influence specificity near the S_2'–S_4' pockets. In contrast there is a small polar group, Ser^{295}, in CATD and a small hydrophobic group, Val^{295}, in YPRA. From the hydrogen bond interactions, we expect Tyr^{189} to influence the specificity at the S_2' pocket, and Gln^{74} to influence the specificity of the barley proteinase at S_2' and S_3' pockets.

3.3. Secondary structure prediction of the 'large insertion' in HvAP

The search against the structural data bank did not yield any structure that could be useful to model the plant-specific insert, but a sequence search revealed that the plant-specific insert shared significant sequence homology to the saposins. Saposins are glycosylated proteins that activate glucosidases that remove the sugar residues from sphingolipids [42], but the mechanism is not understood. We have used the saposin sequence also in the secondary structure prediction.

The results of secondary structure prediction for the region corresponding to the plant-specific insert using the method of SAPIENS are shown in Fig 3a. There is evidence for two large α-helical regions and three β-strands arranged as $loop_1$–$helix_1$–$loop_2$–$strand_1$–$loop_3$–$strand_2$–$loop_4$–$strand_3$–$helix_2$–$loop_5$. $Loop_3$ is likely to be buried whereas $loop_4$ may be exposed. The rice insert, which has 65% sequence identity to the barley insert, and the cardoon insert, which has 54% sequence identity to the barley insert, are predicted to have very similar secondary structures. The results of the overall secondary structure predictions for each of the plant-specific inserts using the PHD method are comparable to those obtained from SAPIENS. The sequence of the plant-specific insert shows some similarity to the saposins; the percentage sequence identity to the barley insert is 33%, to the rice insert is 35% and to the cardoon insert is 31% with five of the six cysteine residues in saposins and plant-specific insert being conserved. The disulphide connectivity, to our knowledge is yet not established. In each case we have used additional information from the saposins; this may assist the identification of the general topology, although the more distant relationship may give rise to problems in the exact lengths of helices. In fact the secondary structure prediction for the barley insert by the SAPIENS method, when the saposins are included, is strikingly similar (see Fig. 3b).

A recent report shows that the association of procathepsin D with prosaposin [43] starts in the rough endoplasmic reticulum and continues in the Golgi. The complex becomes membrane associated in a late Golgi compartment, and the enzyme complex is believed to dissociate in dense lysosomes [43]. The association of these intermolecular lysosomal protein precursors during biosynthesis may play a role in the previously reported mannose-6-phosphate-independent lysosomal targetting of cathepsin D [44], but it is not known which structural parts of the prosaposin or the complex are involved in the targeting. Based on this observation it is probable that there is a similar situation in plants, except that in plants the saposin-like domain is coded as part of the aspartic proteinase and that the saposin-like large insert of HvAP could play a role in the

Table 2
Pairwise percentage sequence identities and pairwise root mean square differences (rmsd) of HvAP with other aspartic proteinases

	Protein*							
	HvAP	5PEP	4CMS	cate	1SMR	1BBS	1LYB	ypra
No. of res.	442	326	320	341	331	331	338	329
% seq. id.	–	46.3	44.1	48.2	45.3	44.1	51.8	47.4
rmsd (Å)	–	0.81	0.81	0.79	0.84	0.75	0.67	0.78
Cα equivalences	–	295	293	308	303	291	316	301
% Cα equivalences		90.4	91.5	90.3	91.5	87.9	93.4	91.4

Note: The 'large insert' in HvAP was not included in the comparisons.
*Protein codes: HvAP, barley-grain aspartic proteinase; 5PEP (PDB Code), porcine pepsin; 4CMS (PDB Code), bovine chymosin; cate, human cathepsin E; 1SMR, mouse renin; 1BBS (PDB Code), human renin; 1LYB (PDB Code), human cathepsin D; ypra, yeast proteinase A.

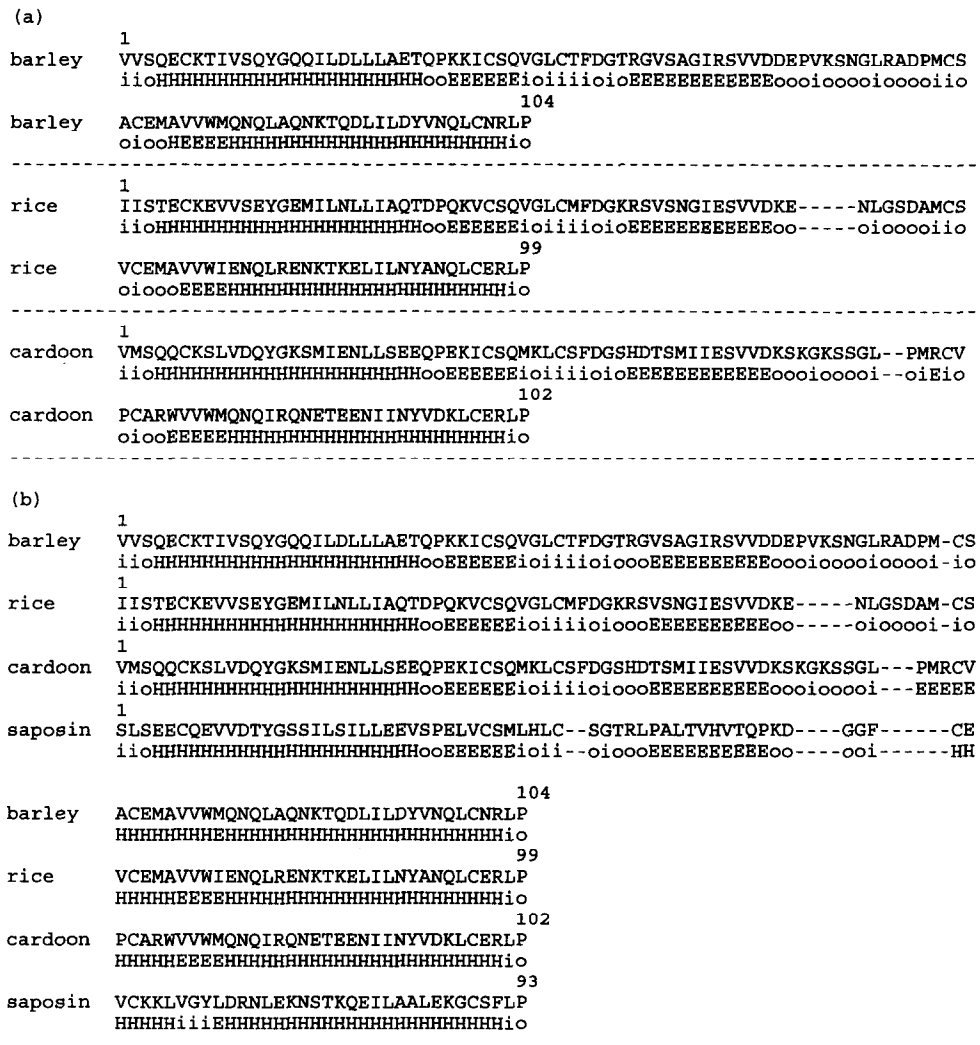

Fig. 3. Secondary structure prediction of the plant-specific insert in barley, rice and cardoon, using (a) the SAPIENS method [34,35]. The amino acid sequence is shown in the first line and the secondary structure prediction is shown in the second line. H = helix, E = sheet, i = buried residue, o = exposed residue. Using the SAPIENS method (b), but including not only the aspartic proteinase sequences but also the sequence of saposin corresponding to the plant-specific insert.

transport of HvAP to plant vacuoles (lysosomes). The role of the plant-specific insert in the lysosomal (vacuolar) targeting is currently under investigation.

4. Conclusions

Computer modelling suggests that several amino acid residues, in particular, Thr^{287}, Ala^{222}, Met^{289}, Arg^{295}, Tyr^{189} and Gln^{74}, are likely to influence the hydrolytic specificity of the barley-grain aspartic proteinase. Studies with site-directed mutants of each or any combinations of these residues should provide a way to test this directly and further our understanding of structure–function relationship on HvAP. The plant-specific insert is probably essential for vacuolar targeting of the plant aspartic proteinases.

Acknowledgements: We thank Prof. John W. Erickson for providing the crystal structure coordinates of human cathepsin D, Dr. Carlos Aguilar for the crystal structure coordinates of yeast proteinase A, and Prof. Peter Brodelius for providing the sequence of the cardoon aspartic proteinase prior to publication. We thank Dr. Mark Johnson for alignment programs, and Prof. John Kay and Dr. Natalia Andreeva for useful discussions. K.G. acknowledges the support of a grant from Parke-Davis.

References

[1] Tang, J. and Wong, R.N.S. (1987) J. Cell. Biochem. 33, 53–63.
[2] Davies, D.R. (1990) Annu. Rev. Biophys. Biophys. Chem. 19, 189–215.
[3] Törmäkangas, K., Kervinen, J., Ostman, A and Teeri, T. (1994) Planta (in press).
[4] Sarkkinen, P., Kalkkinen, N., Tilgmann, C., Siuro, J., Kervinen, J. and Mikola, L. (1992) Planta 186, 317–323.
[5] Runeberg-Roos, P., Törmäkangas, K. and Östman, A. (1991) Eur. J. Biochem. 202, 1021–1027.
[6] Kervinen, J., Sarkkinen, P., Kalkkinen, N., Mikola, L. and Saarma, M. (1993a) Phytochemistry 32, 799–803.
[7] Dunn, B.M., Jimenez, M., Parten, B.F., Valler, M.J., Rolph, C.E. and Kay, J. (1986) Biochem. J. 237, 899–906.
[8] Jupp, R.A., Richards, A.D., Kay, J., Dunn, B.M., Wyckoff, J.B., Samloff, I.M. and Yamamoto, K. (1988) Biochem. J. 254, 895–898.
[9] Scarborough, P.E., Guruprasad, K., Topham, C., Richo, G.R., Conner, G.E., Blundell, T.L. and Dunn, B.M. (1993) Protein Sci. 2, 264–276.

[10] Scarborough, P.E. and Dunn, B.M. (1994) Protein Engng. 7, 495–502.
[11] Sutcliffe, M.J., Haneef, I., Carney, D. and Blundell, T.L. (1987) Protein Eng. 1, 377–384.
[12] Sutcliffe, M.J., Hayes, F.R.F. and Blundell, T.L. (1987) Protein Eng. 1, 385–392.
[13] Blundell, T.L., Carney, D., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J., Singh, D.A., Sibanda, B.L. and Sutcliffe, M. (1988) Eur. J. Biochem. 172, 513–520.
[14] Săli, A., Overington, J.P., Johnson, M.S. and Blundell, T.L. (1990) Trends Biochem. Sci. 15, 235–240.
[15] Cooper, J.B., Khan, G., Taylor, G., Tickle, I.J. and Blundell, T.L. (1990) J. Mol. Biol. 214, 199–222.
[16] Newman, M. Safro, M., Frazao, C., Khan, G., Zdanov, A., Tickle, I.J., Blundell, T.L. and Andreeva, N. (1991) J. Mol. Biol. 221, 1295–1309.
[17] Dhanaraj, V., Dealwis, C.G., Frazao, C., Badasso, M., Sibanda, B.L., Tickle, I.J., Cooper, J.B., Driessen, H.P.C., Newman, M., Aguilar, C., Wood, S.P., Blundell, T.L., Hobart, P.M., Geoghegan, K.F., Ammirati, M.J., Danley, D.E., O'Connor, B.A. and Hoover, D.J. (1992) Nature 357, 466–472.
[18] Dealwis C.G., Frazao, C., Badasso, M., Cooper, J.B., Tickle, I.J., Driessen, H., Blundell, T.L., Murakami, K., Miyazaki, H., Sueiras-Diaz, J., Jones, D.M. and Szelke, M. (1994) J. Mol. Biol. 236, 342–360.
[19] Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C., Cachau, R.E., Collins, J., Silva, A.M. and Erickson, J.W. (1993) Proc. Natl. Acad. Sci. USA 90, 6796–6800.
[20] Aguilar, C., Badasso, M., Dreyer, T., Hoover, D.J., Newman, M., Cooper, J.B., Wood, S.P. and Blundell, T.L. (1994) (submitted).
[21] Johnson, M.S. and Overington, J.P. (1993) J. Mol. Biol. 233, 716 738.
[22] Srinivasan, N. and Blundell, T.L. (1993) Protein Eng. 6, 501–512.
[23] Overington, J.P., Johnson, M.S., Sali, A. and Blundell, T.L. (1990) Proc. R. Soc. Lond. Ser. B 241, 132–145.
[24] Overington, J., Donnelly, D., Johnson, M.S., Sali, A. and Blundell, T.L. (1992) Protein Sci. 1, 216–226.
[25] Topham, C.M., McLeod, A., Eisenmenger, F., Overington, J.P., Johnson, M.S. and Blundell, T.L. (1993) J. Mol. Biol. 229, 194–220.
[26] Hermans Jr., J. and McQueen Jr., J.E. (1974) Acta Crystallogr. 30, 730–739.
[27] Jones, T.A. (1978) J. Appl. Crystallogr. 11, 268–272.
[28] Cooper, J.B., Foundling, S., Hemmings, A., Blundell, T.L., Jones, D.M., Hallet, A. and Szelke, M. (1987) Eur. J. Biochem. 169, 215–221.
[29] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Jr., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[30] Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J. (1987) in: Crystallographic Databases. Information Content, Software Systems, Scientific Applications (Allen, F.H., Bergerhoff, G. and Sievers, R., eds.) pp. 107–132.
[31] Chetana, R.N., Guruprasad, K., Batley, B., Rapundalo, S., Taylor, M.D., Blundell, T.L., Kay, J. and Dunn, B.M. (1994) (submitted).
[32] Johnson, M.S., Overington, J.P. and Blundell, T.L. (1993) J. Mol. Biol. 231, 735–752.
[33] Overington, J.P., Zhu, Z.Y., Sali, A., Johnson, M.S., Sowdhamini, R., Louie, G.V. and Blundell, T.L. (1993) Biochem. Soc. Trans. 21, 597–604.
[34] Wako, H. and Blundell, T.L. (1994) J. Mol. Biol. 238, 682–692.
[35] Wako, H. and Blundell, T.L. (1994) J. Mol. Biol. 238, 693–708.
[36] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[37] Rost, B. and Sander, C. (1993) Proc. Natl. Acad. Sci. USA 90, 7558–7562.
[38] Garnier, J., Osguthorpe, D.J., Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[39] Hashimoto, H., Nishi, R., Uchimiya, H. and Kato, A. (1992) Accession number D12777, submitted (1 August 1992) to the DNA database of Japan.
[40] Brodelius, P.E., Cordeiro, M. and Pais, M.S. (1993) Abstracts L14 and P30 from the proceedings of the 5th International Conference on Aspartic Proteinases held in GIFU, JAPAN, September 19–24.
[41] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Cryst. 26, 283–291.
[42] O'Brien and Kishimoto (1991) FASEB J. 5, 301–308.
[43] Zhu, Y. and Conner G.E. (1994) J. Biol. Chem. 269, 3846–3851.
[44] Rijnboutt, S., Aerts, H.M.F.G., Geuze, H.J., Tager, J.M. and Strous G.J. (1991) J. Biol. Chem. 266, 4862–4868.
[45] Evans, S.E. (1993) J. Mol. Graphics 11, 134–138.