

# Probing the Modelled Structure of Wheatwin1 by Controlled Proteolysis and Sequence Analysis of Unfractionated Digestion Mixtures

Carlo Caporale,<sup>1\*</sup> Carla Caruso,<sup>1</sup> Angelo Facchiano,<sup>2</sup> Monica Nobile,<sup>1</sup> Luca Leonardi,<sup>1</sup> Laura Bertini,<sup>1</sup> Giovanni Colonna,<sup>2</sup> and Vincenzo Buonocore<sup>1</sup>

<sup>1</sup>Dipartimento di Agrobiologia ed Agrochimica, Università della Tuscia, Viterbo, Italy

<sup>2</sup>CRISCEB-Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche, II Università di Napoli, Napoli, Italy

**ABSTRACT** We set up a method to get rapid information on the three-dimensional structure of peptide and proteins of known sequence. Both native and alkylated polypeptide is hydrolyzed with a number of proteases at different digestion times and the resulting mixtures are compared by HPLC analysis to establish the differences in the hydrolysis pathways of the folded and unfolded molecule. Then, the unfractionated digestion mixtures of the native polypeptide are submitted to automatic sequence analysis to identify the hydrolysis sites. The sequence of each fragment present in the mixtures is reconstructed and its amount determined by quantitative data of the sequence analyses. We used this approach to determine the amino acid surface accessibility of wheat-win1, a pathogenesis-related protein from wheat, and constructed a predictive three-dimensional model based on the knowledge of the tertiary structure of barwin, a highly homologous protein from barley. The procedure allowed us to quickly identify and quantify the hydrolysis at the susceptible bonds which could be classified as exposed, partially hidden, or inaccessible. The results were useful to evidentiate and discuss concordances and differences between experimental and model predicted accessibilities of amino acid residues. *Proteins* 1999;36:192–204.

© 1999 Wiley-Liss, Inc.

**Key words:** molecular modelling; peptide mixtures; protein structure; sequence analysis; surface accessibility

## INTRODUCTION

The knowledge of the tertiary structure of proteins in solution is a fundamental goal in establishing their structure-function relationship. In the last years, NMR techniques have been used in tackling this problem [Ludvigsen and Poulsen,<sup>1</sup> Lecompt,<sup>2</sup> and Wutrich,<sup>3</sup> and references therein]. This approach provides spatial atom coordinates which can be used by simulation procedures to determine the three-dimensional structure of homologous proteins.<sup>4–7</sup> Furthermore, useful information on the global folding can be obtained by investigating the amino acid accessibility. This can be done by methods labeling the exposed residues

such as isotopic exchange reactions<sup>8,9</sup> and chemical modifications<sup>10–12</sup> or identifying them by enzymatic hydrolysis of their peptidic bonds.<sup>13–25</sup> In general, the use of proteolytic enzymes has been shown to be an excellent tool in studying protein folding.<sup>26,27</sup> Using this approach, peptidic fragments are usually purified and successively identified. The identification can be accomplished by classical chemistry methodologies<sup>28–32</sup> or by mass spectrometry analysis.<sup>8–11,33–42</sup> In any case, the fragment quantification represents a problem since the yield of purification steps could be different for each product. On the other hand, mass spectrometry analysis, although being a rapid and sensitive method which can be directly performed on digestion mixtures, does not produce data which can be easily quantified. In this view, a fast and sensitive procedure not requiring purification steps and providing quantifiable data should be very useful when fragment isolation is not required for studying particular domains.

In the last years, the automatic sequence analysis of peptide mixtures by modern sequencers equipped on-line with phenylthiohydantoin (PTH) amino acid analyzers has been usefully used to characterize proteins. Different strategies in assessing protein sequences have been reported<sup>43–49</sup> together with methods for determining the hydrolysis pathway of polypeptides by unknown digesting agents.<sup>50–53</sup> These approaches, which have been recently reviewed,<sup>54</sup> do not require any purification step and are based upon the sequence determination of the fragments in mixture; moreover, sequence data being quantitative, the amount of each fragment can be evaluated.<sup>43–54</sup> The method can also be used to get rapid information on the tertiary structure of proteins in solution. In fact, if a native protein of known sequence is digested by a specific protease, it is possible to follow the kinetics of the hydrolysis at the recognized sites by quantitative sequence analysis of the mixture products at different digestion times. In this way, each expected hydrolysis site can be classified as exposed, partially hidden,

Grant sponsor: Ministero per le Politiche Agricole (MIPA, former MIRAAF), Piano Nazionale Biotecnologie Vegetali, Progr. N. 152.

This paper is dedicated to the memory of professor Giacomino Randazzo.

\*Corresponding author: Carlo Caporale, Dipartimento di Agrobiologia ed Agrochimica, Università della Tuscia, Via S. Camillo de Lellis, 01100 Viterbo, Italy. E-mail: caporale@unitus.it

Received 8 December 1998; Accepted 8 March 1999

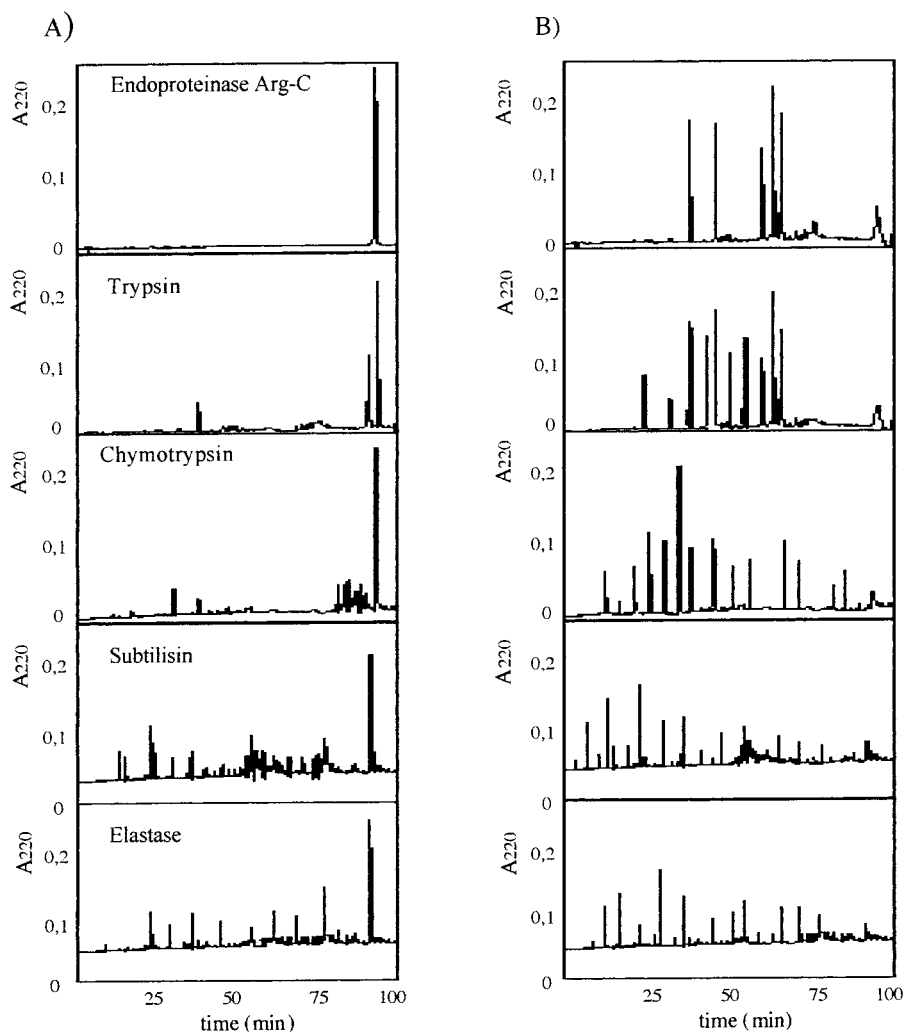


Fig. 1. HPLC analyses of native (A) and alkylated (B) wheatwin1 digestion mixtures with endoproteinase Arg-C, trypsin, chymotrypsin, subtilisin, and elastase.

or inaccessible. Furthermore, if aspecific proteases are utilized, a general information about all the residues present on the protein surface is obtained. A complete mapping can be accomplished since proteases with different specificity can be utilized. This information can be useful to validate predictive structural models based on the knowledge of the tertiary structure of homologous proteins.

We used this approach to investigate the three-dimensional structure in solution of wheatwin1, a pathogenesis-related (PR) protein from wheat, whose sequence is known.<sup>55</sup> Furthermore, a predictive structural model based on the knowledge of the tertiary structure in solution of barwin,<sup>1,56</sup> a highly homologous protein from barley,<sup>57</sup> was constructed. The experimental information on the amino acid accessibility obtained by sequence analysis of the unfractionated digestion mixtures of the native protein was used to verify the concordance with the model structure.

## MATERIALS AND METHODS

### Materials

*Triticum aestivum*, pure variety San Pastore, was kindly supplied from the Istituto Nazionale per la Cerealicoltura

(S. Angelo Lodigiani, Italy). Pulsed liquid-phase automatic sequencer (model 477A) equipped on-line with PTH-amino acid analyzer (model 120A) and relative reagents were from Perkin Elmer-Applied Biosystems, Italia, SPA. Reverse-phase HPLC (RP-HPLC) procedures were carried out on a Beckman GOLD apparatus equipped with a variable-wavelength monitor model 166. Eluent A was aqueous 0.1% trifluoroacetic acid and eluent B was 0.07% trifluoroacetic acid in acetonitrile. The  $\mu$ -Bondapak C<sub>18</sub> column (0.39 × 30 cm) was from Waters-Millipore (Milford, MA). Endoproteinase Arg-C, trypsin, chymotrypsin, subtilisin, and elastase were from Boehringer Mannheim Italia SpA. All other reagents were of analytical grade.

### Protein Purification, Reduction, and Alkylation With 4-Vinylpyridine

Wheatwin1 was purified, reduced, and pyridylethylated as previously described.<sup>58,59</sup> The protein concentration was determined by the Bio-Rad assay, following the manufacturer's instructions and using bovine serum albumin as a standard.

## A) Wheatwin1 - wheat

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  
 Q Q A **T** **N** V R A T Y H Y Y R P A Q N N W D L G A P A V  
 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54  
 S A Y C A T W D A S K P L S W R S **G** Y G W T A F C G P  
 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81  
 A G **A** **H** G Q A **S** C G K C L **Q** V T N P A T G A Q I T A R  
 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108  
 I V D Q C A N G G L D L D W D T V F T K I D T N G I G  
 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125  
 Y Q Q G H L N V N Y Q F V D C R D

## B) Barwin - barley

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  
 Q Q A **N** **D** V R A T Y H Y Y R P A Q N N W D L G A P A V  
 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54  
 S A Y C A T W D A S K P L S W R S **K** Y G W T A F C G P  
 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81  
 A G **P** **R** G Q A **A** C G K C L **R** V T N P A T G A Q I T A R  
 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108  
 I V D Q C A N G G L D L D W D T V F T K I D T N G I G  
 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125  
 Y Q Q G H L N V N Y Q F V D C R D

Fig. 2. Single-literal code amino acid sequence of wheatwin1 (A) and barwin (B). The seven different residues are shown in bold-face.

### Protease Digestions and HPLC Analyses

Proteolysis experiments were accomplished by incubating both native and alkylated wheatwin1 (15 nmol of each) separately with endoproteinase Arg-C, trypsin, chymotrypsin, elastase, and subtilisin. Digestions were performed in 0.6 ml of 0.5% ammonium bicarbonate, pH 7.8 at 37°C. Saline co-factors were added to the incubation mixtures when needed, according to the manufacturer's instructions. The following enzyme/substrate ratio (w/w), necessary for detecting native protein hydrolysis at 1h incubation, were utilized for each enzyme: endoproteinase Arg-C 1:50, trypsin 1:50, chymotrypsin 1:50, elastase 1:25, subtilisin 1:50. Aliquots of the incubation mixtures, corresponding to 2 nmol of the original protein, were withdrawn at 0, 20, 60, 120, 240, and 720 minutes, freeze-dried, dissolved in water (0.2 ml) and lyophilized twice. The samples were then dissolved in 0.1 ml of aqueous 0.1% trifluoroacetic acid. Aliquots of 0.05 ml were submitted to HPLC analysis using a  $\mu$ -Bondapak C<sub>18</sub> column. Eluent A was aqueous 0.1% trifluoroacetic acid and eluent B was 0.07% trifluoroacetic acid in acetonitrile. The elution was performed at a flow rate of 1 ml/min by using a linear gradient from 10% to 40% B in 110 min.

### Sequence Analyses

Aliquots of native wheatwin1 digestion mixtures corresponding to 1 nmol of the original protein were submitted to automatic sequence analysis using a pulsed liquid-

phase automatic sequencer equipped on-line with PTH-amino acid derivatives analyzer. Samples were loaded onto a trifluoroacetic-acid-treated glass-fiber filter, coated with polybrene and washed according to the manufacturer's instructions. Analysis data were processed by both the instrument software and the program "HYDROSITES".<sup>45</sup> The theoretical initial yield of the coupling reaction was 65% using  $\beta$ -lactoglobulin as standard. The average and combined repetitive amino acids yields were not lower than 90%.

### Protein Modelling

Structure prediction of wheatwin1 was based on the availability of the three-dimensional model of the homologous protein barwin<sup>1,56</sup> (pdb code: 1bw3). The alignment of wheatwin1 and barwin did not require deletion or insertion of gap. The program MODELLER<sup>60</sup> and Quanta (Molecular Simulations Inc., San Diego, CA) were used to obtain 20 full-atom models of wheatwin1 according to the comparative protein modelling method. Stereo chemical quality of models was verified with the program PROCHECK.<sup>61</sup> Secondary structure was assigned by the program DSSP.<sup>62</sup> Search for structural classification of barwin was performed on SCOP<sup>63</sup> and CATH<sup>64</sup> databases. Solvent accessibility of amino acids was evaluated by the program NACCESS<sup>65</sup> calculating the atomic accessible surface defined by rolling a probe of 1.40 Angstrom around the van der Waals surface of the protein model. Mean solvent accessibility of residues in the 20 models was calculated. Figures

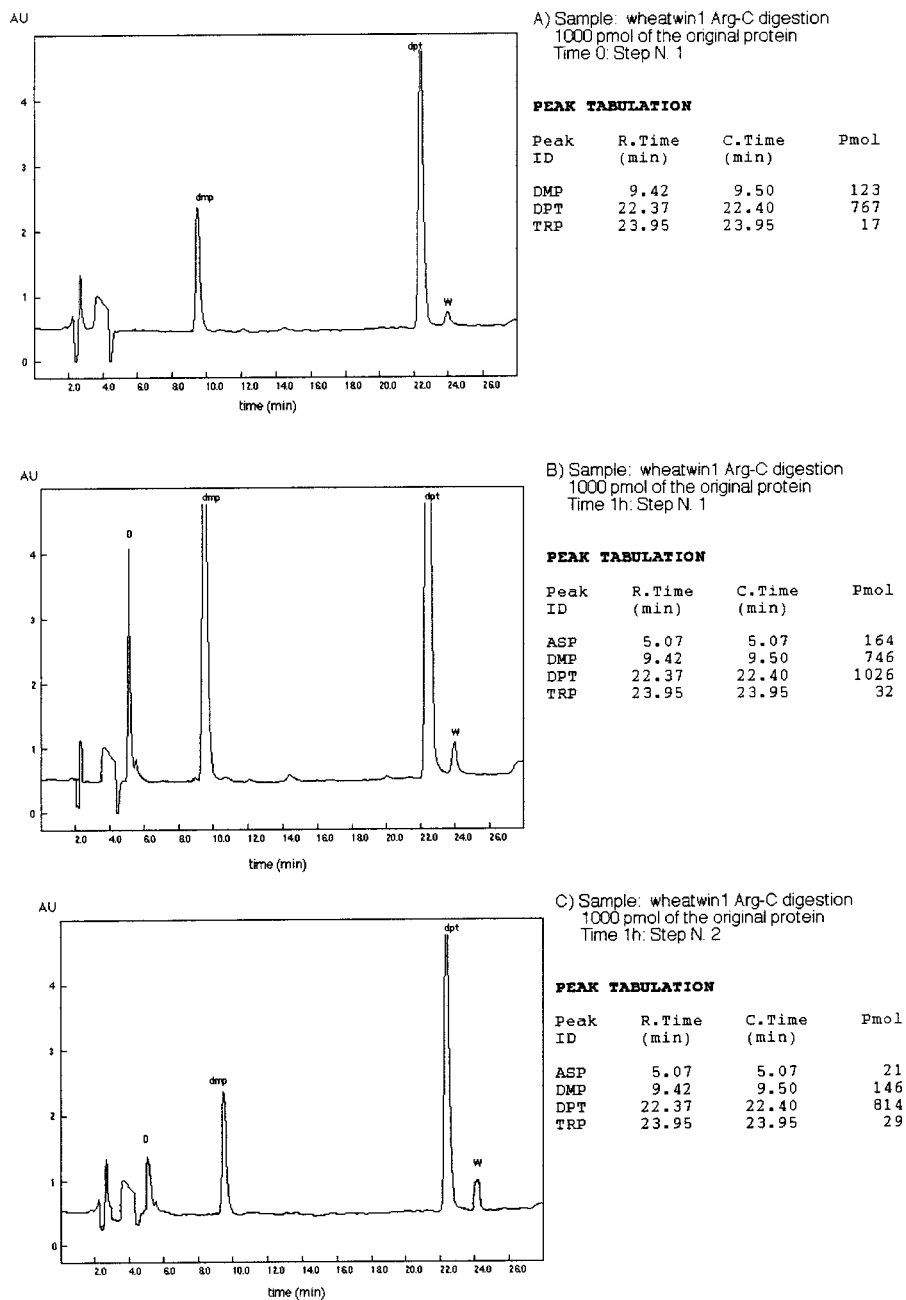


Fig. 3. Sequence analyses of wheatwin1 digestion mixtures with endoproteinase Arg-C; (A) 0 minutes incubation, first degradation step; (B) 1h incubation, first degradation step; (C) 1h incubation, second degradation step.

were drawn with InsightII package (Molecular Simulations, Inc.).

## RESULTS AND DISCUSSION

### Protein Digestions, HPLC, and Sequence Analyses

Wheatwin1 is a monomeric protein of 125 residues containing six cysteines all linked in disulfide bridges.<sup>55</sup> Both native and pyridylethylated protein was separately digested with different proteases. Portions of the digestion mixtures were withdrawn at different times and submitted to HPLC analysis to compare their composition. The results obtained from mixtures at 1 h incubation are reported in Fig. 1. The production of the expected frag-

ments of the alkylated protein was observed with both specific and unspecific proteases, showing a quite complete hydrolysis with each enzyme (Fig. 1B). In comparison, the native protein was very partially digested (Fig. 1A), indicating that such limited hydrolyses were due to its 3D structure. Then, each digestion mixture of the native protein was submitted to automatic sequence analysis to identify the produced fragments. The main difficulties in the interpretation of the sequence analysis of peptide mixtures are due to the different yield of PTH-amino acids and to the carryover of the Edman degradation. The yield of the reaction lowers at progressive steps of the sequence analysis and is not the same for each PTH-amino acid,

**TABLE I. Identification of Fragments in the Digestion Mixtures of Native Wheatwin1<sup>†</sup>**

Enzyme	Hydrolysis site	First residue useful for the identification	Edman step	pmol 0 min	pmol 20 min	pmol 60 min	pmol 120 min	pmol 240 min	pmol 720 min
Arg-C	R <sub>124</sub>	D <sub>125</sub>	1	1	21	164	183	186	188
Trypsin	R <sub>7</sub>	A <sub>8</sub>	1	2	7	66	71	80	87
	R <sub>124</sub>	D <sub>125</sub>	1	1	74	287	321	326	329
Chymotrypsin	A <sub>3</sub>	T <sub>4</sub>	1	0	4	52	55	58	61
	Y <sub>10</sub>	H <sub>11</sub>	1	1	6	56	63	65	66
	W <sub>20</sub>	D <sub>21</sub>	1	1	5	41	49	51	55
Subtilisin	T <sub>9</sub>	H <sub>11</sub>	2	0	16	108	114	117	117
	Y <sub>13</sub>	R <sub>14</sub>	1	1	19	137	142	146	149
	A <sub>16</sub>	Q <sub>17</sub>	1	1	13	119	123	127	129
	N <sub>88</sub>	G <sub>90</sub>	2	1	29	128	134	139	144
	V <sub>98</sub>	F <sub>99</sub>	1	1	6	47	51	54	58
	N <sub>105</sub>	I <sub>107</sub>	2	1	22	158	164	166	170
	N <sub>117</sub>	Q <sub>119</sub>	2	1	7	56	61	65	70
Elastase	V <sub>121</sub>	D <sub>122</sub>	1	2	8	93	99	104	107
	T <sub>9</sub>	Y <sub>10</sub>	1	1	23	108	118	121	124
	A <sub>16</sub>	Q <sub>17</sub>	1	1	18	153	160	167	171
	L <sub>40</sub>	S <sub>41</sub>	1	0	15	168	176	181	185
	S <sub>44</sub>	G <sub>45</sub>	1	2	11	92	97	103	107
	T <sub>79</sub>	A <sub>80</sub>	1	1	5	63	66	71	74
	V <sub>98</sub>	F <sub>99</sub>	1	1	4	49	53	55	58
	T <sub>100</sub>	K <sub>101</sub>	1	0	22	127	134	138	142

<sup>†</sup>Sequence pmol data of the first residue useful to identify the fragments produced in the digestion mixtures of native wheatwin1 at various incubation time with the utilized proteases.

owing to partial destruction (e.g., serine and threonine), slow conversion (e.g., proline and glycine), reduced sensitivity in HPLC detection of positively charged molecules eluted as broad peaks (e.g., arginine and histidine) and low solubilization (e.g., lysine). These difficulties have been greatly reduced using modern pulsed-liquid phase sequencers, since the carryover is a well quantifiable phenomenon and, despite the yield problems, the identification of several amino acid residues at each step requires just a little care and experience.<sup>43–54</sup> Thus, digesting a protein of known sequence by any agent, it is possible to assess the sequence of each fragment present in the mixture and evaluate its amount. Although the absolute quantification is underestimated due to the initial coupling yield of Edman degradation (60–65%), the data can be compared and supply quantitative information since they suffer for a similar error. In the case of wheatwin1, the interpretation of the sequence data of fragments generated by protease digestions was further simplified since the N-terminal sequence of the protein was not detected by Edman degradation, the first residue being blocked (pyroglutamate).<sup>55</sup>

### Digestion With Endoproteinase Arg-C

The single-letter code sequence of the protein is shown in Fig. 2A. Five arginine sites located at positions 7, 14, 43, 81, and 124 could be exposed to the action of endoproteinase Arg-C. As a consequence, the presence of the fragments 1–7 (blocked at the N-terminus), 8–14, 15–43, 44–81, 82–124 and of free aspartic acid due to the hydrolysis of the peptide bond R<sub>124</sub>-D<sub>125</sub> is expected in the digestion mix-

tures. While the HPLC analysis showed a quite complete hydrolysis of the alkylated protein at 1 h incubation (Fig. 1B), the native protein seemed to be completely undigested (Fig. 1A). On the basis of this result, it could appear that no hydrolysis had occurred. In contrast, the sequence analyses of the native mixtures at 0 min and 1 h incubation were different (Fig. 3). In fact, the only peaks recognized at the first step of the analysis of the mixture at 0 min incubation are common to all sequence chromatograms, identifying N,N-dimethyl-N'-phenylthiourea (DMPTU), N,N'-diphenylthiourea (DPTU) and a minor by-product of the Edman degradation having the same retention time of PTH-tryptophan. No amino acid residue was identified (Fig. 3A). On the contrary, 164 pmol of aspartic acid were detected at the first step of the sequence analysis of the mixture at 1 h incubation (Fig. 3B). As can be observed looking at Figure 3C, no further residue was identified at the second step, the detected peak being due to the carry-over of aspartic acid present at the first step. The presence of free aspartic acid at the first step could be due only to the hydrolysis of the peptide bond R<sub>124</sub>-D<sub>125</sub>. The quantitative datum (164 pmol) shows that only a part of the original protein (1,000 pmol) was hydrolyzed. It should be remarked that neither the free aspartic acid, nor the protein lacking of the C-terminal residue were detected by HPLC analysis (Fig. 1A). This demonstrates the ability and sensitivity of the sequence analysis performed directly on the digestion mixture, since the isolation of the produced fragments should be very laborious in this case. Table I shows pmol data of the first residue useful to identify the fragments present in the digestion mixtures produced by each enzyme at the different incubation time.

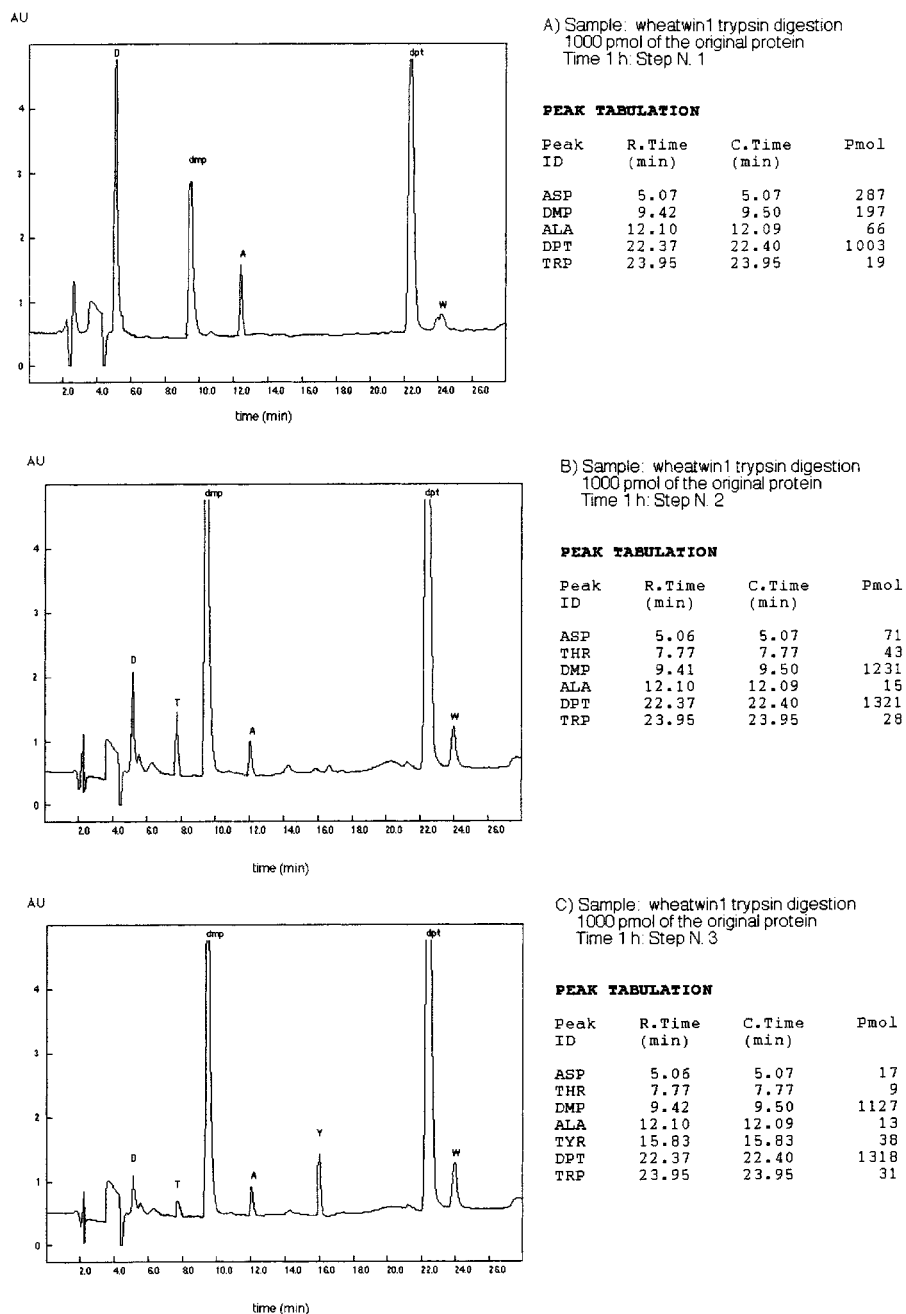


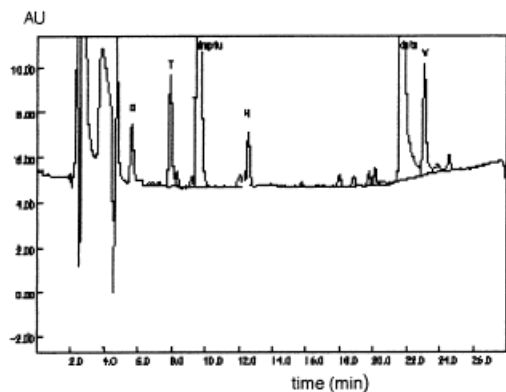
Fig. 4. Sequence analysis of wheatwin1 digestion mixture at 1h incubation with trypsin; first (A), second (B), and third (C) degradation steps.

Endoproteinase Arg-C produced the major amount of  $D_{125}$  during the first incubation hour, while no further fragment was detected up to 12 h incubation. These results indicate that  $R_{124}$  is the most exposed arginine residue, while the enzyme is not able to nick the native protein at any of the other sites.

### Digestion With Trypsin

Endoproteinase Arg-C was able to hydrolyze only the peptide bond involving  $R_{124}$  of the folded protein. In order to verify if this poor digestion was really due to the

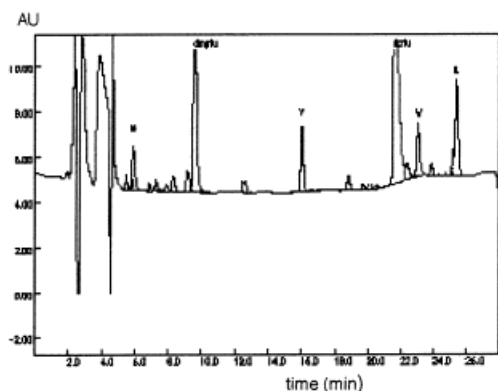
inaccessibility of the remaining sites, wheatwin1 was digested with trypsin. This allowed us to get information also on the lysine residues, since the tryptic digestion of the reduced and alkylated protein produced cleavages at the expected sites (Fig. 1B). Figure 4 shows three of the ten steps of the sequence analysis performed on the native digestion mixture at 1h incubation. The presence of free aspartic acid at the first step (287 pmol) confirmed the hydrolysis at level of  $R_{124}$ . In addition, the hydrolysis at level of  $R_7$  was evident. In fact, the sequence Ala (66 pmol), T (43 pmol), Y (38 pmol) corresponds to the sequence 8–10 of the protein (Fig. 2A). The next steps of the analysis



A) Sample: wheatwin 1  
chymotrypsin digestion  
1000 pmol of the original protein  
Time 1 h : Step N. 1

**PEAK TABULATION**

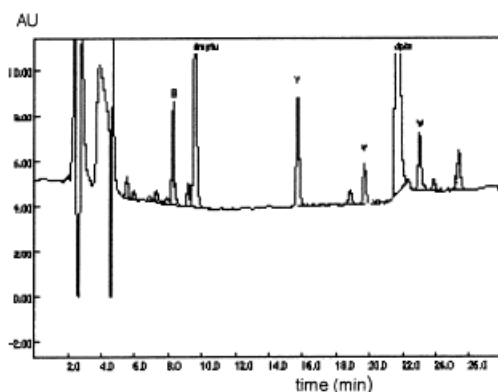
Peak	R. Time (min)	C. Time (min)	Pmol
ASP	5.07	5.07	41
THR	7.76	7.77	52
DMP	9.43	9.50	997
HIS	12.56	12.58	56
DPT	22.35	22.40	1121
TRP	23.95	23.95	42



B) Sample: wheatwin 1  
chymotrypsin digestion  
1000 pmol of the original protein  
Time 1 h : Step N. 2

**PEAK TABULATION**

Peak	R. Time (min)	C. Time (min)	Pmol
ASN	5.82	5.90	40
DMP	9.43	9.50	877
TYR	15.87	15.83	51
DPT	22.37	22.40	976
TRP	23.94	23.96	21
LEU	25.32	25.38	48



C) Sample: wheatwin 1  
chymotrypsin digestion  
1000 pmol of the original protein  
Time 1 h : Step N. 3

**PEAK TABULATION**

Peak	R. Time (min)	C. Time (min)	Pmol
GLY	8.27	8.31	42
DMP	9.43	9.60	989
TYR	15.86	15.83	69
VAL	20.04	20.04	24
DPT	22.37	22.40	988
TRP	23.95	23.95	23

Fig. 5. Sequence analysis of wheatwin1 digestion mixture at 1h incubation with chymotrypsin; first (A), second (B), and third (C) degradation steps.

allowed us to unambiguously identify the sequence up to Y<sub>13</sub> (not shown). The quantitative datum of Ala at the first step (66 pmol) indicates that the hydrolysis at level of R<sub>7</sub> occurred at a minor extent with respect to the hydrolysis at level of R<sub>124</sub>, which was confirmed to be the most exposed site. Moreover, trypsin was more effective than endoproteinase Arg-C. In this case also, the sequence of no further fragment was detected up to 12 h incubation, showing that the remaining arginine and lysine were not accessible in the native conformation of the protein (Table I).

### Digestion With Chymotrypsin

In order to get a more general information about the residues present on the surface, wheatwin1 was digested with chymotrypsin. This enzyme is able to cleave peptide

bonds involving at the C-side aromatic residues and, at a lower rate, Leu, Met, Ala, Asp, and Glu. It hydrolyzed the reduced and alkylated protein at a great number of sites (Fig. 1B). Figure 5 shows three of the ten steps of the sequence analysis performed on the digestion mixture at 1h incubation. Asp, Thr, and His were identified at the first step (Fig. 5A), Asn, Tyr, and Leu at the second one (Fig. 5B), while Gly, Tyr, and Val were present at the third step (Fig. 5C). The sequences T<sub>4</sub>-N<sub>5</sub>-V<sub>6</sub>, H<sub>11</sub>-Y<sub>12</sub>-Y<sub>13</sub>, and D<sub>21</sub>-L<sub>22</sub>-G<sub>23</sub> can be reconstructed by these residues (Fig. 1A), showing that cleavages at level of A<sub>3</sub>, Y<sub>10</sub>, and W<sub>20</sub> occurred. As can be deduced by the quantitative data in Figure 5, these hydrolyses arose to an extent ranging between 4–6% of the total molecules and no further significant amounts of the fragments were produced up to 12 h incubation (Table I).

## A)

Edman steps	Residues (pmoles)																			
	D	N	S	Q	T	G	E	A	H	Y	R	P	M	V	C	W	F	I	K	L
1	<b>93</b>	3	2	<b>119</b>	2	<b>281</b>	11	4	1	<b>183</b>	<b>137</b>	2	---	2	---	14	<b>47</b>	3	2	4
2	19	<b>195</b>	---	<b>56</b>	<b>43</b>	<b>208</b>	9	3	<b>108</b>	17	16	<b>107</b>	---	1	---	15	9	<b>158</b>	1	3
3	15	<b>86</b>	---	10	9	<b>128</b>	7	<b>83</b>	15	<b>89</b>	<b>64</b>	16	---	---	---	15	<b>39</b>	17	<b>35</b>	<b>173</b>
4	<b>170</b>	11	1	8	5	19	5	9	9	<b>138</b>	9	11	---	<b>32</b>	---	<b>69</b>	7	<b>29</b>	8	19
5	<b>98</b>	9	1	<b>92</b>	4	11	5	6	6	15	7	10	---	8	---	21	6	8	7	<b>103</b>
6	<b>61</b>	8	1	<b>88</b>	17	9	5	6	3	11	5	9	---	6	---	17	5	5	6	<b>44</b>
7	11	<b>15</b>	1	13	6	<b>66</b>	4	5	3	8	3	8	---	5	---	<b>55</b>	5	4	5	8
8	<b>63</b>	7	---	11	5	7	4	5	<b>39</b>	6	3	8	---	4	---	18	4	3	5	7
9	14	6	---	9	<b>48</b>	6	3	4	7	6	2	7	---	4	---	14	4	3	4	<b>28</b>
10	10	<b>29</b>	---	7	10	6	3	4	5	6	2	6	---	<b>38</b>	---	14	3	2	3	6

## B)

Residues identified at each step									
1	2	3	4	5	6	7	8	9	10
Y (183)	H (108)	Y (89)	Y (138)	D (98)	L (44)	G (66)	H (39)	L (28)	N (29)
R (137)	P (107)	A (83)	W (69)	Q (92)	Q (88)	W (55)	D (63)	T (48)	V (38)
Q (119)	N (195)	N (86)	V (32)	L (103)	D (61)	N (15)			
D (93)	Q (56)	R (64)	D (170)		T (17)				
G (281)	G (128)	L (173)	I (29)						
F (47)	T (43)	K (35)							
	I (158)	G (128)							
		F (39)							

## C)

Identified sequences	Trait	Site	Hydrolysis (%)						
				1	2	3	4	5	6
a) <b>Y-H-Y-Y</b>	10-13	T <sub>9</sub>	11						
b) <b>R-P-A</b>	14-16	Y <sub>13</sub>	14						
c) <b>Q-N-N-W-D-L</b>	17-22	A <sub>16</sub>	15						
d) <b>G-G-L-D-L-D-W-D-T-V</b>	89-98	N <sub>88</sub>	21						
e) <b>F-T-K-I-D-T-N</b>	99-105	V <sub>98</sub>	5						
f) <b>G-I-G-Y-Q-Q-G-H-L-N</b>	106-115	N <sub>105</sub>	16						
g) <b>Y-Q-F-V</b>	118-121	N <sub>117</sub>	6						
h) <b>D-X-R-D</b>	122-125	V <sub>121</sub>	9						

Fig. 6. Sequence analysis of wheatwin1 digestion mixture with subtilisin at 1h incubation; **A)** Pmol data of all PTH-amino acids obtained at each step of degradation; the data of the identified residues are in bold-face. **B)** Residues identified at each step and their pmol data. **C)** Sequences of the fragments present in the mixture and identification of the hydrolysis sites.

### Digestion With Subtilisin

The digestion of wheatwin1 with subtilisin, a protease with a broad specificity, yielded a more complex hydrolysis pathway for both native and alkylated protein (Fig. 1). Figure 6 shows the information achieved from the sequence analysis of the native mixture at 1h incubation. Pmol data of all PTH amino acids obtained at each of ten degradation steps are shown in Fig. 6A. The data in bold-face identify the residues present at each step, which are listed in Fig. 6B. The sequences of the corresponding fragments present in the mixture, deduced from the sequence of the intact protein (Fig. 2A), are shown in Figure 6C. Cysteine 123 is indicated by the symbol "X" since it was not detectable in the sequence analysis of the native protein. All peptides are unambiguously identified by the residues shown in bold-face which can belong only to the sequence of a single fragment. The corresponding pmol data are useful to determine the amounts in the mixture. For example, the first residue useful to identify and quantify the peptide 10-13 is histidine (H, 108 pmol) at the second step, since tyrosine (Y, 183 pmol) at the first step is common to both the fragments 10-13 and 118-121

(Fig. 6C, position 1). In this way, eight cleavage sites were distinguished and the relative hydrolyses quantified with respect to the amount of the original protein (Fig. 6C). The hydrolyses arose to an extent ranging between 5% (Valine 98) and 21% (Asparagine 88). Also in this case, the major hydrolyses occurred during the first incubation hour and no further fragment was produced up to 12 h (Table I).

### Digestion With Elastase

Further information was obtained digesting the protein with elastase, which preferentially cleaves peptide bonds involving at the C-side uncharged nonaromatic residues. The alkylated protein was extensively digested also in this case (Fig. 1B), while the enzyme produced limited hydrolysis of the native protein at 1 h incubation (Fig. 1A). Pmol data of all PTH amino acids of the native mixture sequence analysis are shown in Figure 7A. The data in bold-face identify the residues present at each step (Fig. 7B), allowing the characterization of the fragments in the mixture and the identification of the hydrolysis sites (Fig. 7C). Like subtilisin, elastase nicked the protein at level of T<sub>9</sub>, A<sub>16</sub>, and V<sub>98</sub>. Furthermore, new hydrolyses at level of

## A)

Edman steps	Residues (pmoles)																			
	D	N	S	Q	T	G	E	A	H	Y	R	P	M	V	C	W	F	I	K	L
1	4	2	<b>168</b>	<b>153</b>	1	<b>92</b>	13	<b>63</b>	1	<b>108</b>	4	3	---	2	---	14	<b>49</b>	3	<b>127</b>	3
2	15	<b>121</b>	19	17	<b>32</b>	11	7	11	<b>69</b>	<b>88</b>	51	1	---	2	---	<b>131</b>	14	<b>94</b>	17	2
3	<b>85</b>	<b>98</b>	9	8	<b>8</b>	<b>69</b>	5	7	13	<b>62</b>	<b>103</b>	1	---	2	---	23	7	<b>49</b>	8	2
4	10	17	<b>83</b>	5	<b>72</b>	13	4	5	6	<b>49</b>	12	1	---	<b>38</b>	---	<b>84</b>	5	16	5	2
5	<b>75</b>	<b>66</b>	12	4	<b>44</b>	8	3	5	5	9	<b>34</b>	1	---	7	---	19	4	9	4	1
6	8	11	8	<b>35</b>	10	<b>57</b>	10	<b>38</b>	4	5	7	<b>27</b>	---	5	---	15	4	7	3	<b>39</b>
7	6	9	5	9	6	14	6	<b>21</b>	4	4	4	8	---	4	---	13	<b>31</b>	<b>44</b>	3	8
8	5	7	3	6	5	<b>39</b>	4	<b>15</b>	3	3	2	5	---	3	---	13	9	14	2	4
9	7	<b>19</b>	3	5	3	<b>26</b>	3	7	2	<b>28</b>	2	4	---	3	---	12	6	7	1	2
10	4	8	2	<b>21</b>	3	<b>21</b>	7	3	2	7	2	<b>18</b>	---	2	---	12	5	5	1	2

## B)

Residues identified at each step									
1	2	3	4	5	6	7	8	9	10
Y (108)	H (69)	Y (62)	Y (49)	R (34)	P (27)	A (21)	A (15)	G (26)	Q (21)
Q (153)	N (121)	N (98)	W (84)	D (75)	L (39)	F (31)	G (39)	N (19)	P (18)
A (63)	R (51)	I (49)	V (38)	T (44)	Q (35)	I (44)	Y (28)	G (21)	
S (168)	W (131)	R (103)	S (83)	N (66)	A (38)				
G (92)	Y (88)	G (69)	T (72)	G (57)					
F (49)	T (32)	D (85)							
K (127)	I (94)								

## C)

Identified sequences	Trait	Site	Hydrolysis (%)
1 2 3 4 5 6 7 8 9 10			
a) <b>Y-H-Y-Y-R-P-A</b>	10-16	T <sub>9</sub>	11
b) <b>Q-N-N-W-D-L</b>	17-22	A <sub>16</sub>	15
c) <b>S-W-R-S</b>	41-44	L <sub>40</sub>	17
d) <b>G-Y-G-W-T-A-F-X-G-P</b>	45-54	S <sub>44</sub>	9
e) <b>A-R-I-V-D-Q-X-A-N-G</b>	80-89	T <sub>79</sub>	6
f) <b>F-T</b>	99-100	V <sub>98</sub>	5
g) <b>K-I-D-T-N-G-I-G-Y-Q</b>	101-110	T <sub>100</sub>	12

Fig. 7. Sequence analysis of wheatwin1 digestion mixture with elastase at 1h incubation; **A)** Pmol data of all PTH-amino acids obtained at each step of degradation; the data of the identified residues are in bold-face. **B)** Residues identified at each step and their pmol data. **C)** Sequences of the fragments present in the mixture and identification of the hydrolysis sites.

L<sub>40</sub>, S<sub>44</sub>, T<sub>79</sub>, and T<sub>100</sub>, ranging between 6–17%, occurred. Kinetic data confirmed the results obtained with the other enzymes. No cleavage was complete and the major hydrolysis increments were detected at 1 h incubation (Table I).

### Comparison With the Model

Three-dimensional structure of wheatwin1 was modelled using the barwin 3D structure as reference.<sup>1</sup> Alignment of wheatwin1 and barwin sequences shows only seven amino acid differences, without deletion or insertion (Fig. 2). The program MODELLER<sup>60</sup> was used to generate 20 wheatwin1 modelled structures based on the twenty NMR barwin structures stored in Protein Data Bank. Differences among modelled structures were very subtle, and loop regions very similar in all of them. The DSSP program<sup>62</sup> recognized the same secondary structure for barwin and wheatwin1, so wheatwin1 can be classified as an all-beta protein, similarly to barwin. Its structure consists of a main beta-sheet of four anti-parallel strands, two short parallel beta strands constituting a little independent beta-sheet, and few short helices (Fig. 8).

In Table II, summarizing all the hydrolyses, the accessibility average and standard deviation calculated for each wheatwin1 residue in the 20 models is reported, together

with the hydrolysis percent with the proteases used. The data refer to 1h incubation since the major increment was observed with all the enzymes at this time. Of course, useful comparisons can be correctly made just among data deriving from the same enzyme. Each cleavage datum should be compared with the correspondent datum of solvent accessibility to evidenciate concordance or difference between experimental and model predicted information. On the basis of their specificity, the proteases can recognize 103 amino acid residues out of 125 as potential hydrolysis sites, and were able to nick the native protein at level of 17 of them, while each enzyme fully digested the alkylated protein. Major cuts of native protein occurred at well exposed sites, while a lower hydrolysis percent was measured at less accessible residues, showing a good agreement between experimental and model predicted data (Table II and Fig. 9). In any event, no hydrolysis was observed at the level of the other model predicted exposed sites, indicating that the surface accessibility is not the only parameter to be considered. In fact, in vitro hydrolyses occur only when structural requirements are satisfied: i.e., sites must be exposed on the protein surface and must adopt a conformation suited to interact with the catalytic site on the surface of the proteinase. Conformational

Fig. 8. Secondary structure model of wheatwin1. The green ribbon reproduces the backbone trace. Yellow arrows indicate beta strands, red cylinders represent helices. The location of Thr<sub>9</sub>, Thr<sub>79</sub> and Thr<sub>70</sub> within beta sheet is evidenced by spacefill representation. N- and C- termini are also indicated.

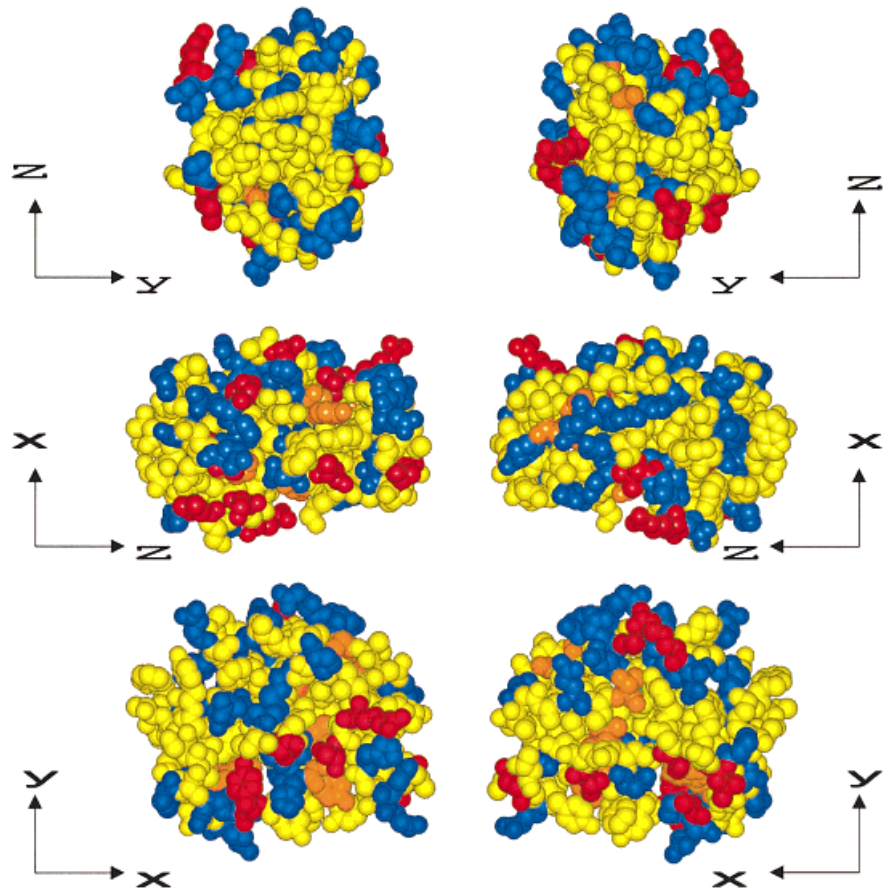
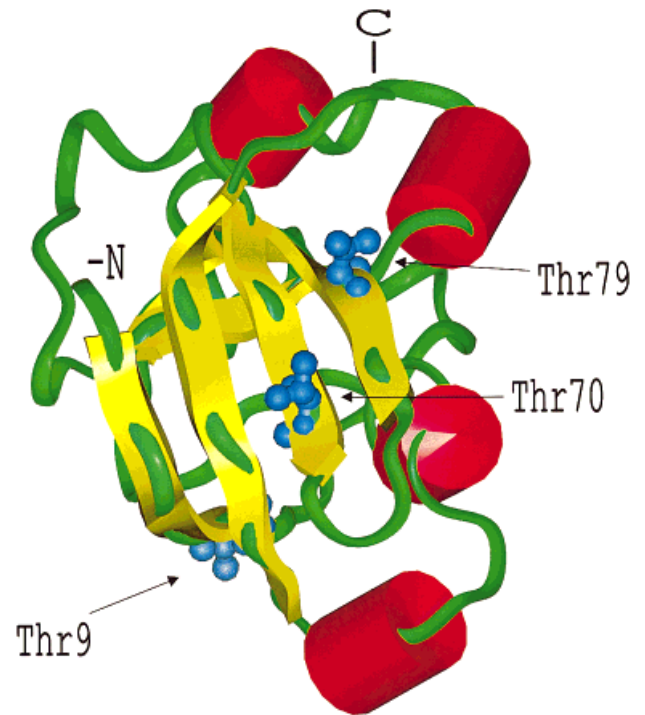


Fig. 9. Spacefill view of wheatwin1 model. Each couple of images shows the protein surface along cartesian axes (positive and negative semi-axis). Colors evidence sites with high (red) and medium (orange) hydrolysis rate. Nonhydrolyzed sites are colored in yellow, while blue residues are not recognized by the proteinases utilized in digesting the protein.

TABLE II. Accessibility and Hydrolysis of Native Wheatwin1 Residues<sup>†</sup>

RES	Access average	SD	%Cut 1h	RES	Access average	SD	%Cut 1h	RES	Access average	SD	%Cut 1h	RES	Access average	SD	%Cut 1h	RES	Access average	SD	%Cut 1h
GLN <sub>1</sub>	85,875	5,252	nd	VAL <sub>27</sub>	27,125	11,264	—	GLY <sub>53</sub>	33,885	7,522	—	THR <sub>79</sub>	27,605	5,000	6	ASN <sub>105</sub>	75,85	5,338	16
GLN <sub>2</sub>	73,695	4,946	nd	SER <sub>28</sub>	67,07	5,942	—	PRO <sub>54</sub>	95,125	15,880	nd	ALA <sub>80</sub>	0,665	0,941	—	GLY <sub>106</sub>	31,775	5,526	—
ALA <sub>3</sub>	13,405	2,130	5	ALA <sub>29</sub>	8,795	1,381	—	ALA <sub>55</sub>	35,86	2,816	—	ARG <sub>81</sub>	5,32	2,031	—	ILE <sub>107</sub>	60,065	5,348	—
THR <sub>4</sub>	56,595	4,638	—	TYR <sub>30</sub>	56,365	7,707	—	GLY <sub>56</sub>	14,74	3,425	—	ILE <sub>82</sub>	1,89	0,903	—	GLY <sub>108</sub>	2,51	1,584	—
ASN <sub>5</sub>	70,9	3,450	—	CYS <sub>51</sub>	1,26	1,622	nd	ALA <sub>57</sub>	88,5	7,268	—	VAL <sub>83</sub>	2,975	1,088	—	TYR <sub>109</sub>	47,275	4,621	—
VAL <sub>6</sub>	28,145	4,685	—	ALA <sub>32</sub>	10,05	1,394	—	HIS <sub>58</sub>	26,94	10,221	nd	ASP <sub>84</sub>	17,115	2,863	nd	GLN <sub>110</sub>	87,09	5,894	nd
ARG <sub>7</sub>	77,795	5,010	7	THR <sub>33</sub>	75,615	5,924	—	GLY <sub>59</sub>	39,52	7,246	—	GLN <sub>85</sub>	65,675	5,525	nd	GLN <sub>111</sub>	76,345	5,003	nd
ALA <sub>8</sub>	9,595	3,783	—	TRP <sub>34</sub>	67,185	3,373	—	GLN <sub>60</sub>	89,195	4,804	nd	CYS <sub>86</sub>	24,925	4,403	nd	GLY <sub>112</sub>	27,63	4,250	—
THR <sub>9</sub>	30,28	4,871	11	ASP <sub>35</sub>	10,745	3,659	nd	ALA <sub>61</sub>	80,625	6,570	—	ALA <sub>87</sub>	97,31	3,862	—	HIS <sub>113</sub>	48,94	3,831	nd
TYR <sub>10</sub>	16,575	2,046	6	ALA <sub>36</sub>	38,235	4,602	—	SER <sub>62</sub>	5,035	2,133	—	ASN <sub>88</sub>	89,15	5,891	21	LEU <sub>114</sub>	14,89	3,224	—
HIS <sub>11</sub>	39,05	12,926	nd	SER <sub>37</sub>	78,98	2,389	—	CYS <sub>63</sub>	35,505	4,950	nd	GLY <sub>89</sub>	46,82	8,593	—	ASN <sub>115</sub>	49,565	4,459	—
TYR <sub>12</sub>	70,94	5,420	—	LYS <sub>38</sub>	44,825	6,526	—	GLY <sub>64</sub>	73,47	8,711	—	GLY <sub>90</sub>	16,93	7,119	—	VAL <sub>116</sub>	4,35	1,547	—
TYR <sub>13</sub>	51,175	7,868	14	PRO <sub>39</sub>	63,18	2,479	nd	LYS <sub>65</sub>	48,055	5,091	—	LEU <sub>91</sub>	14,57	3,834	—	ASN <sub>117</sub>	16,505	2,622	6
ARG <sub>14</sub>	65,77	3,364	—	LEU <sub>40</sub>	53,175	3,906	17	CYS <sub>66</sub>	0,46	0,912	nd	ASP <sub>92</sub>	9,155	4,915	nd	TYR <sub>118</sub>	6,22	1,682	—
PRO <sub>15</sub>	20,55	4,885	nd	SER <sub>41</sub>	75,6	1,726	—	LEU <sub>67</sub>	3,635	2,311	—	LEU <sub>93</sub>	13,675	4,061	—	GLN <sub>119</sub>	25,335	3,628	nd
ALA <sub>16</sub>	47,92	4,138	15	TRP <sub>42</sub>	27,745	2,458	—	GLN <sub>68</sub>	27,415	5,005	nd	ASP <sub>94</sub>	15,135	4,210	nd	PHE <sub>120</sub>	24,935	2,365	—
GLN <sub>17</sub>	85,39	4,056	nd	ARG <sub>43</sub>	4,485	1,769	—	VAL <sub>69</sub>	5,125	1,587	—	TRP <sub>95</sub>	37,315	6,788	—	VAL <sub>121</sub>	29,625	6,426	9
ASN <sub>18</sub>	59,855	4,306	—	SER <sub>44</sub>	51,42	4,436	9	THR <sub>70</sub>	22,805	3,874	—	ASP <sub>96</sub>	49,995	12,366	nd	ASP <sub>122</sub>	61,55	6,582	nd
ASN <sub>19</sub>	51,615	4,078	—	Gly <sub>45</sub>	76,85	4,088	—	ASN <sub>71</sub>	2,08	0,806	—	THR <sub>97</sub>	11,545	3,475	—	CYS <sub>123</sub>	14,815	3,108	nd
TRP <sub>20</sub>	8,005	1,759	4	TYR <sub>46</sub>	24,03	3,586	—	PRO <sub>72</sub>	58,235	2,928	nd	VAL <sub>98</sub>	6,715	2,462	5	ARG <sub>124</sub>	86,21	3,685	29
ASP <sub>21</sub>	12,395	2,856	nd	GLY <sub>47</sub>	9,745	2,338	—	ALA <sub>73</sub>	82,765	9,168	—	PHE <sub>99</sub>	5,16	1,275	—	ASP <sub>125</sub>	96,145	3,958	nd
LEU <sub>22</sub>	2,195	1,278	—	TRP <sub>48</sub>	1,345	0,672	—	THR <sub>74</sub>	37,19	6,135	—	THR <sub>100</sub>	32,19	4,563	12				
GLY <sub>23</sub>	75,515	1,834	—	THR <sub>49</sub>	2,125	1,087	—	GLY <sub>75</sub>	66,28	4,354	—	LYS <sub>101</sub>	59,67	3,255	—				
ALA <sub>24</sub>	17,445	2,023	—	ALA <sub>50</sub>	7,155	2,739	—	ALA <sub>76</sub>	29,245	1,992	—	ILE <sub>102</sub>	2,58	1,039	—				
PRO <sub>25</sub>	96,135	2,091	nd	PHE <sub>51</sub>	20,42	6,856	—	GLN <sub>77</sub>	59,9	5,880	nd	ASP <sub>103</sub>	12,675	1,571	nd				
ALA <sub>26</sub>	88,59	11,932	—	CYS <sub>52</sub>	41,04	7,979	nd	ILE <sub>78</sub>	22,305	5,416	—	THR <sub>104</sub>	74,07	3,074	—				

<sup>†</sup>Residues accessibilities were calculated as average in the 20 models of wheatwin1 based on the barwin atom coordinates stored in PDB. The standard deviations and hydrolysis percent at 1h incubation are also reported. The 22 residues not recognized by the enzymes are indicated as "nd" (not determined).

comparison of protein proteolytic sites and serine-proteinase inhibitors shows that proteolytic sites do not possess a conformation similar to the inhibitors binding loop, at least in their crystal structures.<sup>65</sup> However, in solution, such sites can be involved in structural changes. Modelling studies verified how proteolytic sites can be designed to adopt an inhibitor-like conformation. Besides loops, only helices could adopt such conformation, whereas beta-strands were shown to be less suitable.<sup>66</sup> In our study, the most hydrolyzed sites occur on the wheatwin1 surface according to their predicted high accessibility (Table II and Fig. 9). Two buried hydrolyzed sites (Trp<sub>20</sub>, Val<sub>98</sub>) occur in loop regions, each of them being 13 residues long. Loop regions are in general highly flexible in proteins, so they can easily fluctuate adapting their conformation to fit the proteinase catalytic site, and modifying residue accessibility, too. In our study, we found that: i) 45% of proteolytic sites occur in loop regions, ii) 25% in helices, and iii) 30% in beta-strands. The hydrolyzed sites occurring in loop regions show good agreement between the increase of hydrolysis rates and the increase of accessibilities. As recently reported by Hubbard et al.,<sup>66</sup> helices can also modify their backbone structure to adopt a conformation suitable to interact with proteinases. In the case of wheatwin1 (as well as in barwin, the reference structure), helices are very short, so that barwin is classified as an all-beta protein in both CATH and SCOP databases.<sup>63,64</sup> Moreover, short helices (5–6 amino acids) are stabilized by just a couple of main-chain H-bonds, so that their conformation may be easily disrupted. Sites found in helices are well exposed and show high hydrolysis rates. Regarding beta-strands,

four sites occur in the middle of a strand, namely Ala<sub>3</sub>, Arg<sub>7</sub>, Thr<sub>79</sub>, Asn<sub>117</sub> and are hydrolyzed at a low rate. Other two sites (Thr<sub>9</sub> and Val<sub>121</sub>), at the extreme residue of a strand, are similarly exposed but show a higher rate of hydrolysis. All these six sites occur in edge strands, in perfect agreement with the expectation of modelling studies.<sup>66</sup> The reason why some sites are not hydrolyzed can be explained by further considerations about the secondary structure. As an example, ten threonine residues are potential sites for elastase or subtilisin; while two are quite hidden (Thr<sub>49</sub> and Thr<sub>97</sub>), eight of them show a high accessibility, but only three (Thr<sub>9</sub>, Thr<sub>79</sub>, Thr<sub>100</sub>) are effectively hydrolyzed (Table II). It is interesting to highlight that Thr<sub>100</sub> is enclosed in a short helix, Thr<sub>9</sub> and Thr<sub>79</sub> are in edge beta-strands, while the not hydrolyzed Thr<sub>70</sub> occurs in the central strand of the main beta-sheet (Fig. 8), in agreement with previous considerations.<sup>66</sup> Furthermore, while elastase was able to nick the protein at level of Thr<sub>9</sub>, Thr<sub>79</sub>, and Thr<sub>100</sub> (Fig. 7), subtilisin produced cleavages just at level of Thr<sub>9</sub> (Fig. 6), showing that, although the enzymes utilized can recognize common residues on the basis of their specificity, the conformation that such residues adopt in solution is not suitable to interact with the catalytic sites of all the proteases. It is also to be stressed that some putative sites occurring in well-exposed loop regions are not hydrolyzed. Thr<sub>4</sub>, Thr<sub>74</sub>, and Thr<sub>104</sub> are located in short loops. One reasonable explanation could be that they are constrained in their flexibility by two close secondary structures. Finally, no cut was observed at level of Thr<sub>33</sub>, although this residue is well exposed (Table II) and located in the middle of a short helix, similarly to

Thr<sub>100</sub>. We can suppose that the protease's inefficiency in nicking the protein at this site may be due to the closeness of the voluminous side chain of Trp<sub>34</sub>.

### CONCLUSION

The aim of the present work is the setting up of a rapid method to test a 3D model of proteins of unknown structure based on spatial coordinates of homologous molecules stored in data banks. This should allow the users to develop various applications, even if the reference protein is not materially available. Our data evidence that only when conformational as well as exposure requirements are satisfied, can high hydrolysis rate be attained. It is also important to stress that the rapid and quantitative approach we utilized should represent a useful tool to supplement NMR analysis in studying protein structure. In fact, it appeared to be very effective in defining the wheatwin1 surface accessibility and probing the conformation obtained by its modelled structure.

### REFERENCES

- Ludvigsen S, Poulsen FM. Three-dimensional structure in solution of barwin, a protein from barley seeds. *Biochemistry* 1992;31: 8783–8789.
- Lecomte JT. NMR/X-ray workshop: an overview. In: Villafranca JJ, editor. *Techniques in protein chemistry II*. San Diego: Academic Press; 1991. p 337–345.
- Wütrich K. The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. *Acc Chem Res* 1989;22:36–44.
- Lesk AM, Tramontano A. Computational analysis of protein structure: ponder the known, model the unknown. *Binary* 1992;4: 15–16.
- Lesk AM, Tramontano A. The computational analysis of protein structures: Sources, methods, systems and results. *J Res Natl Institute of Standards and Technology* 1989;94:85–93.
- Chothia C, Lesk AM. The relation between divergence of sequence and structure in proteins. *EMBO J* 1986;5:823–826.
- Blundell TL, Sibanda LB, Sternberg MJ, Thornton JM. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* 1987;326:347–352.
- Wagner SD, Melton LG, Yan Y, Erickson BW, Anderegg RJ. Deuterium exchange of  $\alpha$ -helices and  $\beta$ -sheets monitored by electrospray ionization mass spectrometry. *Protein Sci* 1994;3: 1305–1314.
- Zhang Z, Smith DL. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. *Protein Sci* 1993;2:522–531.
- Zappacosta F, Ingallinella P, Scaloni A, et al. Surface topology of Minibody by selective chemical modifications and mass spectrometry. *Protein Sci* 1997;6:1901–1909.
- Suckau D, Mak M, Przybylski M. Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc Natl Acad Sci USA* 1992;89:5630–5634.
- Ohguro H, Palczewski K, Walsh KA, Johnson RS. Topographic study of arrestin using differential chemical modification and hydrogen/deuterium exchange. *Protein Sci* 1994;3:2428–2434.
- Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. Probing the conformational state of apomyoglobin by limited proteolysis. *J Mol Biol* 1997;266:223–230.
- Fontana A, Zambonin M, De Filippis V, Bosco M, Polverino de Laureto P. Limited proteolysis of cytochrome c in trifluoroethanol. *FEBS Lett* 1995;362:266–270.
- Polverino de Laureto P, Scaramella E, De Filippis V, Bruix M, Rico M, Fontana A. Limited proteolysis of ribonuclease A with thermolysin in trifluoroethanol. *Protein Sci* 1997;6:860–872.
- Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M, Fontana A. Limited proteolysis of lysozyme in trifluoroethanol. Isolation and characterization of a partially active enzyme derivative. *Eur J Biochem* 1995;230:779–787.
- Polverino de Laureto P, De Filippis V, Di Bello M, Zambonin M, Fontana A. Probing the molten globule state of alpha-lactalbumin by limited proteolysis. *Biochemistry* 1995;34:12596–12604.
- Polverino de Laureto P, Toma S, Tonon G, Fontana A. Probing the structure of human growth hormone by limited proteolysis. *Int J Pept Protein Res* 1995;45:200–208.
- Loomes KM. Structural organisation of human bile-salt-activated lipase probed by limited proteolysis and expression of a recombinant truncated variant. *Eur J Biochem* 1995;230:607–613.
- McGinley MD, Narhi LO, Kelley MJ, Davy E, Robinson J, Rohde MF, Wright SD, Lichenstein HS. CD14: physical properties and identification of an exposed site that is protected by lipopolysaccharide. *J Biol Chem* 1995;270:5213–5218.
- Pfeil W, Nolting BO, Jung C. Apocytochrome P450cam is a native protein with some intermediate-like properties. *Biochemistry* 1993;32:8856–8862.
- Gentile S, Salvatore G. Preferential sites of proteolytic cleavage of bovine, human and rat thyroglobulin. *Eur J Biochem* 1993;218: 1847–1851.
- Tozawa K, Miyauchi M, Yoshida M. Structure of the alpha subunit of F1-ATPase probed by limited proteolysis. *J Biol Chem* 1993;268: 19044–19054.
- Calvete JJ, Mann K, Alvarez MV, Lopez MM, Gonzalez-Rodriguez J. Proteolytic dissection of the isolated platelet fibrinogen receptor, integrin GPIIb/IIIa. Localization of GPIIb and GPIIIa sequences putatively involved in the subunit interface and in intrasubunit and intrachain contacts. *Biochem J* 1992;282:523–532.
- Arnone MI, Birolo L, Giamberini M, et al. Limited proteolysis as a probe of conformational change in aspartate aminotransferase from *Sulfolobus solfataricus*. *Eur J Biochem* 1992;204:1183–1189.
- Rudolph R. The use of limited proteolysis in studies on protein folding and association. *Biochem Soc Trans* 1985;13:308–311.
- Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin M. Probing the partly folded states of proteins by limited proteolysis. *Fold Des* 1997;2:R17–R26.
- Vindigni A, De Filippis V, Zanotti G, Visco C, Orsini G, Fontana A. Probing the structure of hirudin from *Hirudinaria manillensis* by limited proteolysis. Isolation, characterization and thrombin-inhibitory properties of N-terminal fragments. *Eur J Biochem* 1994;226:323–333.
- Martini F, Aceto A, Sacchetta P, Bucciarelli T, Dragani B, Di Ilio C. Investigation of intra-domain and inter-domain interactions of glutathione transferase P1–1 by limited chymotryptic cleavage. *Eur J Biochem* 1993;218:845–851.
- Sun AQ, Yuksel KU, Gracy RW. Limited proteolysis of triosephosphate isomerase and characterization of the catalytically active peptide complex. *J Biol Chem* 1993;268:26872–26878.
- Chen SX, Hardin CC, Swaisgood HE. Purification and characterization of beta-structural domains of beta-lactoglobulin liberated by limited proteolysis. *J Protein Chem* 1993;12:613–625.
- Sharma AK, Minke-Gogl V, Gohl P, Siebendritt R, Jaenicke R, Rudolph R. Limited proteolysis of gamma II-crystallin from calf eye lens. Physicochemical studies on the N-terminal domain and the intact two-domain protein. *Eur J Biochem* 1990;194:603–609.
- Bothner B, Dong XF, Bibbs L, Johnson JE, Siuzdak G. Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J Biol Chem* 1998;273:673–676.
- Zappacosta F, Pessi A, Bianchi E, et al. Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody. *Protein Sci* 1996;5:802–813.
- Arnold U, Rucknagel KP, Schierhorn A, Ulbrich-Hofmann R. Thermal unfolding and proteolytic susceptibility of ribonuclease A. *Eur J Biochem* 1996;237:862–869.
- Stoller G, Tradler T, Rucknagel KP, Rahfeld JU, Fischer G. An 11.8 kDa proteolytic fragment of the E. coli trigger factor represents the domain carrying the peptidyl-prolyl cis/trans isomerase activity. *FEBS Lett* 1996;384:117–122.
- Cohen SL, Ferré-D'Amaré AR, Burley KS, Chait BT. Probing the solution structure of the DNA-binding protein Max by a combina-

- tion of proteolysis and mass spectrometry. *Protein Sci* 1995;4:1088–1099.
38. Seielstad DA, Carlson KE, Kushner PJ, Greene GL, Katzenellenbogen JA. Analysis of the structural core of the human estrogen receptor ligand binding domain by selective proteolysis/mass spectrometric analysis. *Biochemistry* 1995;34:12605–12615.
  39. Monks SA, Gould AR, Lumley PE, et al. Limited proteolysis study of structure-function relationships in Sh I, a polypeptide neurotoxin from a sea anemone. *Biochim Biophys Acta* 1994;1207:93–101.
  40. Massotte D, Yamamoto M, Scianimanico S, et al. Structure of the membrane-bound form of the pore-forming domain of colicin A: a partial proteolysis and mass spectrometry study. *Biochemistry* 1993;32:13787–13794.
  41. Brockerhoff SE, Edmonds CG, Davis TN. Structural analysis of wild-type and mutant yeast calmodulins by limited proteolysis and electrospray ionization mass spectrometry. *Protein Sci* 1992;1:504–516.
  42. Scaloni A, Miraglia N, Orru' S, et al. Topology of the calmodulin-mellitin complex. *J Mol Biol* 1998;277:945–958.
  43. Caporale C, Sepe C, Caruso C, Buonocore V. A computer program to compare sequence fingerprints of homologous proteins for the rapid assessment of their primary structure differences. *J Protein Chem* 1998;17:867–873.
  44. Caporale C, Caruso C, Buonocore V. Determination of the primary structure of homologous proteins by sequence analyses of peptide mixtures. *J Protein Chem* 1996;15:405–412.
  45. Caporale C, Sepe C, Caruso C, Petrilli P, Buonocore V. An algorithm to determine protein sequence alignment by utilizing data obtained from a peptide mixture and individual peptides. *Comput Appl Biosci* 1994;10:489–494.
  46. Caporale C, Caruso C, Petrilli P, Sepe C, Poerio E, Buonocore V. A computer program to determine the amino acid sequence of proteins by utilizing data obtained from peptide mixtures. *Protein Seq Data Anal* 1993;5:337–344.
  47. Johnson RS, Walsh KA. Sequence analysis of peptide mixtures by automated integration of Edman and mass spectrometric data. *Protein Sci* 1992;1:1083–1091.
  48. Henzel WJ, Rodriguez H, Watanabe C. Computer analysis of automated Edman degradation and amino acid analysis data. *J Chromatogr* 1987;404:41–52.
  49. Nielsen PR, Landis B, Svoboda M, Schneider K, Przybylski M. N-terminal sequence determination of polypeptides and peptide mixtures by Edman degradation combined with californium-252 plasma desorption mass spectrometry. *Anal Biochem* 1990;191:302–308.
  50. Caporale C, Sepe C, Caruso C, Garzillo AMV, Buonocore V. An algorithm to analyse the hydrolysis pathway of peptides and proteins by sequence analyses of unfractionated digestion mixtures. *Comput Appl Biosci* 1996;12:81–88.
  51. Caporale C, Garzillo AMV, Caruso C, Buonocore V. Characterization of extracellular proteases from *Trametes trogii*. *Phytochemistry* 1996;41:385–393.
  52. Petrilli P, Caporale C, Caruso C. Hydrolysis pattern of procasomorphin by gut proteases from plant parasite *Heliothis Zea* determined by sequence analyses performed on the unfractionated digestion mixtures. *Int J Pept Protein Res* 1994;43:201–204.
  53. Masiarz FR, Malcom BA. Rapid determination of endoprotease specificity using peptide mixtures and Edman degradation analysis. *Methods Enzymol* 1994;241:302–310.
  54. Caporale C. Analytical techniques and computer algorithms combined for the rapid characterisation of structural peptide and protein features. *J Pept Res* 1998;52:421–429.
  55. Caruso C, Caporale C, Poerio E, Facchiano A, Buonocore V. The amino acid sequence of a protein from wheat kernel closely related to proteins involved in the mechanisms of plant defence. *J Protein Chem* 1993;12:379–386.
  56. Ludvigsen S, Poulsen FM. Secondary structure in solution of barwin from barley seed using <sup>1</sup>H nuclear magnetic resonance spectroscopy. *Biochemistry* 1992;31:8771–8782.
  57. Svensson B, Svendsen I, Hojrup P, Roepstorff P, Ludvigsen S, Poulsen FM. Primary structure of barwin: a barley seed protein closely related to the C-terminal domain of proteins encoded by wound-induced plant genes. *Biochemistry* 1992;31:8767–8770.
  58. Carrano L, Nitti G, Buonocore V, Caporale C, Poerio E. An effective purification procedure of amylase and trypsin inhibitors from wheat flour; isolation of a new water-soluble protein. *Plant Sci* 1989;65:25–31.
  59. Caruso C, Caporale C, Chilosi G, et al. Structural and antifungal properties of a pathogenesis-related protein from wheat kernel. *J Protein Chem* 1996;15:35–44.
  60. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
  61. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - A program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26:283–291.
  62. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–2637.
  63. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;247:536–540.
  64. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH - A hierarchic classification of protein domain structures. *Structure* 1997;5:1093–1108.
  65. Hubbard SJ, Campbell SF, Thornton JM. Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J Mol Biol* 1991;220:507–630.
  66. Hubbard SJ, Eisenmenger F, Thornton JM. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci* 1994;3:757–768.