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

Carlo Caporale,1,∗ Carla Caruso,1 Angelo Facchiano,2 Monica Nobile,1 Luca Leonardi,2 Laura Bertini,1 Giovanni Colonna,2 and Vincenzo Buonocore1
1Dipartimento di Agrobiologia ed Agrochimica, Università della Tuscia, Viterbo, Italy
2CRISCEB-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 barley, 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,1 Lecompt,2 and Wutrich,3 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.4–7 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 reactions8,9 and chemical modifications10–12 or identifying them by enzymatic hydrolysis of their peptidic bonds.13–25 In general, the use of proteolytic enzymes has been shown to be an excellent tool in studying protein folding.26,27 Using this approach, peptidic fragments are usually purified and successively identified. The identification can be accomplished by classical chemistry methodologies28–32 or by mass spectrometry analysis.8–11,33–42 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 reported43–49 together with methods for determining the hydrolysis pathway of polypeptides by unknown digesting agents.50–53 These approaches, which have been recently reviewed,54 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.43–54 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,
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.55 Furthermore, a predictive structural model based on the knowledge of the tertiary structure in solution of barley1,56 a highly homologous protein from barley, 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 Cerealcoltura (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 µ-Bondapak C18 column (0.39 x 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.58,59 The protein concentration was determined by the Bio-Rad assay, following the manufacturer’s instructions and using bovine serum albumin as a standard.
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 µ-Bondapak C18 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”. The theoretical initial yield of the coupling reaction was 65% using β-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 barwin1 (pdb code: 1bw3). The alignment of wheatwin1 and barwin did not require deletion or insertion of gap. The program MODELLER 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. Secondary structure was assigned by the program DSSP. Search for structural classification of barwin was performed on SCOP and CATH databases. Solvent accessibility of amino acids was evaluated by the program NACCESS 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
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.55 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 fragments 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 hydrolysies 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,
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.43–54 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).55

**Digestion With Endoproteinase Arg-C**

The single-literal 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 R124–D125 is expected in the digestion mixtures. 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 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). Table I shows 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.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Hydrolysis site</th>
<th>First residue useful for the identification</th>
<th>Edman step</th>
<th>pmol 0 min</th>
<th>pmol 20 min</th>
<th>pmol 60 min</th>
<th>pmol 120 min</th>
<th>pmol 240 min</th>
<th>pmol 720 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg-C</td>
<td>R124</td>
<td>D125</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>164</td>
<td>183</td>
<td>186</td>
<td>188</td>
</tr>
<tr>
<td>Trypsin</td>
<td>R7</td>
<td>A8</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>66</td>
<td>71</td>
<td>80</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>R124</td>
<td>D125</td>
<td>1</td>
<td>1</td>
<td>74</td>
<td>287</td>
<td>321</td>
<td>326</td>
<td>329</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>A3</td>
<td>T4</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>52</td>
<td>55</td>
<td>58</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Y10</td>
<td>H11</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>56</td>
<td>63</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>W20</td>
<td>D21</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>41</td>
<td>49</td>
<td>51</td>
<td>55</td>
</tr>
<tr>
<td>Subtilisin</td>
<td>T9</td>
<td>H11</td>
<td>2</td>
<td>0</td>
<td>16</td>
<td>108</td>
<td>114</td>
<td>117</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>Y13</td>
<td>R14</td>
<td>1</td>
<td>1</td>
<td>19</td>
<td>137</td>
<td>142</td>
<td>146</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>A16</td>
<td>Q17</td>
<td>1</td>
<td>1</td>
<td>13</td>
<td>119</td>
<td>123</td>
<td>127</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>N88</td>
<td>G90</td>
<td>2</td>
<td>1</td>
<td>29</td>
<td>128</td>
<td>134</td>
<td>139</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>V88</td>
<td>F99</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>47</td>
<td>51</td>
<td>54</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>N105</td>
<td>I107</td>
<td>2</td>
<td>1</td>
<td>22</td>
<td>158</td>
<td>164</td>
<td>166</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>N127</td>
<td>Q119</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>56</td>
<td>61</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>V122</td>
<td>D122</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>93</td>
<td>99</td>
<td>104</td>
<td>107</td>
</tr>
<tr>
<td>Elastase</td>
<td>T9</td>
<td>Y10</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>108</td>
<td>118</td>
<td>121</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>A16</td>
<td>Q17</td>
<td>1</td>
<td>1</td>
<td>18</td>
<td>153</td>
<td>160</td>
<td>167</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>L40</td>
<td>S41</td>
<td>1</td>
<td>0</td>
<td>15</td>
<td>168</td>
<td>176</td>
<td>181</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>S44</td>
<td>G45</td>
<td>1</td>
<td>2</td>
<td>11</td>
<td>92</td>
<td>97</td>
<td>103</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>T79</td>
<td>A80</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>63</td>
<td>66</td>
<td>71</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>V88</td>
<td>F99</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>49</td>
<td>53</td>
<td>55</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>T300</td>
<td>K103</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>127</td>
<td>134</td>
<td>138</td>
<td>142</td>
</tr>
</tbody>
</table>

*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.*

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.
Endoproteinase Arg-C produced the major amount of R_{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
allowed us to unambiguously identify the sequence up to Y_{13} (not shown). The quantitative datum of Ala at the first step (66 pmol) indicates that the hydrolysis at level of R_{7} occurred at a minor extent with respect to the hydrolysis at level of R_{124}, 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_{4}-N_{5}-V_{6}, H_{11}-Y_{12}-Y_{13}, and D_{21}-L_{22} G_{23} can be reconstructed by these residues (Fig. 1A), showing that cleavages at level of A_{3}, Y_{10}, and W_{20} 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).
### 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 sequence information achieved from the sequence analysis of the native mixture at 1h incubation. Pmol data of all PTH amino acids obtained at each step of degradation 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 and their pmol data. 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).
L40, S44, T79, and T100, 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 modeled using the barwin 3D structure as reference. Alignment of wheatwin1 and barwin sequences shows only seven amino acid differences, without deletion or insertion (Fig. 2). The program MODELLER was used to generate 20 wheatwin1 modeled structures based on the twenty NMR barwin structures stored in Protein Data Bank. Differences among modeled structures were very subtle, and loop regions very similar in all of them. The DSSP program 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 I, 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 1 h 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 evidentiate 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 Thr9, Thr79 and Thr70 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 Wheat1 Residues

<table>
<thead>
<tr>
<th>RES</th>
<th>Access average</th>
<th>SD</th>
<th>% Cut 1h RES</th>
<th>Access average</th>
<th>SD</th>
<th>% Cut 1h RES</th>
<th>Access average</th>
<th>SD</th>
<th>% Cut 1h RES</th>
<th>Access average</th>
<th>SD</th>
<th>% Cut 1h RES</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLN1</td>
<td>85,875</td>
<td>5,252</td>
<td>nd</td>
<td>GLN1</td>
<td>85,875</td>
<td>5,252</td>
<td>nd</td>
<td>GLN1</td>
<td>85,875</td>
<td>5,252</td>
<td>nd</td>
<td>GLN1</td>
</tr>
<tr>
<td>VAL1</td>
<td>72,125</td>
<td>11,264</td>
<td>11.644</td>
<td>VAL1</td>
<td>72,125</td>
<td>11,264</td>
<td>11.644</td>
<td>VAL1</td>
<td>72,125</td>
<td>11,264</td>
<td>11.644</td>
<td>VAL1</td>
</tr>
<tr>
<td>GLY1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>GLY1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>GLY1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>GLY1</td>
</tr>
<tr>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
</tr>
<tr>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
</tr>
<tr>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
</tr>
<tr>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
</tr>
<tr>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
</tr>
<tr>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
</tr>
<tr>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
<td>28,145</td>
<td>4,078</td>
<td>nd</td>
<td>GLU1</td>
</tr>
<tr>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
<td>33,885</td>
<td>7,522</td>
<td>11.644</td>
<td>LYS1</td>
</tr>
<tr>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
<td>17,445</td>
<td>2,023</td>
<td>nd</td>
<td>LYS1</td>
</tr>
</tbody>
</table>

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.65 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.66 In our study, the most hydrolyzed sites occur on the wheat1 surface according to their predicted high accessibility (Table II and Fig. 9). Two buried hydrolyzed sites (Trp20, Val98) occur in loop regions, each of them being 13 residues long. Loop regions are in general highly flexible in proteins, so that 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 accessibility. As recently reported by Hubbard et al.,66 helices can also modify their backbone structure to adopt a conformation suitable to interact with proteinases. In the case of wheat1 (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.63,64 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 Ala9, Arg7, Thr79, Asn117 and are hydrolyzed at a low rate. Other two sites (Thr9 and Val121), 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.66 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 (Thr40 and Thr97), eight of them show a high accessibility, but only three (Thr9, Thr79, Thr100) are effectively hydrolyzed (Table II). It is interesting to highlight that Thr100 is endo at a short helix, Thr9 and Thr79 are in edge beta-strands, while the not hydrolyzed Thr70 occurs in the central strand of the main beta-sheet (Fig. 8), in agreement with previous considerations.66 Furthermore, while elastase was able to nick the protein at level of Thr9, Thr79, Thr100 (Fig. 7), subtilisin produced cleavages just at level of Thr9 (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 proteinases. It is also to be stressed that some putative sites occurring in well-exposed loop regions are not hydrolyzed. Thr4, Thr74, and Thr100 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 Thr33, although this residue is well exposed (Table II) and located in the middle of a short helix, similarly to
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 Trp34.

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 modeled structure.

REFERENCES

10. Zappacosta F, Ingallinella P, Scaloni A, et al. Surface topology of Thr100. 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 Trp34.
1099.
103.
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.
302–308.