MS/MS identification strategies for the arginine and lysine protein group when searches of spectra were carried out against different databases.

INTRODUCTION

1. Wheat gluten proteins represent ~80% of the protein in wheat flour and confer the unique viscoelastic properties that enable the production of bread, noodles and many other food products [1].

2. The gluten proteins are made up of a large number of proteins with very similar sequences that fall into five major groups, the α-, γ- and ω-gliadins and the high- and low-molecular weight glutenin subunits (HMW-GS and LMW-GS).

3. Protein within the α-, γ- and ω-gliadin groups are monomeric proteins that are associated with extensibility properties while proteins from HMW-GS and LMW-GS groups are linked by cysteine residues into large polymers that confer elasticity properties to wheat flour doughs.

4. All gluten proteins have high percentages of proline and glutamine and low percentages of arginine and lysine making it difficult to produce suitable and sufficient trypsin fragments for mass spectrometry (Table 1).

5. There is considerable heterogeneity of gluten protein sequences among different wheat cultivars.

6. Individual proteins within the major gluten protein groups respond differentially to elevated temperature and fertilizer during grain development, altering the protein composition and functionality of wheat flour [2,3]. To better understand how crop management and environment impact wheat flour quality, we developed MS/MS methods to improve identification of these highly similar proteins in the US wheat cultivar Butte 86.

DATA ANALYSIS

1. Initial searches of spectra were carried out against the “SuperWheat” (S_W) database using MASCOT and XTandem.

2. The “Superwheat” database contained 2,049,765 entries and included proteins encoded by all NCBI nonredundant green plant sequences, proteins encoded by contigs from three public EST assemblies, and proteins encoded by contigs assembled from ESTs from the wheat cultivar Butte 86 that was used for these experiments [5].

3. Search engine results from Mascot, XTandem or both Mascot and XTandem were analyzed separately for each enzyme or combination of enzymes using Scaffold [6] (Table 2 & Fig 2).

RESULTS

1. The database size influenced the number of hits. 48 gluten proteins passed the filter setting for the initial S_W database search. 12 of these were sequences from Butte 86.

2. A greater number of gluten proteins were identified when the database size was reduced to less than 30,000 sequences. 59 gluten proteins were identified using the Ta_B86 database (that did not contain the Subset database) and 16 of these were assigned to Butte 86 sequences.

3. Adding the “subset” database to the search (Ta_B86 Reverse Decoy) increased the total protein number passing the filter to 62, but decreased the number of Butte 86 proteins identified.

4. The number of proteins passing the filter ranged from 64 to 70 when subset database searches were performed and more of those proteins were assigned to sequences from Butte 86. The greatest number of Butte 86 sequences was obtained when M. mazei was used as the decoy.

CONCLUSIONS

To maximize the number of cultivar-specific wheat gluten proteins identified by MS/MS:

- Digest proteins separately with chymotrypsin, thermolysin and trypsin.
- Use both MASCOT and XTandem search engines for analysis of spectra.
- Perform first pass searches on a large SuperWheat database followed by second pass searches on a subset database that contains M. mazei or random decoy sequences.

REFERENCES


Table 1. Number of proteins identified within each gluten protein group when searches of spectra were carried out against different databases.

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Enzyme</th>
<th>Chymo+</th>
<th>Thermo</th>
<th>All</th>
<th>Mascot</th>
<th>X!Tandem</th>
<th>Both</th>
<th>Mascot</th>
<th>X!Tandem</th>
<th>Both</th>
<th>Mascot</th>
<th>X!Tandem</th>
<th>Both</th>
<th>Mascot</th>
<th>X!Tandem</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-gliadin</td>
<td>53</td>
<td>101</td>
<td>141</td>
<td>24</td>
<td>23</td>
<td>56</td>
<td>7</td>
<td>46</td>
<td>67</td>
<td>18</td>
<td>15</td>
<td>10</td>
<td>18</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>γ-gliadin</td>
<td>55</td>
<td>85</td>
<td>121</td>
<td>102</td>
<td>75</td>
<td>152</td>
<td>32</td>
<td>108</td>
<td>162</td>
<td>65</td>
<td>71</td>
<td>69</td>
<td>65</td>
<td>65</td>
<td>68</td>
<td>64</td>
</tr>
<tr>
<td>ω-gliadin</td>
<td>46</td>
<td>101</td>
<td>141</td>
<td>24</td>
<td>23</td>
<td>56</td>
<td>7</td>
<td>46</td>
<td>67</td>
<td>18</td>
<td>15</td>
<td>10</td>
<td>18</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>LMW-GS</td>
<td>40</td>
<td>104</td>
<td>144</td>
<td>24</td>
<td>23</td>
<td>56</td>
<td>7</td>
<td>46</td>
<td>67</td>
<td>18</td>
<td>15</td>
<td>10</td>
<td>18</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>145</td>
<td>310</td>
<td>455</td>
<td>48</td>
<td>48</td>
<td>96</td>
<td>13</td>
<td>70</td>
<td>123</td>
<td>29</td>
<td>23</td>
<td>16</td>
<td>23</td>
<td>18</td>
<td>21</td>
<td>20</td>
</tr>
</tbody>
</table>