Addressing The Practical Challenges Of Recombinant Protein Production In Corn: Purification Strategies

**Introduction**

Corn offers many advantages as a transgenic expression host: -Capable of targeted expression (e.g. in the germ, endosperm) with proven fractionation methods (e.g. dry- or wet-milling) to reduce containment load -Ease & economic scalability is high -At least half of the world’s food protein is derived from corn -High protein production per plant vs. other plants

**Background**

Previously a simple and relatively inexpensive method was developed that characterizes an individual HCP present in a mixture using only three physical/chemical properties: pI (isoelectric point), MW (molecular weight) and surface hydrophobicity (SH). The first two are determined using 2D electrophoresis (2DE) and the latter is quantified from the protein's partition coefficient (K) across aqueous two-phase partitioning (ATP) and when coupled together allow for the 3D-characterization of a protein (4). In addition, retention times for a set of model proteins were able to be accurately predicted during chromatography on CEX (using a partial least squares (PLS) regression model developed with only the 3 properties as inputs (5)). The work presented here applies the 3D characterization method to a corn germ protein (TCGP) each individual protein accounted for, creating a database of 4 characterized proteins & univariate characterization properties.

**Statistical Analysis Background: Multivariate Random Forest**

Multivariate Random Forest (MVRF): An expansion of the univariate classification and regression tree (CART) methodology, an example of which is shown below. The data are divided into subsets using binary (yes or no) questions that are expressed in the form of a decision tree. Each split is phrased to maximize between subset variance and is guided by a complex split function algorithm used to grow trees for MVRF.

**Experimental Methods**

**3D Characterization Method (4):**

- **Stain, Scan & Analyze Gels:** 2D gel stained with Blue Coomassie (Bio-Rad) [84:4:2:50] or Sypro Ruby (Invitrogen) (Progenics)

**3Dmesh plots illustrating accuracy of MVRF method by gradient fraction between % of each protein that was observed to elute (experimental) & % predicted to elute in each elution fraction for all 3 characterization properties.**

**Results**

**MVRF Prediction Performance for CEX Using Model Protein Dataset**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Data Set</th>
<th>50% Eluted</th>
<th>75% Eluted</th>
<th>90% Eluted</th>
<th>95% Eluted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aa1</td>
<td>Training</td>
<td>0.0305</td>
<td>0.0304</td>
<td>0.0303</td>
<td>0.0302</td>
</tr>
<tr>
<td>Aa2</td>
<td>Training</td>
<td>0.0312</td>
<td>0.0310</td>
<td>0.0309</td>
<td>0.0308</td>
</tr>
</tbody>
</table>

**References**


**Conclusion**

- The ability to accurately predict the chromatographic behavior of a host cell protein mixture would improve elution efficiency while at the same time minimizing resources.

- Advantages to this approach include: -No use of radioactive materials -Low cost -Can be performed on large scale mixtures -Easy to perform post-translational modifications

**Ongoing Work**

- Apply 3D characterization method to more data in order to determine the optimal combination of characterized methods (diagram shown below)

**Figure:** A 3D mesh plot illustrating the accuracy of MVRF method by gradient fraction between % of each protein that was observed to elute (experimental) & % predicted to elute in each elution fraction for all 3 characterization properties.