Application of Spectrophotometric Fingerprint in Cluster Analysis for Starch Origin Determination

Nikola Sakač1, Maja Karnas2, Jasmina Dobša3, Marija Jozanović4**, Vlatka Gvozdić4, Elvira Kovač-Andrić4, Marija Kraševac Sakač4 and Bojan Šarkanj5

1University of Zagreb, Faculty of Geotechnical Engineering, Hallerova 7, HR-42000 Varaždin, Croatia
2Josip Juraj Strossmayer University of Osijek, Faculty of Agriculture, Vladimira Preloga 1, 31000 Osijek, Croatia
3University of Zagreb, Faculty of Organization and Informatics, Pavlinska 2, 42000 Varaždin, Croatia
4Josip Juraj Strossmayer University of Osijek, Department of Chemistry, Cara Hadrijana 8/A, 31000 Osijek, Croatia
5University Nord, Trg dr. Žarka Dolinara 1, 43000 Koprivnica, Croatia

SUMMARY

The botanical origin of starch is of importance in industrial applications and food processing because it may influence the properties of the final product. Current microscopic methods are time-consuming. Starch consists of an origin-dependent amylose/amylopectin ratio. Triiodide ions bind characteristically to the amylose and amylopectin depending on the starch’s botanical origin. The absorbance of the starch-triiodide complex was measured for: wheat, potato, maize, rye, barley, rice, tapioca and unknown origin starch; and within the different cultivars. Each starch sample provided specific parameters: starch-triiodide peak wavelength maximum ($\lambda_{\text{max}}$/nm), maximum absorbance change for $\lambda_{\text{max}}$ (ΔA) and wavelength ($\lambda_{\text{max}}$) shift towards the unknown origin starch sample. The Vis spectra (500-800 nm) for each starch sample were used as an unique fingerprint, and then elaborated by Cluster analysis. The Cluster analysis managed to distinguish data on two cluster, a cereal type...
cluster and a potato/tapioca/starch cluster. The cereal subclusters extensively distinguished wheat/barley/rye starches from maize starches. Data for cultivars exhibited mostly good agreement within the same subclaster. The proposed method of combining Cluster analysis and Vis absorbance data for starch triiodide complex was able to distinguish starches of different botanical origins and cultivars within the same species. Presented method is simpler and more convenient compared to standard time-consuming methods.

Key words: starch, botanical origin, spectrophotometry, cluster analysis

INTRODUCTION

Starch is one of the major natural polysaccharides. Starch is widely employed in numerous branches of industry, e.g., in food, paper, adhesive, textile, cosmetic and biorefinery (1-4). One of the most important factors that dictate starch processing and final product quality is the botanical origin of starch (5), where appeared a need for a fast and reliable method for starch origin determination. Determination of starch origin in artificial food is also one of current issues.

Starch is found in nature as granules, an immense and highly organized structure. Botanical origin of starch dictates the ratio and the way amylose and amylopectin are associated and packed (6-8). The branched molecules of amylose and amylopectin from various origins have their own characteristic structures such as molecular size, inner chain length, and the number of side chains. For example, wheat amylose probably contains only a small number of very large branched molecules, whereas sweet potato amylose has a small number of relatively large unbranched molecules (9).

Many physicochemical properties of amylose and amylopectin (10,11), such as iodine binding capacity and degree of polymerisation (DP) depend on the botanical origins of the starch. The amylose content and the amylose-amylopectin ratio have traditionally been measured by iodine-binding procedures (12) using various techniques (13,14), especially spectroscopic techniques (15). These procedures are based on the capacity of amylose and amylopectin to form helical inclusion complexes with iodine, which exhibit a blue color. Different spectroscopic (visible light absorption) characteristics of the starch–triiodide complex vary with chain length which is dependent on the botanical origin of starch (16). It is also known that plant species inside one variety have similar amylose/amylopectin ratio (within certain deviations) and knowledge of their spectroscopic properties could contribute to starch origin identification (17,18).
The typical origin analysis of starch includes indirect techniques that measure the differences in the physical and chemical properties of the starches (20). The starch origin identification was based on enzymatic hydrolysis of starch granule (21,22). Recently, a group of authors (5) proposed SDS-PAGE for starch granule-associated protein (SGAPs) determination and use of peptide mass finger printing of granule-bound starch synthase (23) for providing reliable results in starch origin determination. Mentioned methods are time consuming and require specific instrumentation. Therefore, the main goal of the proposed investigations is development of a fast and inexpensive method for identification of starch origin.

The method previously described (24) related to the determination of starch’s origin based on direct potentiometric measurements of starch-triiodide complexes and the use of principal components data analysis (PCA). This method that could distinguish between starches of different botanical origins was based on their chemical differences, amylose/amylopectin ratios and specific triiodide ion binding affinities to amylose and amylopectin.

In this paper a new method was proposed that uses the spectral, instead of potentiometric properties of the starch-triiodide complex. To speed up and simplify the determination, the botanical origin specific triiodide ion binding affinities to amylose and amylopectin were used, instead of time consuming determination of amylose and amylopectin content and property of iodine binding capacity for each sample. Ability of proposed method to distinguish starch types of different botanical origins is verified by statistical data analysis using hierarchical clustering of the complete visible spectra data of the starch-triiodide complex. The method is particularly suitable for use in laboratories equipped with simple UV-Vis spectrophotometers.

MATERIALS AND METHODS

Samples and reagents
Starch samples were isolated from wheat (Srpanjka and Golubica), maize, rye (Eho, Danovski and Conduct), barley (Barun, Vanessa), rice and tapioca, which were obtained at the local market store in Croatia.

Waxy maize, wheat and potato commercial starches were purchased from Sigma-Aldrich (Darmstadt, Germany). Commercial starch of unknown origin was purchased from Kemika (Zagreb, Croatia). Potassium triiodide solution was prepared using iodine (I$_2$) and potassium iodide (KI), both purchased from Sigma--Aldrich (Darmstadt, Germany). Buffer components: glacial acetic acid was purchased from Panreac (Barcelona, Spain) and sodium acetate trihydrate was purchased from J.T. Baker (Deventer, Netherland).
Starch sample preparation
The samples’ seed coats were peeled off, and an alkali steeping method \( (25,26) \) was used to separate the starches. The preparation steps included pH adjustment, blending at low speed (BOSCH blender model: MMBH4P3W, 1600 W), filtrations, three steps of resuspension in deionised water, drying (Instrumentaria ST-05, Zagreb, Croatia), grounding with a mortar and pestle to pass through 60mesh sieve and dry storage.

Starch solution preparation
The starch was dried for 90 minutes at 130 °C by spreading approximately 2 g of air-dried soluble starch in a thin layer over the bottom of a weighing bottle with a lid.
The starch solution was prepared by dissolving the amount of starch equivalent to 0.2 g of anhydrous starch (concentration 2 g/L) in a previously prepared acetate buffer solution (pH 6.0) in a closed 100 mL volumetric flask. After the solution was heated and stirred for 10 minutes in a sonic bath, it was allowed to cool to room temperature and was then diluted to 100 mL with deionized water in a volumetric flask. For each starch type, five independent solutions were prepared. Starch solutions were freshly prepared each day to avoid microbial degradation.

Potassium triiodide solution preparation
The potassium triiodide solution was prepared by dissolving solid iodine (600 µM) in a 0.03 M potassium iodide solution. Iodine is very toxic and easily sublimates, which makes it difficult to weigh – the weighing procedure should be fast but accurate. Iodine has low water solubility; thus, it was dissolved in the following order. First, potassium iodide was dissolved in a small volume of water, making it possible to slowly dissolve the iodine. After the iodide solution was vigorously stirred, the iodine was completely dissolved, and the volumetric flask was filled to the mark.

Procedure
The spectrophotometric measurements were performed on an Avantes optical system -AvaSpec-ULS3648 high resolution spectrophotometer with 400µm Ø optical fibers, measuring cuvette holder and AvaLight-DH-S light source; with AvaSoft 7.0. software (all from Avantes, Apeldoorn, Netherland) \( (27) \). Sonic bath (Bandelin Sonorex, Berlin, Germany) was used for sample solution preparations.
For spectrophotometric measurements, five independent series of starch triiodide solutions were prepared. First, six 50 mL volumetric flasks were incrementally filled with starch in the following volumes: 0, 0.25, 0.50, 0.75, 1.0 and 1.25 mL. The previously prepared potassium triiodide solution was then added to the starch-filled volumetric flasks in opposite incremental volume order: 1.25, 1.0,
0.75, 0.5, 0.25 and 0 mL. The volumetric flasks were filled to the mark with deionized water, stirred in
a sonic bath for 5 minutes and were ready for further investigation. Subsequently, the
spectrophotometric measurement data were collected and analyzed using licensed Statistica 12
software (TIBCO Statistica, Palo Alto, USA) (28).

RESULTS AND DISCUSSION

Spectrophotometric characterization of starch-triiodide complex

Starch samples were characterized by measuring starch triiodide complex absorption spectra. Each
starch sample was measured in a range from 0 to 1.25 mL of added starch. Fig. 1 represents the
spectra of the commercial starch of unknown origin that was used for comparison with other starch
types.

>Fig. 1<

The spectra exhibited three different peaks in the recorded UV-Vis region. The first one, with an
absorption maximum at 564 nm, corresponded to a starch triiodide complex. The second two, with
maxima at 342 and 285 nm, were assigned to triiodide/iodide ions.

With an increase of starch concentration, the starch-triiodide complex concentration also increased,
resulting in absorbance enhancement at the starch-triiodide wavelength maximum. At the same time,
the absorbance of the other two peaks decreased due to the binding of triiodide to starch.

Because the amylose-amyllopectin ratio in starch and specific triiodide ion binding affinities to amylose
and amylopectin depends directly on botanical starch origin, there is a considerable difference in the
starch-triiodide spectra for different starch types (Fig. 2).

>Fig. 2<

The difference is noticeable in the wavelength area of the starch-triiodide complex where peaks vary
in their heights and maximum wavelength values.

This was supported by the literature data (14), where the concentration of triiodide ion is consumed by
the inner cavities of helical structures of amylose and amyllopectin. The inclusion complex of amylose
with triiodide exhibits the adsorption spectra with the maximum wavelengths peak at wavelengths
higher than 620 nm. On the other hand, amyllopectin inclusion complex with triiodide ion exhibits the
adsorption spectra with the maximum wavelengths peak at lower wavelengths at approximately 540
nm, and interferes with the amylose determination colorimetric methods. (12,13)

The parameters presented in Table 2: maximum peak wavelength (λmax/nm), change of absorbance
value at maximum peak wavelength ΔA(λmax) for the same measuring condition for each starch
sample, and a maximum peak wavelength shift towards the unknown sample starch maximum peak
wavelength (δλ/nm (shift λmax)); were obtained from the starch-triiodide complex absorbance spectra for each starch sample. These values have been used for raw pre-statistical starch type differentiation (Table 2).

The commercial starch of unknown origin was used as a zero for λmax comparison to other starch types (shown as δλ). Rye (Danovski) showed the highest positive shift (+71 nm), waxy maize and maize starch the highest negative (-44 and -46) and potato starch (+7) the lowest shift towards the commercial starch of unknown origin. Rye, barley and wheat varieties exhibited λmax in very narrow λ region. The absorbance increase (ΔA) at each λmax was calculated. Potato and rice starch showed the highest absorbance increase, 0.474 and 0.422, respectively. This indicates that the highest amount of triiodide is complexed with starch.

In this paper we used the usually negative influence of amylopectin (that immobilizes some triiodide ion) on amylose determination (false positive result) as an advantage to get the whole absorbance spectra and actually get the individual fingerprint of each starch sample. These fingerprint datasets were then for the first time analysed and compared by the use of the Cluster analysis.

>Table 2<

Statistical data evaluation using Cluster analysis
Clustering aims at discovering natural groupings of items revealing interesting data patterns concerning relations between items (29,30). A matrix of similarities between items is used to define groups of items. In this research we use the most common Euclidean distance between items as a measure of similarity. Clustering methods are divided in two groups: hierarchical clustering and non-hierarchical clustering methods. Further, hierarchical methods can be agglomerative or divisive. Clustering by agglomerative methods starts with the individual item and groups the most similar items until all objects are contained in the same cluster, while divisive hierarchical methods work in the opposite direction: an initial single group of items is divided into two subgroups such that the items in two groups are dissimilar. The result of hierarchical clustering is displayed in a form of a diagram called dendrogram. The way of merging clusters together in agglomerative hierarchical method is determined by linkage method. The most common methods are single linkage (groups are merged according the distance between their nearest members of groups), complete linkage (groups are merged according to distance between their farthest members) and average distance (groups are merged according to average distance between their members). In this paper we will use agglomerative hierarchical method with complete linkage as a linkage method. The advantage of complete linkage method is that it is efficient in discerning poorly separated clusters.
In addition to visual inspections of starch spectra and pre-statistical raw starch type differentiation (Table 2), we examined the spectra by the Cluster analysis method. Although we used complete visible spectra of the starch-triiodide complex (fingerprints) for all starch increments, only the data for starch volume increments of 1.25 mL allowed starches to be distinguished using Cluster analysis, and this concentration data were used for further evaluation.

On Fig. 3 is shown dendrogram of hierarchical clustering of average values (centroids) for starch samples. Here we can observe that starches are divided in two main clusters (Cluster 1 and 2). The first cluster consists of three smaller clusters: first two (Cluster 1.1 and 1.2) contain cereals (wheat, rye and barley) with model starch (US) attached to these two clusters, while third cluster (Cluster 1.3) contains maize starches. The second cluster contains potato, tapioca and rice. Sources from the second cluster seem to be well separated from others.

The unknown starch (US) sample was placed within the Cluster 1, but it was separated from Clusters 1.1 and 1.2, but is also significantly separated from Cluster 1.3. The reason for this could be explained through the origin of the unknown origin starch. On our efforts, the producer provided us the information that this starch (US) was a wheat origin starch but slightly debranched by the enzymes. When observing the position of the US in Cluster 1, this makes sense, it is close enough to cereal type starches in Cluster 1.1. and 1.2. and still quite distinguished from maize type starches in Cluster 1.3.

On Fig. 4 is shown dendrogram of hierarchical clustering of samples belonging to starches from the first two sub-clusters (Cluster 1.1. and 1.2.) of the first cluster in dendrogram shown on Fig. 3 with source of model starch (US) which is attached to these sub-clusters. Each source is presented by three samples. Generally speaking, it can be observed that samples from the same species (origin) but different cultivars mainly are well grouped which confirms our hypothesis that the proposed method, using the Vis spectra (dependant on amylose/amylopectin ratio) as a starch fingerprint and their evaluation with Cluster analysis, is capable to distinguish starches based on their origin but also based on their cultivars. As explained in previous part the position of US starch is distanced and makes a separate subcluster, but it is still positioned in the cereal type Cluster 1 due to its natural origin.

On Fig. 5 is shown dendrogram of hierarchical clustering of samples of sources contained in the third sub-cluster of the first cluster shown on Fig. 3 (maize and waxy maize). The maize samples were distinguished due to their unique Vis spectra fingerprint, which is moved to lower wavelengths because of their high amylopectin nature. But it is interesting to observe the position of "maize" Cluster 1.3, which is still within the "cereal" Cluster 1.
On opposite, (Fig. 6) hierarchical clustering of samples of sources contained in the second cluster on Fig. 3 (potato – pure sigma, tapioca and rice) are totally different and separated in their own cluster based on their unique Vis spectra fingerprint. Tapioca and potato have indeed different properties compared to cereal type of starches, and rice starch granules are the smallest known to exist in cereal grains (31). This shows that proposed method is very certain in recognizing starch samples from observed sources.

The observed distance in different starch spectra separated in clusters (see Fig. 2) were in agreement with our assumption about the Vis absorbance spectra fingerprint is a result of an origin-dependent amylose–amylopectin ratio (17,18) and specific triiodide binding to different starches (9), also shown in Table 1. Thus, specific differences in the VIS absorbance spectra of starch-triiodide (fingerprint) could be exploited to distinguish starches based on their botanical origins (Fig. 3). It should be noted that among all starch samples, the absorption spectra of potato, tapioca and rice are remarkably different. This difference can be explained by the different amylose–amylopectin ratios, degrees of polymerization, different helical structures, granule size and other physical and chemical properties of starch. The presented method offers a potential impact in artificial food determination and also in food processing industry for starch origin control and increase in quality of final product.

CONCLUSION
The presented research successfully combined Vis spectrophotometric data obtained from measuring the starch-triiodide complex (as an unique fingerprint of each starch sample), and statistical Cluster analysis, to distinguish the starch samples based on different origin and among different cultivars. The proposed methodology uses simple instrumentation, is more convenient and easier to use than the standard microscopic methods, and provides an alternative to our recently developed electrochemical methods. In the future work more comprehensive research on testing the performance of proposed method will be conducted. The method will be tested by supervised machine learning method of automatic classification of data samples and compared to other methods for recognition of starch source. The presented method offers a potential impact in artificial food determination and also in food processing industry for starch origin control and increase in quality of final product, but potentially, also in botanic determinations and origin relations of different species.

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.
ORCID ID
N. Sakač https://orcid.org/0000-0001-5515-2368
M. Karnaš https://orcid.org/0000-0001-5742-0012
J. Dobša https://orcid.org/0000-0002-1684-1010
M. Jozanović https://orcid.org/0000-0002-9509-8105
V. Gvozdić https://orcid.org/0000-0002-1412-3701
E. Kovač-Andrić https://orcid.org/0000-0001-7473-5964
M. Kraševac Sakač https://orcid.org/0000-0001-7292-2379
B. Šarkanj https://orcid.org/0000-0002-1567-6455

REFERENCES


https://doi.org/10.1080/07408170500232784

Table 1. Iodine affinity, amylose and amylopectin content in starch from various origins. (adapted from 9, 14, 19)

<table>
<thead>
<tr>
<th>Starch</th>
<th>Iodine affinity</th>
<th>Amylose content (%)</th>
<th>Amylopectin content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>4</td>
<td>17.5</td>
<td>82.5</td>
</tr>
<tr>
<td>Wheat</td>
<td>4.86</td>
<td>21.7</td>
<td>78.3</td>
</tr>
<tr>
<td>Barley</td>
<td>6.08</td>
<td>27.5</td>
<td>72.5</td>
</tr>
<tr>
<td>Maize</td>
<td>5.18</td>
<td>21.5</td>
<td>78.5</td>
</tr>
<tr>
<td>Potato</td>
<td>4.44</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>Sweet potato</td>
<td>4.18</td>
<td>20.7</td>
<td>79.3</td>
</tr>
<tr>
<td>Chestnut</td>
<td>4.32</td>
<td>21.6</td>
<td>78.4</td>
</tr>
<tr>
<td>Sago</td>
<td>5.16</td>
<td>25.8</td>
<td>74.2</td>
</tr>
<tr>
<td>Rye</td>
<td>N.A.</td>
<td>31.1</td>
<td>68.9</td>
</tr>
<tr>
<td>Tapioca</td>
<td>N.A.</td>
<td>19.7</td>
<td>80.3</td>
</tr>
</tbody>
</table>
Table 2. Measured parameters from the absorbance spectra for all starch types measured. The commercial starch of unknown origin was used as a zero for comparison.

<table>
<thead>
<tr>
<th>Starch origin (cultivars)</th>
<th>$\lambda_{\text{max}}$/nm*</th>
<th>$\Delta \bar{A}(\lambda_{\text{max}})$* for 1.25mL starch increment</th>
<th>$\Delta \lambda$/nm (shift $\lambda_{\text{max}}$)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starch of unknown origin</td>
<td>564</td>
<td>0.337</td>
<td>0</td>
</tr>
<tr>
<td>Potato</td>
<td>571</td>
<td>0.474</td>
<td>+7</td>
</tr>
<tr>
<td>Maize</td>
<td>510</td>
<td>0.221</td>
<td>-46</td>
</tr>
<tr>
<td>Waxy maize</td>
<td>520</td>
<td>0.155</td>
<td>-44</td>
</tr>
<tr>
<td>Tapioca</td>
<td>600</td>
<td>0.398</td>
<td>+36</td>
</tr>
<tr>
<td>Rye (Eho)</td>
<td>630</td>
<td>0.310</td>
<td>+66</td>
</tr>
<tr>
<td>Rye (Danovski)</td>
<td>635</td>
<td>0.318</td>
<td>+71</td>
</tr>
<tr>
<td>Rye (Conduct)</td>
<td>623</td>
<td>0.255</td>
<td>+59</td>
</tr>
<tr>
<td>Wheat (Golubica)</td>
<td>633</td>
<td>0.312</td>
<td>+69</td>
</tr>
<tr>
<td>Wheat (Srpanjka)</td>
<td>630</td>
<td>0.197</td>
<td>+66</td>
</tr>
<tr>
<td>Wheat</td>
<td>632</td>
<td>0.270</td>
<td>+68</td>
</tr>
<tr>
<td>Barley (Barun)</td>
<td>632</td>
<td>0.326</td>
<td>+68</td>
</tr>
<tr>
<td>Barley (Vanessa)</td>
<td>623</td>
<td>0.246</td>
<td>+59</td>
</tr>
<tr>
<td>Rice</td>
<td>590</td>
<td>0.422</td>
<td>+26</td>
</tr>
</tbody>
</table>

* Average values for 5 measurements
Fig. 1. Absorption spectra for 0 - 1.25 mL commercial starch of unknown origin increments and a) starch-triodide complex peak, b) and c) triiodide peaks.
Fig. 2. Absorption spectra for starches from different botanical origins. The arrow represents a vertical separation (at a given wavelength) for a more straightforward presentation of starch absorption lines. From bottom: waxy maize, maize, wheat (Srpanjka), rye (Conduct), wheat, barley (Vanessa), rye (Eho), rye (Danovski), barley (Barun), wheat (Golubica), commercial starch of unknown origin, tapioca, rice, potato.

Fig. 3. A dendrogram of hierarchical clustering of centroids for starch samples belonging to the same source. Linkage distance is shown relatively according to maximal linkage distance between the groups. Here and in later figures abbreviations for starches from different origin are used for clarity: Barley Barun (BB), barley Vanessa (BV), rye Eho (RE), rye Danovski (RD), rye Conduct (RC), wheat Golubica (WG), wheat (W), wheat Srpanjka (WS), waxy maize (MW), maize (M), potato (PS), tapioca (T), rice (R) and commercial starch of unknown origin (US).
Fig. 4. A dendrogram of hierarchical clustering of samples belonging to sources in the first two sub-clusters of the first cluster on dendrogram shown on Figure 3 (Cluster 1.1 and 1.2). Samples of commercial starch of unknown origin are attached to these sub-clusters. Abbreviations: barley Barun (BB), barley Vanessa (BV), rye Eho (RE), rye Danovski (RD), rye Conduct (RC), wheat Golubica (WG), wheat (W), wheat Srpanjka (WS) and commercial starch of unknown origin (US).
Fig. 5. A dendrogram of hierarchical clustering of samples belonging to sources in the second sub-cluster of the first cluster on dendrogram shown on Figure 3 (Cluster 1.3). Abbreviations: waxy maize (MW) and maize (M).

Fig. 6. A dendrogram of hierarchical clustering of samples belonging to sources in the second cluster of dendrogram shown on Figure 3 (Cluster 2). Abbreviations: potato (PS), tapioca (T) and rice (R).