Food Chemistry 173 (2015) 482–488

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Feasibility in multispectral imaging for predicting the content of bioactive compounds in intact tomato fruit

Changhong Liu [a], Wei Liu [b], Wei Chen [a], Jianbo Yang [c,*], Lei Zheng [a,d,*]

[a] School of Biotechnology and Food Engineering, Hefei University of Technology, Hefei 230009, China
[b] Intelligent Control and Compute Vision Lab, Hefei University, Hefei 230601, China
[c] Rice Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230031, China
[d] School of Medical Engineering, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Article history:
Received 29 June 2014
Received in revised form 7 October 2014
Accepted 9 October 2014
Available online 18 October 2014

Keywords:
Multispectral imaging
Tomato fruit
Bioactive compounds
Non-destructive determination
Chemometric

ABSTRACT

Tomato is an important health-stimulating fruit because of the antioxidant properties of its main bioactive compounds, dominantly lycopene and phenolic compounds. Nowadays, product differentiation in the fruit market requires an accurate evaluation of these value-added compounds. An experiment was conducted to simultaneously and non-destructively measure lycopene and phenolic compounds content in intact tomatoes using multispectral imaging combined with chemometric methods. Partial least squares (PLS), least squares-support vector machines (LS-SVM) and back propagation neural network (BPNN) were applied to develop quantitative models. Compared with PLS and LS-SVM, BPNN model considerably improved the performance with coefficient of determination in prediction (R_P^2) = 0.938 and 0.965, residual predictive deviation (RPD) = 4.590 and 9.335 for lycopene and total phenolics content prediction, respectively. It is concluded that multispectral imaging is an attractive alternative to the standard methods for determination of bioactive compounds content in intact tomatoes, providing a useful platform for infield fruit sorting/grading.

1. Introduction

Tomato fruit (*Solanum lycopersicum*) is the second most important vegetable crop worldwide and consumed either fresh or in the form of processed products (Viuda-Martos et al., 2013). It is well known as a health-stimulating fruit because of the antioxidant properties of its main bioactive compounds. The most important bioactive compounds in tomato fruit are lycopene and phenolic compounds (Canene-Adams, Campbell, Zaripheh, Jeffery, & Erdman, 2005; Lenucci, Cadinu, Taurino, Piro, & Dalessandro, 2006). Lycopene, exhibiting the highest antioxidant activity and singlet oxygen quenching ability of all dietary carotenoids (Gärtner, Stahl, & Sies, 1997), is the most abundant carotenoid in ripe tomato fruit, comprising approximately 80–90% of the total carotenoids content (Shi & Le Maguer, 2000). Phenolic compounds are famous group of secondary metabolites in plants, which are vital determinants in the nutrition and sensory quality in fruits. Vinson et al. (1998) studied the amount of phenolics in commonly consumed vegetables and, on the basis of their average consumption, tomatoes were identified as the most important suppliers of phenolics in human diet. The main phenolic compounds in tomato fruit are namely the flavonoids rutin (quercetin-3-rutinoside), kaempferol-3-*O*-rutinoside, and naringenin chalcone (Long et al., 2006) and the phenolic acids caffeic acid, *p*-coumaric acid and ferulic acid (Luthria, Mukhopadhyay, & Krizek, 2006). Differentiation of the final product in the market requires an accurate evaluation of these value-added compounds. However, the conventional methods to determine these bioactive components are generally based on either colorimetric or chromatographic techniques (liquid chromatography coupled with tandem mass spectrometry, LC-MS; high-performance liquid chromatography, HPLC). These techniques are mostly time-consuming, require preparation of sample and the use of chemical products, and they are laborious and expensive. Moreover, since conventional methods are destructive, they only enable quality control of a few samples per batch, rather than each individual fruit. In order to overcome these disadvantages, developing a rapid, non-contaminant and non-destructive determination methods, preferably based on optical properties, is urgently required.

At present, various optical/electromagnetic methods have been developed for rapid, accurate, and non-destructive determination of lycopene content in tomato fruit, which included visible-near infrared (vis–NIR) spectroscopy (Clément, Dorais, & Vernon,

* Corresponding authors at: School of Biotechnology and Food Engineering, Hefei University of Technology, Hefei 230009, China. Tel.: +86 551 62919398 (L. Zheng).
E-mail addresses: yjianbo@263.net (J. Yang), lzheng@hfut.edu.cn, lei.zheng@aliyun.com (L. Zheng).

http://dx.doi.org/10.1016/j.foodchem.2014.10.052
0308-8146/© 2014 Elsevier Ltd. All rights reserved.

2008a), hyperspectral reflectance imaging (Polder, van der Heijden, van der Voet, & Young, 2004), attenuated total reflection infrared (ATR-IR) spectroscopy (Baranska, Schütze, & Schulz, 2006), surface colour measurement (Arias, Lee, Logendra, & Janes, 2000), Raman chemical imaging (Qin, Chao, & Kim, 2011), and magnetic resonance imaging (Cheng, Wang, Chen, & Lin, 2011). In addition, near infrared (NIR) reflectance spectroscopy has been used for the determination of total phenolic content in many other fruits or vegetables, such as blueberries (Sinelli, Spinardi, Di Egidio, Mignani, & Casiraghi, 2008), and grapes (Ferrer-Gallego, Hernández-Hierro, Rivas-Gonzalo, & Escribano-Bailón, 2011). Regarding the chemometric techniques, most of these methods were developed using partial least square (PLS) analysis. New chemometric analysis such as least square-support vector machine (LS-SVM) and back propagation neural network (BPNN) appear promising in that they enable the non-linearity of data to be modelled using local or specific equations which could improve prediction models.

Multispectral imaging is an emerging non-destructive technology that integrates conventional imaging and spectroscopy to obtain both spatial and spectral information from an object simultaneously. Multispectral imaging analyses are non-destructive, rapid, simple to perform, and require no sample pre-treatment, which makes this technology ideally suited for on-line process monitoring and quality control (Feng & Sun, 2012; Gowen, O'Donnell, Cullen, Downey, & Frias, 2007). More importantly, this technique has the great potential to measure the multiple components at the same time for quality assurance. Recently, this technology has been applied as a powerful process analytical tool for rapid, non-destructive inspection of internal and external attributes in various fruits and vegetables such as apple (Lu, 2004; Lunadei, Galleguillos, Diezma, Lleo, & Ruiz-Garcia, 2011; Peng & Lu, 2006, 2007), peach (Lleó, Barreiro, Ruiz-Altisent, & Herrero, 2009), fresh-cut spinach leaves (Lunadei et al., 2012), packaged wild rocket (Løkke, Seefeldt, Skov, & Edelenbos, 2013), and strawberry (Liu et al., 2014). In regards of tomato fruit, multispectral imaging was originally applied to prediction of unripe tomatoes and an accuracy of over 85% was achieved (Hahn, 2002). And it has also recently emerged as a powerful approach for identification of cherry-tomato varieties (Yang, Nie, Feng, He, & Chen, 2010). However, to our knowledge, there is no published data on the multispectral imaging for determination of the content of bioactive compounds (lycopene and total phenolics) in tomato fruit. Therefore, the main objective of this study was to assess the application of multispectral imaging for predicting the content of lycopene and total phenolics in tomato fruit, and compare the performance of prediction models obtained using PLS, LS-SVM and BPNN.

2. Materials and methods

2.1. Tomato fruit sampling

A total of 162 tomato fruit (*S. lycopersicum*, cv. Wanza 15) were harvested at the following maturity stages: mature green, breakers, turning, pink, light red and red according to the *Standards for Grades of Fresh Tomatoes* (USDA, 1991), in a commercial greenhouse in Hefei City, China in July 2013 and transported to our laboratory within 1 h after harvest. For each stage, 27 fruit with uniform size, weight, disease free and no other defects were selected and washed with tap water, air-dried at ambient temperature, and then analysed by a multispectral imaging system.

2.2. Spectra collection

The multispectral imaging analysis was performed using a VideometerLab equipment (Videometer A/S, Hørsholm, Denmark) which acquires multispectral images at 19 different wavelengths in the ultra-violet A (UVA), visual (VIS) and the lower wavelengths of the NIR region: 405, 435, 450, 470, 505, 525, 570, 590, 630, 645, 660, 700, 780, 850, 870, 890, 910, 940 and 970 nm. The acquisition system records surface reflections with a standard monochrome charge coupled device chip, nested in a Point Grey Scorpion camera. The object of interest is placed inside an integrating sphere with a matte white coating to ensure that the light is scattered evenly with a uniform, diffuse light at illumination. At the rim of the sphere light emitting diodes (LEDs) are positioned side by side in a pattern which distributes the LEDs belonging to each wavelength uniformly around the entire rim. The LEDs are strobing successively, resulting in an image for each LED of dimensionality 2056×2056. The system is first calibrated radiometrically using both a diffuse white and dark target followed by a light setup based on the type of object to be recorded. The system is geometrically calibrated with a geometric target to ensure pixel correspondence for all spectral bands. Segmenting images into distinct regions is an important pre-processing step in image analysis. Image segmentation was performed using the VideometerLab software version 2.12.23. To remove the image background, all items except the tomato fruit were removed by a Canonical Discriminant Analysis (CDA) (Cruz-Castillo et al., 1994) and segmented using a simple threshold. The image of tomato fruit sample without the background could be transformed to spectra based on a mean calculation. Thus each image contributed with a single spectrum for the model calibration. After spectra collection, the edible pericarp of the sampled fruit was excised. After complete removal of the jelly-like parenchyma, the pericarp was cut into small pieces using a sharp knife and frozen in liquid nitrogen. Frozen samples were kept at −70 °C for subsequent analysis of various bioactive components. All analyses were conducted in triplicate.

2.3. Analyses of bioactive compounds

Lycopene content was extracted and determined according to the methods of Clément, Dorais, and Vernon (2008b). Lycopene from fruit sample was extracted with hexane:ethanol:acetone (2:1:1), containing 2.5% BHT (butylated hydroxy toluene). Optical density of the hexane extract was measured spectrophotometrically at 503 nm against a hexane blank. The following relation was then used for estimation of lycopene content: lycopene (mg/kg) = $(A_{503} \times 31.2)$/(quantity of tissue used). Results were expressed as mg/kg fresh weight (FW).

Total phenolics was extracted and determined according to the methods of Toor and Savage (2005). Total phenolic contents were determined with spectrophotometer using Folin–Ciocalteu reagent and the results were expressed as gallic acid equivalents, in mg gallic acid/100 g fresh weight (FW).

2.4. Chemometrics: multivariate analysis

Multivariate analyses including PLS, LS-SVM and BPNN were used to estimate calibration models between the extracted spectra and reference values (chemically determined).

PLS is a linear regression method for multivariate calibration. It has been widely applied in fruits and vegetables analysis with favourable results. The optimal number of latent variables (LVs) was determined by the lowest value of predicted residual error sum of squares (PRESS). LVs can eliminate noises and random errors in the original data and account as much as possible for the variability of the original variables.

Support vector machine (SVM) proposed by Cortes and Vapnik (1995) is a learning algorithm and powerful for solving problems in nonlinear classification, function estimation, and density estimation. Compared with other methods, SVM does not require a