World Applied Sciences Journal 15 (10): 1465-1468, 2011
ISSN 1818-4952

Quantitative Structure-Retention Relationship Analysis of Polyphenols in Carob Fibre

[1]Hamid Karimi, [2]Abbas Farmani and [2]Hadi Nourizadeh

[1]Petroleum Engineering and Development Company (PEDEC), Ostad Najatollai Avenue, Shahid Kalantari Avenue, No. 61, Tehran, Iran
[2]Department of Chemistry, Faculty of Sciences, Arak Branch, Islamic Azad University, Arak, Iran

Abstract: Genetic algorithm and partial least square (GA-PLS) technique was used to investigate the correlation between retention time (RT) and descriptors for 24 polyphenols in carob fibre which obtained by high performance liquid chromatography (HPLC). The applied internal (leave-group-out cross validation (LGO-CV)) and external (test set) validation methods were used for the predictive power of model. It is easy to notice that there was a good prospect for the GA-PLS application in the quantitative structure-retention relationship (QSRR) modeling. This indicates that RT of essential oils possesses some linear characteristics. It can also be used successfully to estimate the RT for new compounds or for other compounds whose experimental values are unknown.

Key words: Antioxidants · Carob fibre · Polyphenols · QSRR · Genetic algorithm

INTRODUCTION

Carob is the beanlike fruit of *Ceratonia siliqua*, which grows widely in the Mediterranean region and belongs to the genus leguminosae. Historically, due to its high content of sugars, the brown pod was consumed as food, especially in ancient times as a candy for children or in emergency situations such as war. The carob pulp contains 40-60% of low-molecular-weight carbohydrates, mainly sucrose, depending on carob species, origin and climate [1]. Apart from carbohydrates, high amounts of dietary fibre and polyphenols are characteristic of this Mediterranean food. Dietary fibre itself or a diet rich in dietary fibre is known to exert a variety of physiological effects, including improved digestion and attenuation of blood cholesterol and glucose levels [2, 3]. Dietary fibre probably also has chemopreventive potential against certain cancers, in particular those of the gastro-intestinal tract [4-6].

Often the physiological effects of diets or foods rich in dietary fibre cannot be linked to a single component because of the presence of additional health-promoting constituents such as phytochemicals, among which the polyphenols can exhibit antioxidative, antimutagenic, anticarcinogenic, antiproliferative, or antioestrogenic activity.

Polyphenols are a very heterogeneous group and include simple phenolic acids, cinnamic acid and its derivatives, flavonoids, iso- flavones, lignans, anthocyans and tannins. Therefore, when included in the daily diet, these compounds may play an important role in the prevention and reduction of cancer and heart disease [7, 8].

Quantitative structure–retention relationship (QSRR) is statistically derived relationships between chromatographic parameters and descriptors related to the molecular structure of the analytes. QSRR on the retention time have been reported for several compounds [9-11].

There is a trend to develop QSRR from a variety of methods. In particular, genetic algorithm (GA) is frequently used as search algorithms for variable selection in chemometrics and QSRR. GA is a stochastic method to solve the optimization problems defined by fitness criteria, applying the evolution hypothesis of Darwin and different genetic functions, i.e. crossover and mutation [12, 13]. Partial least square (PLS) is the most commonly used multivariate calibration method [14, 15]. In the present work, a QSRR study has been carried out on the retention times (RT) for major individual polyphenols in carob fibre.

Corresponding Author: Hadi Nourizadeh, Department of Chemistry, Faculty of Sciences, Arak Branch, Islamic Azad University, Arak, Iran. Tel: + 989183452507.

MATERIALS AND METHODS

Data Set: The molecules in the data set, including the 24 major individual polyphenols in carob fibre, are shown in Table 1. Carob fibre was found to contain a rich variety of phenolic compounds and the elucidated structures of 24 isolated substances are presented in Fig. 1 [16]. Analytical HPLC was conducted on a Hewlett-Packard (HP) 1090 liquid chromatograph fitted with a C-18, reversed-phase (5 µm) column (25 cm × 4 mm I.D.; Latex, Eppelheim, Germany). Analyses were performed using a HP 5973 mass spectrometer coupled to a HP 6890 gas chromatograph. Prior to Gas chromatography coupled with mass spectrometry (GC–MS), dried methanolic extracts (10 µl) were derivatized by addition of BSTFA (100 ml) at 37°C for 30 min. Liquid-chromatography electrospray-ionisation mass spectrometry (LC-ESI) was conducted on an Agilent 1100 HPLC coupled to an Agilent LC/MSD (HP 1101). Most of the isolated phenolic compounds were identified by comparison of their UV spectra and HPLC retention times to those of reference compounds and by mass spectrometry, including Liquid chromatography-Nano-electrospray-ionisation mass spectrometry in the negative- ion mode, nano-ESI-MS in both negative- and positive-ion modes, MS-MS experiments where useful, as well as GC-MS.

Table 1: The data set and corresponding observed RT values

No		Name	RT_{Exp}
		Training Set	
1	V	Gallic acid	5.1
2	VI	Methyl gallate	10.1
3	XXII	1,6-Di-O-galloyl-β-d-glucose	11.1
4	XXIII	1,2,6-Tri-O-galloyl-β-d-glucose	13.4
5	XXIV	1,2,3,6-Tetra-O-galloyl-β-d-glucose	15.9
6	II	p-Coumaric acid	17.3
7	XIX	Ferulic acid	18.8
8	XVIII	Myricetin glucoside	19.7
9	XIII	Myricetin	21.4
10	XVI	Myricetin	24.4
11	III	Quercetin	25.4
12	XV	Kaempferol	29.3
13	I	Cinnamic acid	29.5
14	XI	Quercetin	29.7
15	XX	Naringenin	30.1
16	XII	Isorhamnetin	32.6
17	XXI	Genistein	32.7
18	XIV	Kaempferol	34.4
19	VII	Apigenin	35.1
20	IX	Tricetin 30,50 dimethyl ether	36.1
		Test Set	
21	IV	Syringic acid	14.4
22	XVII	Quercetin	24.5
23	X	Luteolin	31.8
24	VIII	Chrysoeriol	35.6

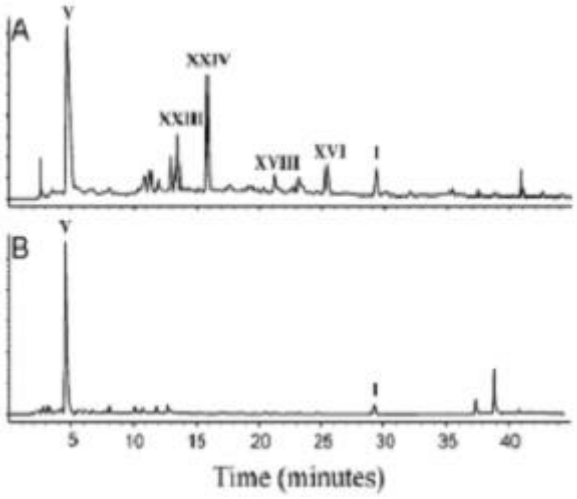

Fig. 1: Analytical HPLC chromatograms monitored by UV absorption l278 for the methanol (A) and aqueous (B) extracts of carob ?bre. Roman numerals refer to individual compounds

RESULTS AND DISCUSSION

Results of the GA-PLS Model: To reduce the original pool of descriptors to an appropriate size, the objective descriptor reduction was performed using various criteria. To reduce the original pool of descriptors to an appropriate size, objective descriptor reduction was performed using various criteria. Reducing the pool of descriptors eliminates those descriptors which contribute either no information or whose information content is redundant with that of other descriptors present in the pool. Any descriptor that had identical or zero values for greater than 90% of the compounds was eliminated. The remained descriptors were employed to generate the models with the GA-PLS program. The best model is selected on the basis of the highest square correlation coefficient (R^2) and relative error (RE) of prediction and simplicity of the model. These parameters are probably the most popular measure of how well a model fits the data [17]. The best GA-PLS model contains 16 selected descriptors in 11 latent variables space. The R^2 and RE for training and test sets were (0.811, 0.629) and (13.03, 25.59), respectively. The GA-PLS linear model has good statistical quality with low prediction error. Plots of predicted RT versus experimental RT values by GA-PLS for training and test set are shown Fig. 2. Obviously, there is a close agreement between the experimental and predicted RT and the data represent a very low scattering around a straight line with respective slope and intercept close to one and zero.

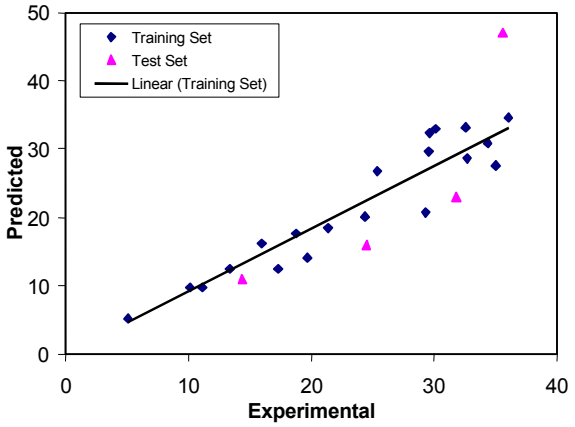

Fig. 2: Plot of predicted RT obtained by GA-PLS against the experimental values

Model Validation and Statistical Parameters: The applied internal (leave-group-out cross validation (LGO-CV)) and external (test set) validation methods were used for the predictive power of models. In the leave-group-out procedure one compound was removed from the data set, the model was trained with the remaining compounds and used to predict the discarded compound. The process was repeated for each compound in the data set. The predictive power of the models developed on the selected training set is estimated on the predicted values of test set chemicals. The data set should be divided into three new sub-data sets, one for calibration and prediction (training) and the other one for testing. The calibration set was used for model generation. The prediction set was applied deal with overfitting of the network, whereas test set which its molecules have no role in model building was used for the evaluation of the predictive ability of the models for external set. In the other hand by means of training set, the best model is found and then, the prediction power of it is checked by test set, as an external data set. In this work, randomly in each running program, from all 24 components, 15 components are in calibration set, 5 components are in prediction set and 4 components are in test set.

CONCLUSION

In this research, an accurate QSRR model for estimating the retention time of major individual polyphenols in carob fibre was developed by employing the GA-PLS technique. This model has good predictive capacity and excellent statistical parameters. It is easy to notice that there was a good prospect for the GA-PLS application in the QSRR modeling. It can also be used successfully to estimate the RT for new compounds or for other compounds whose experimental values are unknown.

REFERENCES

1. Marakis, S., G. Marakis and M. Lambraki, 1997. Tannins of eight carob varieties from the island of Lefkada, Greece. Chimika Chronika, New Seriesm, 26: 57-66.
2. Nàdia Ortega, Alba Macià, Maria-Paz Romero, Jordi Reguant and Maria-Jose Motilva, 2011. Food Chemistry, 124: 65-71.
3. Vitali, D., I. Vedrina Dragojević and B. Šebečić, 2009. Food Chemistry, 114: 1462-1469.
4. Whiteley, L.O. and D.M. Klurfeld, 2000. Nutr. Cancer 36: 131-149.
5. Bingham, S.A., N.E. Day, R. Luben, P. Ferrari, N. Slimani, T. Norat, R. Clavel-Chapelon, V. Krogh, H.B. Bueno-de-Mesquita, P.H.M. Peeters, G. Berglund, G. Hallmans, E. Lund, G. Skeie, R. Kaaks and E. Riboli, 2003. Lancet, 361: 1496-1501.
6. Peters, U., R. Sinha, N. Chatterjee, A.F. Subar, R.G. Ziegler, M. Kulldorff, R. Bresalier, J.L. Weissfeld, A. Flood, A. Schatzkin and R.B. Hayes, 2003. Lancet, 361: 1491-1495.
7. Yamagishi, M., M. Natsume, A. Nagaki, T. Adachi, N. Osakabe, T. Takizawa, H. Kumon and T. Osawa, 2000. J. Agric. Food. Chem., 48: 5074-5078.
8. Manthey, J.A. and N. Guthrie, 2002. J. Agric. Food. Chem., 50: 5837-5843.
9. Chen, J., T. Yang and S.M.J. Cramer, 2008. Chromatogr. A., 1177: 207.
10. Noorizadeh, H. and A. Farmany, 2010. Chromatographia, 72: 563.
11. Hemmateenejad, B., K. Javadnia and M. Elyasi, 2007. Anal. Chim. Act., 592: 72.
12. Karimi, H., A. Farmany and H. Noorizadeh, 2011. World Appl. Sci. J., 15(8): 1086-1088.
13 Aires-de-Sousa, J., M.C. Hemmer and J. Casteiger, 2002. Prediction of H-1 NMR chemical shifts using neural networks, Anal. Chem., 74: 80-90.
14. Riahi, S., E. Pourbasheer, M.R. Ganjali and P.J. Norouzi, 2009. Haz. Mat., 166: 853.
15. Bodzioch, K., A. Durand, R. Kaliszan, T. Bączek, T. Heyden and Y.V. Talanta, 2010, 81: 1711.
16. Niazi, A., S. Jameh-Bozorghi and D.J. Nori-Shargh, 2008. Hazard. Mater., 151: 603.
14. Woo, S.H., O. Jeon, Ch., Y.S. Yun, H. Choi, S. Lee, Ch., D.S.J. Lee, 2009. Hazard. Mater., 161: 538.

15. Krämer, N., A.L. Boulesteix and G. Tutz, 2008. Chemom. Intell. Lab. Syst., 94: 60.
16. Owen, R.W., R. Haubner, W.E. Hull, G. Erben, B. Spiegelhalder, H. Bartsch and B. Haber, 2003. Food Chem. Toxicol., 41: 1727-1738.
17. Noorizadeh, H. and A. Farmany, 2011. World Appl. Sci. J., 14(4): 603-606.