International Journal of Biomedical and Clinical Sciences
Vol. 5, No. 4, 2020, pp. 382-387
http://www.aiscience.org/journal/ijbcs

Analysis of Glycated Hemoglobin by Near-Infrared Spectroscopy

Yun Han, Huihui Zhou*

Department of Data Science, Guangdong Ocean University, Zhanjiang, China

Abstract

In this study, the relative indicator glycated hemoglobin (HbA1c) was determined by near-infrared (NIR) spectroscopy in human hemolysate samples. Because HbA1c is a percentage indicator, it was indirectly determined via Hb and Hb•HbA1c (absolute HbA1c content). Equidistant combination multiple linear regression (EC-MLR) and moving window partial least squares (MW-PLS) methods were employed for screening key wavelengths. Using the EC-MLR, 6 and 14 wavelengths were selected for Hb and Hb•HbA1c, respectively. Using the MW-PLS, wavebands 940–1750 nm and 1492–1858 nm were selected for Hb and Hb•HbA1c, respectively. The EC-MLR method adopted fewer wavelengths. The HbA1c predicted values were further calculated by predicted values of Hb and Hb•HbA1c. The obtained root mean square error and correlation coefficients of prediction (V_SEP, V_R_P) for HbA1c in validation set were 0.49% and 0.909 with EC-MLR method and 0.41% and 0.919 with MW-PLS method. Both methods achieved good prediction results. The results show that the strategy of measuring relative indicator HbA1c by NIR is feasible, which provides a wider application space for NIR spectroscopy. In addition, the technique is fast and simple compared to traditional methods, so it is a promising tool for screening diabetes in large populations.

Keywords

Glycated Hemoglobin, Near-Infrared Spectroscopy, Wavelength Selection

Received: September 4, 2020 / Accepted: September 27, 2020 / Published online: November 23, 2020

1. Introduction

Diabetes is a chronic metabolic disease characterized by hyperglycemia due to defects in insulin secretion or impaired insulin action, which seriously endangers human health. Therefore, it is of great significance to actively explore new treatment methods and strengthen preventive measures for diabetes. Studies have shown that good blood glucose control can delay the occurrence of diabetes and its complications. Glycosylated hemoglobin (HbA1c) is the “gold standard” for evaluating long-term blood glucose levels. It is not easily affected by other factors (eating, medication, etc.) and can objectively reflect the state of hyperglycemia. Therefore, the rapid and accurate detection of HbA1c is important for diabetes screening, diagnosis and treatment.

HbA1c is formed via the non-enzymatic glycation of glucose and hemoglobin in the blood. This process is irreversible and involves a series of Maillard reactions [1]. The HbA1c value is a relative percentage, equaling to the ratio of the HbA1c absolute content to total Hb. The clinical cut-off value of HbA1c is 6.0%, that is, HbA1c >6.0% is a diabetic phenotype positive patient [2].

Near-infrared (NIR) spectroscopy is mainly produced by frequency doubling and combination absorption of molecular vibration, and only the stretching vibration of hydrogen-containing functional groups such as C-H, N-H, S-H, and O-H can be detected. This technique can fast and directly detect samples without reagents. The molecules of glucose and Hb contain hydrogen-containing functional groups that have marked absorption in the NIR region, so NIR spectroscopy was used to analyze glucose and Hb in many previous studies [3–9]. The process of Hb saccharification

* Corresponding author
E-mail address: 15702096261@163.com (Yun Han), 352907520@qq.com (Huihui Zhou)

(Maillard reactions) includes some hydrogen-containing functional groups; therefore, NIR spectroscopy has the theoretical basis for obtaining the information about HbA1c.

The HbA1c value and its spectral absorption value do not conform to Beer's law because it is a percentage. Therefore, HbA1c cannot be directly analyzed by NIR spectroscopy. As a result, there are few related studies on HbA1c detection by NIR spectroscopy. Indirect determination of HbA1c is considered in this study. Since Hb can be determined by NIR [6–9], we judge Hb•HbA1c can be determined with the same method. The predicted values of HbA1c can be obtained by simple calculation of the predicted values of Hb and Hb•HbA1c.

Since human blood is a multi-component system, when one component is analyzed by NIR spectroscopy, it will be interfered by the noise of other components. Thus, wavelength selection is essential for the rapid and accurate measurement of the human blood with NIR spectroscopy. The moving window PLS (MW-PLS) and the equidistant combination multiple linear regressions (EC-MLR) are widely-used and well-performed wavelength selection methods [9–15], so both of them were used for selecting information wavebands and wavelength combinations. Furthermore, Savitzky–Golay (SG) smoothing, as an efficient spectral preprocessing method, was employed for the spectral data pretreatment.

2. Materials and Methods

2.1. Experiment

The experimental study was done on 240 cases of participants who provide informed consent. Experiments were conducted in accordance with relevant laws and institutional guidelines and approved by local medical institutions. The reference values in modeling and validation process of Hb and HbA1c for these samples were obtained by traditional methods. Table 1 is the analysis of the reference values of these samples. Based on the HbA1c cut-off value (6.0%), 104 negative and 136 positive samples were obtained. Considering that the peripheral blood is thick with high noise interference, 2 × dilute hemolytic solutions were adopted.

Table 1. Analysis of reference values of Hb and HbA1c for 240 samples.

Sample types	Number of samples	Hb (g L^{-1})				HbA1c (%)			
		Min	Max	Mean	SD	Min	Max	Mean	SD
All samples	240	77	165	130.8	14.2	4.6	10.8	6.32	0.98
Negative	104	78	165	128.0	14.2	4.6	6.0	5.57	0.35
Positive	136	77	156	133.0	13.9	6.1	10.8	6.90	0.92

Note: SD is the abbreviations of standard deviation

Spectrum was collected with XDS Rapid Content™ Liquid Grating Spectrometer (FOSS, Denmark). The spectral scanning range was 780–2498 nm, and the wavelength interval was 2 nm. The temperature and relative humidity of the laboratory were 25±1°C and 46±1%RH, respectively. Each sample was measured three times, and the mean value of the three measurements was used for modeling.

2.2. Sample Set Division

All samples were divided into two parts, one was called modeling set, the other was called validation set, and then the modeling set were further divided into calibration and prediction sets multiple times so as to avoid the contingency of the results of a single division. Calibration and prediction processes were conducted for each division and the optimal model were determined based on the mean value of predicted root mean square error for all divisions. Finally, the optimal models were revalidated against the validation samples. It should be noted that the validation samples were randomly selected and did not participate in the modeling process. Figure 1 is frame diagram of sample set division, which shows the number of samples and type of calibration, prediction, and validation sets.

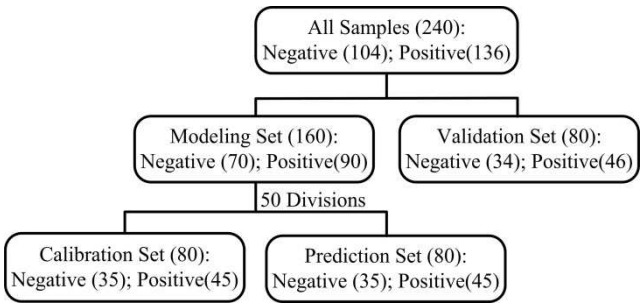

Figure 1. Frame diagram of sample set division.

Calibration and prediction models were established for each division i, $i=1, 2, \cdots, 50$. The root mean square errors for calibration and prediction for modeling set are respectively denoted as M_SEC$_i$ and M_SEP$_i$, and the corresponding correlation coefficients are respectively denoted as M_$R_{C,i}$ and M_$R_{P,i}$. The mean value and standard deviation of the M_SEP$_i$ and M_$R_{P,i}$ of all divisions were further calculated and respectively denoted as M_SEP$_{Ave}$, M_$R_{P,Ave}$, M_SEP$_{SD}$, and M_$R_{P,SD}$. These values were used to evaluate prediction effect of the model. The root mean square error and correlation coefficients of prediction for validation samples were then calculated and respectively denoted as V_SEP and V_R_P.

Hb and Hb•HbA1c were independently quantified through the same process. The predicted value of HbA1c can be obtained indirectly via simultaneous determination of the above two indicators.

2.3. EC-MLR Method

EC-MLR is a low key and practical method of selecting equidistant discrete wavelengths. For the method, the spectral data of N equidistant discrete wavelengths are designated as a window. By moving and changing the size of the window, all combinations of equidistant discrete wavelengths can be obtained. The MLR models are established for all combinations, and the optimal wavelengths combination is selected accroding to the prediction effect. The parameters of the method are as follows: (1) beginning wavelength (B), (2) number of wavelengths (N), and (3) number of wavelength gaps (G) [15]. The search range of the method can be full spectrum or partial spectrum.

The search range for this study was the full spectrum (780–2498 nm) with 860 wavelengths. After repeated computer experiments, parameters B, N, and G were set to be $B \in \{780, 782, \cdots, 2498\}$, $N \in \{1, 2, \cdots, 50\}$, and $G \in \{1, 2, \cdots, 100\}$, respectively. For each combination (B, N, G) of all divisions of calibration and prediction sets, MLR model was established, and then the corresponding M_SEP$_{Ave}$, M_$R_{P,Ave}$, M_SEP$_{SD}$, and M_$R_{P,SD}$ values were calculated. Finally, the optimal wavelength combination was screened out based on minimum M_SEP$_{Ave}$. Figure 2 is a schematic diagram of the moving window screening for equidistant wavelength combination when (B, N, G) = (1554, 14, 9).

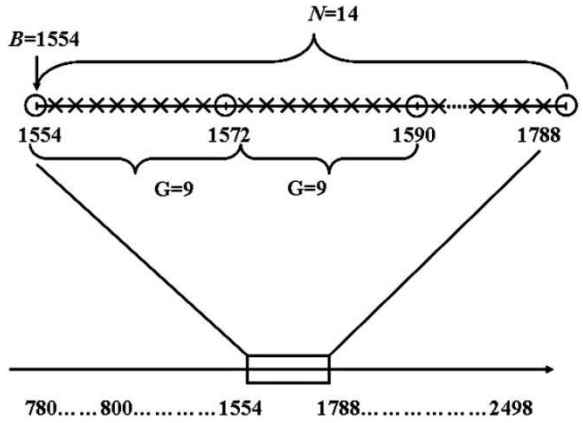

Figure 2. Sketch map for equidistant wavelength combinations and screening modes of the moving window.

2.4. MW-PLS Method

The MW-PLS is an effective variables selection method with high prediction ability for NIR analysis. For the method, N consecutive spectral data are designated as a window. The PLS models are established for all windows in the predetermined search area of the spectrum, and information wavebands are screened via moving and changing the size of the window. Considering the location and length of the waveband and the PLS factor are variable, the search parameters are set as follows: (1) beginning wavelength (B), (2) number of wavelengths (N), and (3) number of PLS factors (F) [12–15]. The search range of parameters B, N, and F can be determined according to the actual chemical and physical properties. For all divisions, PLS model was established for each combination of (B, N, F). The corresponding M_SEP$_{Ave}$, M_$R_{P,Ave}$, M_SEP$_{SD}$, and M_$R_{P,SD}$ values were then calculated. The optimal waveband with minimum M_SEP$_{Ave}$ was selected to achieve stable results.

The search range of MW-PLS was also full spectrum. Further, taking into account both the workload reduction and the representativeness, B and F were set to be $B \in \{780, 782, \cdots, 2498\}$ and $F \in \{1, 2, \cdots, 30\}$ for Hb and Hb•HbA1c. N were respectively set to be $N_{Hb} \in \{1, 2, \cdots, 450\} \cup \{460, 470, \cdots, 860\}$ and $N_{Hb•HbA1c} \in \{1, 2, \cdots, 200\} \cup \{215, 230, \cdots, 860\}$.

The algorithm codes of the two methods mentioned above were written using MATLAB 7.6.

3. Results and Discussion

3.1. Results with MW-PLS Method

In order to show the importance of wavelength selection in NIR spectroscopy, the PLS models of Hb and Hb•HbA1c were first established in full spectrum. The SG smoothing model with first-order derivative, third-degree polynomial and 13 smoothing points was used in the PLS models for reducing the interference of water or other factors on the modelling. The prediction results are shown in Table 2. As can be seen from the results, the Hb predicted values are highly correlated with reference values, but the correlation between the predicted values and reference values of Hb•HbA1c is poor. In addition, the number of wavelengths used was 860, and the model was complex.

Table 2. Prediction results of PLS models in the whole scanning area (780–2498 nm) for Hb and Hb•HbA1c.

Indicator	Waveband (nm)	N	F	M-SEP$_{Ave}$	M-SEP$_{SD}$	M-$R_{P,Ave}$	M-$R_{P,SD}$
Hb	780-2498	860	6	4.2	0.3	0.962	0.005
Hb•HbA1c	780-2498	860	8	1.27	0.08	0.699	0.038

To achieve higher model prediction accuracy with fewer wavelengths, the MW-PLS method was used for further waveband optimization. According to the minimum M_SEP_{Ave} value, the optimal MW-PLS models were determined for Hb and Hb•HbA1c. Table 3 summarizes the corresponding prediction effects (M_SEP_{Ave}, $M\_R_{P,Ave}$, M_SEP_{SD}, and $M\_R_{P,SD}$). The results show that the optimal values of *B* and *N* are 940 nm and 406 for Hb and 1492 nm and 184 for Hb•HbA1c, respectively. The corresponding waveband were 940–1750 nm for Hb and 1492–1858 nm for Hb•HbA1c, which within the NIR overtone region. The number of wavelengths for the former waveband was less than half of that for the full spectrum and for the latter one; it was less than quarter of that for the full spectrum. Therefore, the complexity of model was considerably reduced. Tables 2 and 3 show that M_SEP_{Ave} values of the optimal MW-PLS models were significantly lower than those of the full spectrum for the two indicators. Therefore, the prediction accuracy and stability of the optimal MW-PLS models for both the indicators were significantly improved, specifically for Hb•HbA1c.

Table 3. Prediction results of the optimal MW-PLS models for Hb and Hb•HbA1c.

Indicator	Waveband (nm)	*B*	*N*	*F*	M_SEP_{Ave}	M_SEP_{SD}	$M\_R_{P,Ave}$	$M\_R_{P,SD}$
Hb	940-1750	940	406	5	3.9	0.2	0.968	0.004
Hb•HbA1c	1492-1858	1492	184	7	0.63	0.06	0.933	0.013

3.2. Results with EC-MLR Method

The EC-MLR method, which was able to select equidistant discrete wavelength combination, requires a relatively wide data gap to overcome the spectral colinearity of MLR. The optimal models were screened out for Hb and Hb•HbA1c according to the minimum M_SEP_{Ave} value. Table 4 summarizes the corresponding parameters *B*, *N*, and *G* and the prediction results (M_SEP_{Ave}, $M\_R_{P,Ave}$, M_SEP_{SD}, and $M\_R_{P,SD}$).

Table 4. Prediction results of the optimal EC-MLR models for Hb and Hb•HbA1c.

Indicator	Wavelength (nm)	*B*	*N*	*G*	M_SEP_{Ave}	M_SEP_{SD}	$M\_R_{P,Ave}$	$M\_R_{P,SD}$
Hb	1520, 1552, 1584, 1616, 1648, 1680	1520	6	16	3.7	0.3	0.971	0.004
Hb•HbA1c	1554, 1572, 1590, 1608, 1626, 1644, 1662, 1680, 1698, 1716, 1734, 1752, 1770, 1788	1554	14	9	0.76	0.05	0.902	0.011

As can be seen from tables 3 and 4, the prediction effects of the two methods are very close. Furthermore, compared with MW-PLS, the EC-MLR optimal model contained only 6 wavelengths for Hb and 14 wavelengths for Hb•HbA1c; thus, the complexity of the model is significantly reduced. It should be noted that for the two indicators, the optimal wavelengths combination screened out by the EC-MLR method was within the optimal band selected out by the MW-PLS method. Therefore, the spectral absorption regions screened by the two methods are consistent, indicating that the wavelength selection is reasonable.

3.3. Model Validation

In order to further test the accuracy and stability of the selected optimal models, 80 validation samples which were not involved in the modeling process were used for the corresponding experimental verification. The regression coefficients were calculated using the spectral data and reference values for entire modeling set based on the parameters of the corresponding optimal models. Then, based on the obtained regression coefficients and the spectra of validation samples, the predicted values of the validation samples could be calculated.

Table 5. Validation results of the optimal MW-PLS and EC-MLR models.

Indicator	Wavelength (nm)	*N*	V_SEP	V_R_P
MW-PLS method				
Hb	940-1750	406	3.2	0.967
Hb•HbA1c	1492-1858	184	0.62	0.948
EC-MLR method				
Hb	1520, 1552, 1584, 1616, 1648, 1680	6	3.4	0.962
Hb•HbA1c	1554, 1572, 1590, 1608, 1626, 1644, 1662, 1680, 1698, 1716, 1734, 1752, 1770, 1788	14	0.70	0.941

The verification results (V_SEP and V_R_P) for validation set are presented in Table 5. The results show that the two prediction models reach high prediction accuracy. The predicted values of Hb and Hb•HbA1c of the validation samples are close to the reference values. Satisfactory validation results were obtained for the random validation samples because stability was taken into account in the modeling optimization process.

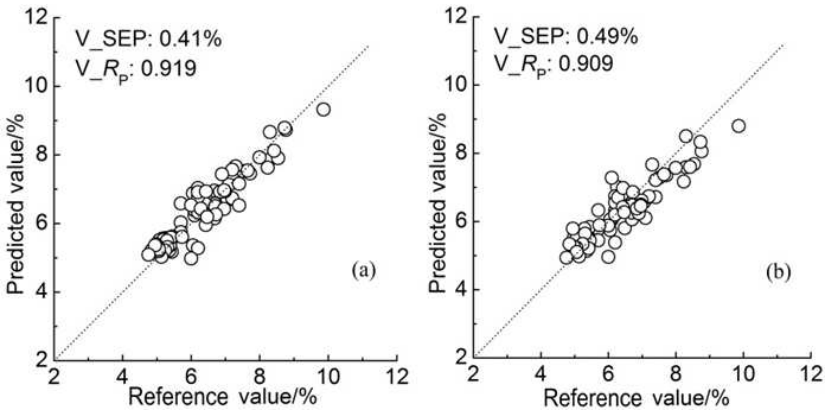

Figure 3. Correlation between the predicted value and the reference value of validation samples for HbA1c based on the optimal (a) MW-PLS and (b) EC-MLR models.

The predicted values of relative indicator HbA1c were calculated according to the predicted values of Hb and Hb•HbA1c. The corresponding V_SEP and V_R_P values of HbA1c were 0.41% and 0.919 with the MW-PLS method and 0.49% and 0.909 with the EC-MLR method, respectively. The correlation between the predicted values and the reference values for HbA1c of the 80 validation samples is shown in Figure 3. The results indicate that the predicted value and the reference value of HbA1c were also highly correlated. The experimental results confirm the feasibility of using NIR spectroscopy to quantitatively analyze HbA1c by simultaneous quantitative analysis of Hb and Hb•HbA1c.

MW-PLS and EC-PLS are two variable selection methods commonly used in NIR spectral analysis. Among them, MW-PLS selects continuous wavelengths and EC-MLR selects equidistant discrete wavelengths. The detection of relative indicator is complicated and difficult for NIR spectroscopy, so indirect analysis strategy was employed in this paper. Based on this strategy and multiple modeling, good predicated results can be obtained for relative indicator HbA1c by the commonly used variable selection method without using complicated and tedious methods. In addition, the strategy can also be applied to the analysis of relative indicators in other fields. Therefore, the analysis strategy for relative indicator is simple and practical, which provides a wider application space for NIR spectroscopy.

4. Conclusion

The relative percentage HbA1c was successfully determined by NIR spectroscopy based on simultaneous determination of absolute indicators Hb and Hb•HbA1c. In addition, the strategy of multiple modeling was adopted to generate objective and stable models. Using MW-PLS and EC-MLR methods, appropriate wavebands and wavelength combinations were screened out for Hb and Hb•HbA1c based on the multiple divisions of calibration and prediction sets. All the wavelengths selected by the two methods were within the NIR overtone region, and corresponding optimal models achieved satisfactory validation effects. In addition, the optimal wavelengths combination screened out by the EC-MLR method was within the optimal band selected out by the MW-PLS method. Thus, the spectral absorption regions selected by the two methods were consistent, which indicated that the wavelength selections were reasonable.

The results show that the indirect determination of HbA1c based on simultaneous quantitative analysis of Hb and Hb HbA1c by NIR spectroscopy is feasible. The strategy of determining the relative indicator by NIR spectroscopy can be applied to the other fields.

Acknowledgements

This work was supported by Youth Innovation Talents Project of Colleges and Universities in Guangdong Province (No. Q18285) and Guangdong Ocean University Scientific Research Start-up Funding for the Doctoral Program (No. R17057).

References

[1] Y. Han, T. Pan, HH Zhou, R. Yuan, ATR-FTIR spectroscopy with equidistant combination PLS method applied for rapid determination of glycated hemoglobin, Anal. Methods. 10 (2018) 3455-3461.

[2] American Diabetes Association, Standards of Medical Care for Patients with Diabetes Mellitus, Diabetes Care. 26 (2003) 33-50.

[3] J. T. Xue, L. M. Ye, C. Y. Li, M. X. Zhang, Rapid and nondestructive measurement of glucose in a skin tissue phantom by near-infrared spectroscopy, Optik. 170 (2018) 30-36.

[4] M. Goodarzi, W. Saeys, Selection of the most informative near infrared spectroscopy wavebands for continuous glucose monitoring in human serum, Computational and mathematical methods in medicine. 146 (2016) 155-165.

[5] J. Xie, T. Pan, J. M. Chen, et al, Joint Optimization of Savitzky-Golay Smoothing Models and Partial Least Squares Factors for Near-infrared Spectroscopic Analysis of Serum Glucose, Chinese J. Anal. Chem. 38 (2010) 342-346.

[6] Y. Zhao, L. N. Qiu, Y. L. Sun, et al, Optimal hemoglobin extinction coefficient data set for near-infrared spectroscopy, Biomedical optics express. 8 (2017) 5151-5159.

[7] H. Tian, M. Li, Y. Wang, et al, Optical wavelength selection for portable hemoglobin determination by near-infrared spectroscopy method, Infrared physics & technology. 86 (2017) 98-102.

[8] L. Pollonini, Optical properties and molar hemoglobin concentration of skeletal muscles measured in vivo with wearable near infrared spectroscopy, Ieee sensors journal. 18 (2018) 2326-2334.

[9] J. M. Chen, L. J. Peng, Y. Han, et al, A rapid quantification method for the screening indicator for β-thalassemia with near-infrared spectroscopy, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 193 (2018) 499-506.

[10] Y. P. Du, Y. Z. Liang, J. H. Jiang, et al, Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares, Anal. Chim. Acta. 501 (2004) 183-191.

[11] J. M. Chen, Z. W. Yin, Y. Tang, T. Pan, Vis-NIR spectroscopy with moving-window PLS method applied to rapid analysis of whole blood viscosity, Anal. Bioanal. Chem. 409 (2017) 2737-2745.

[12] J. M. Chen, T. Ai, T. Pan, L. J. Yao, AO-MW-PLS method applied to rapid quantification of teicoplanin with near-infrared spectroscopy, Journal of Innovative Optical Health Sciences. 9 (2016) 1650029.

[13] L. J. Yao, N. Lyu, J. M. Chen, et al, Joint analyses model for total cholesterol and triglyceride in human serum with near-infrared spectroscopy, Spectrochimica Acta Part A: Molecular And Biomolecular Spectroscopy. 159 (2016) 53-59.

[14] S. H. Wang, Y. Zhao, R. Hu, et al., Analysis of near-infrared spectra of coal using deep synergy adaptive moving window partial least square method based on genetic algorithm, Chinese Journal of analytical chemistry. 47 (2019) E19034-E19044.

[15] T. Pan, M. M. Li, J. M. Chen, Selection Method of Quasi-Continuous Wavelength Combination with Applications to the Near-Infrared Spectroscopic Analysis of Soil Organic Matter, Appl. Spectrosc. 68 (2014) 263-271.