Multispectral Analysis for Quantitative Measurements of Myoglobin Oxygen Fractional Saturation in the Presence of Hemoglobin Interference

LORILEE S. L. ARAKAKI and DAVID H. BURNS*
Center for Bioengineering, WD-12, University of Washington, Seattle, Washington 98195

Quantitative values for myoglobin oxygen fractional saturation were extracted from visible absorption spectra of myoglobin and hemoglobin solutions by analysis with three algorithms: classical least-squares, partial least-squares, and stagewise multiple linear regression. In an effort to mimic *in vivo* conditions, oxygen tensions and concentrations of myoglobin and hemoglobin solutions in separate cuvettes were varied independently. Transmission measurements were made through both cuvettes so that spectra contained contributions from both myoglobin and hemoglobin. Oxygen tensions in the myoglobin solutions spanned the rapidly varying region of the myoglobin oxygen saturation curve with pO_2 ranging from 0 to 4.79 Torr, corresponding to fractional saturation values between 0 and 0.903. A range of hemoglobin oxygenations from fully oxygenated to fully deoxygenated was used. Estimation of myoglobin fractional saturation by the classical least-squares algorithm had a standard error (SE_{est}) of 0.094, while the partial least-squares method resulted in an SE_{est} of 0.070. Partial least-squares estimations resulted in an SE_{est} of 0.041 when a limited wavelength range was used. The stagewise multiple linear regression method had an SE_{est} of 0.052. Results indicate that stagewise regression and partial least-squares yielded estimates of myoglobin fractional saturation that were more accurate than those obtained from classical least-squares.

Index Headings: Myoglobin; Oxygen; Absorption spectroscopy; Chemometrics; UV-visible spectroscopy.

INTRODUCTION

In the visible spectral region, the chromogenic species found in muscle are hemoglobin, myoglobin, and the cytochromes. Hemoglobin is contained within the vascular compartment, while myoglobin and cytochromes are located within muscle cells. Myoglobin is of particular interest since it is a direct probe of intracellular pO_2 in myocytes. The magnitudes of spectral contributions from the oxygenated and deoxygenated forms of myoglobin correlate directly with the pO_2 of the intracellular medium.

Myoglobin is a respiratory protein found in striated and cardiac muscle. These tissues perform sustained work and require large oxygen supplies. The primary function of myoglobin is to facilitate the intracellular transport of oxygen from the capillaries to the mitochondria. Myoglobin also serves as an intracellular oxygen store: the myoglobin-bound oxygen content is estimated to be thirty times that of free O_2 in normal muscle cells.[1] Inhibition of myoglobin function results in decreased oxygen uptake by the muscle and decreased isometric twitch tension.[2]

The degree of oxygen binding to myoglobin can be expressed as myoglobin oxygen fractional saturation (S), with

$$S = [MbO_2]/([MbO_2] + [Mb]) \quad (1)$$

where $[MbO_2]$ is the concentration of oxygenated myoglobin, and [Mb] is the concentration of deoxygenated myoglobin. S depends on both oxygen tension and temperature. The strong affinity that the single heme group of myoglobin has for oxygen results in low p50 values (oxygen tension when $S = 50\%$). At 37°C, the p50 for myoglobin is 2.5 Torr, while at 25°C it is 1 Torr.[3] Myoglobin oxygen binding is independent of pH in the range of 5.5 to 11.3.[3]

Detection of the myoglobin signal is impaired by the presence of hemoglobin *in vivo*. In aerobic mammalian species, the hemoglobin concentration in skeletal muscle at rest (calculated from capillary density)[4] is on the order of 50 μM. Since each hemoglobin molecule contains four heme groups, the heme concentration is approximately 200 μM. This value is comparable to myoglobin concentrations (4.35 mg myoglobin/g wet weight skeletal muscle[5] corresponds to approximately 250 μM myoglobin). Therefore, in skeletal muscle at rest, the hemoglobin contribution to a measured spectrum will be approximately equal to the myoglobin contribution. When muscles perform work, additional capillaries are recruited to carry blood to the tissue. Larger blood volumes result in increased hemoglobin interference. A similar interference problem exists in cardiac muscle. Capillary density, and thus, hemoglobin concentration, is greater in cardiac muscle than in skeletal muscle.[4] However, the myoglobin concentration in cardiac muscle is generally lower than in skeletal muscle.[1] In contrast to hemoglobin, cytochromes do not present a large interference problem. The cytochrome concentration is at least an order of magnitude lower than that of myoglobin.[6]

Hemoglobin further confounds the myoglobin spectral measurement because the absorbance spectrum of hemoglobin is similar to that of myoglobin. Figure 1 shows the spectra of oxygenated myoglobin (MbO_2), deoxygenated myoglobin (Mb), oxygenated hemoglobin (HbO_2), and deoxygenated hemoglobin (Hb) individually. The main spectral feature that distinguishes the oxygenated species is the 4-nm red shift[7] of the α-peak of MbO_2 relative to HbO_2. The large degree of overlap between these spectra is the major obstacle to resolving myoglobin oxygenation when hemoglobin is present.

Several studies have measured combined myoglobin and hemoglobin signals from muscle *in vivo* using both visible spectroscopy[8,9] and near-infrared spectrosco-

Received 15 June 1992.
* Author to whom correspondence should be sent.

0003-7028/92/4612-1919$2.00/0

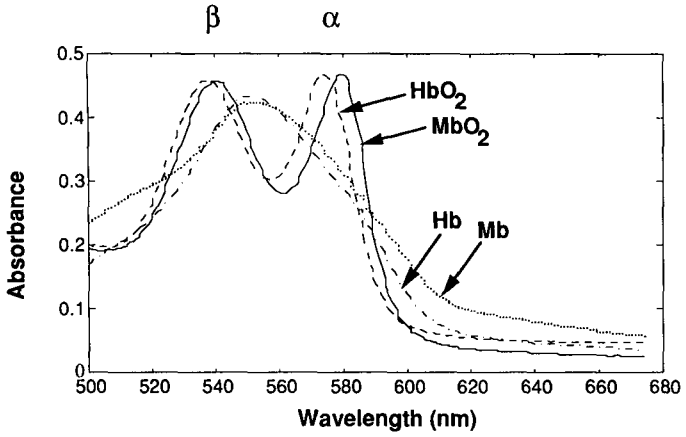

FIG. 1. Absorbance spectra of MbO_2 (——), Mb (·····), HbO_2 (– –), and Hb (–·–·). The main spectral feature distinguishing MbO_2 and HbO_2 is the 4-nm shift in the α-peak. The α-peak of MbO_2 occurs at 581 nm, while that of HbO_2 falls at 577 nm. The β-peaks in spectra of the oxygenated species and the single peaks of the deoxygenated species are highly overlapping.

py.[10,11] However, because of the large spectral overlap between myoglobin and hemoglobin absorbance spectra, no attempt was made to distinguish between the oxygen saturations of the two components. Myoglobin oxygen saturation measurements from tissue were possible when the tissue was perfused with a hemoglobin-free saline solution.[12–14] Unfortunately, the perfusion limits the usefulness of the technique for physiological measurements. Honig and co-workers obtained hemoglobin-free reflectance spectra of myoglobin by freezing tissue specimens and using microspectrophotometry to study individual muscle cells.[15,16] Although quantitative myoglobin fractional saturation values were obtained, the effects of the rapid-freezing procedure on the measured myoglobin absorbances are unknown. Also, only a single time-point measurement can be made, which greatly limits physiological studies.

Recently, a full-spectrum classical least-squares (CLS) analysis has yielded simultaneous myoglobin and hemoglobin fractional saturation estimations *in vitro*.[17] Reasonable agreement with known values was obtained from spectra of mixtures of myoglobin and hemoglobin solutions in one cuvette. These results are encouraging, and demonstrate that multispectral analysis is capable of distinguishing myoglobin and hemoglobin spectra. However, the experimental setup used does not model the *in vivo* situation where the two proteins are contained in separate compartments (myocyte and vascular) at different oxygen tensions. The p50 of myoglobin (2.5 Torr at 37°C) is much lower than the p50 of hemoglobin (26 Torr at pH = 7.4 and 37°C).[18] Therefore, when spectra are acquired from myoglobin and hemoglobin mixtures in the same cuvette, hemoglobin remains almost completely desaturated in the pO_2 range where myoglobin saturation changes rapidly. *In vivo* myoglobin and hemoglobin fractional saturations, as well as their concentrations, can change simultaneously and independently. The increased uncertainty that independent changes in oxygenation levels and concentration would cause in spectra may be problematic for the CLS method. It has been shown that least-squares coefficients of highly overlapped components are especially prone to errors due to noise in the data.[19]

Several other multispectral analysis methods have been developed which provide more robust estimations of analyte concentration than CLS. Partial least-squares (PLS) is a full-spectrum technique that has gained wide acceptance. Use of the optimal number of factors can result in excellent predictions by PLS because spectral contributions that are uncorrelated to concentration are weighted lightly. In contrast, CLS weighs all of the spectral information evenly in its prediction process, including noise and uncorrelated background.

An alternative approach to acquiring accurate estimations from spectra with varying background noise is stagewise multiple linear regression (SMLR). This method has been used to identify the wavelengths in a spectrum that are highly correlated with concentration.[20] Using absorbance values at only these wavelengths may improve estimations of concentration, since most of the uncorrelated wavelengths in the spectrum are not included.

The goal of the present experiments was to obtain accurate, quantitative estimates of myoglobin fractional saturation (S) from visible spectra of myoglobin and hemoglobin solutions. In an effort to mimic *in vivo* conditions, the oxygen tensions and concentrations of the myoglobin and hemoglobin solutions were varied independently. Two full-spectrum analysis techniques (CLS and PLS) and a selected-wavelength method (SMLR) were applied to the spectra, and the accuracies of the estimates of myoglobin S were compared.

THEORY

Classical Least-Squares (CLS). The CLS technique uses individual spectra of the chromogenic species in a sample to determine the fraction of each one present from an observed spectrum. If there are multiple spectral components in a sample, the absorbance spectrum of the sample is assumed to be a linear combination of the component spectra. In matrix notation, this can be expressed as

$$\mathbf{R} = \mathbf{KA} + \epsilon, \qquad (2)$$

where $\mathbf{R}$ is the $n \times m$ response matrix of sample spectra with n wavelengths per spectrum and m samples. $\mathbf{K}$ is the $n \times p$ spectral matrix of pure components, with p indicating the number of pure components. $\mathbf{A}$ ($p \times m$) contains the fractional coefficients reflecting concentrations of individual components in each sample, and ϵ ($n \times m$) is the error matrix.

An estimate of $\mathbf{A}$ ($\hat{\mathbf{A}}$) can be obtained with the use of the least-squares criterion, which minimizes the sum of squared errors between the sample spectra and the pure component spectra. The solution to the classical least-squares equation is given by

$$\hat{\mathbf{A}} = (\mathbf{K}^t\mathbf{K})^{-1}\mathbf{K}^t\mathbf{R} \qquad (3)$$

where t denotes the transpose of a matrix, and -1 the inverse.

When the pure spectra of MbO_2 and Mb are included in $\mathbf{K}$, myoglobin fractional saturation is calculated for each sample by using coefficients in $\hat{\mathbf{A}}$ and Eq. 1. Coef-

ficients represent percentages of myoglobin and hemoglobin present in the oxygenated, deoxygenated, or oxidized form. Negative coefficients are nonsensical. To constrain the results to nonnegative values, one can use an iterative approach based on the Kuhn-Tucker theorem. A detailed description of the method is given by Lawson and Hanson.[21] Considerable improvement in component quantitation has been achieved with this approach for other least-squares analyses.[19]

Partial Least-Squares (PLS). PLS analysis has demonstrated significant success for constituent estimation in complicated mixtures. It is based on an inverse model, where constituent values are expressed as a function of spectral responses. A multispectral myoglobin saturation data set can be written in matrix notation as

$$\mathbf{C} = \mathbf{R}\mathbf{B}^t \tag{4}$$

where **C** contains constituent values of myoglobin S for m samples, **R** is a matrix of spectral responses at various wavelengths for each sample ($m \times n$), and $\mathbf{B}^t$ is a vector of n constants of proportionality between **R** and **C**. $\mathbf{B}^t$ contains weighting constants for each wavelength, which result in estimated myoglobin saturation when multiplied by a myoglobin absorbance spectrum.

The object in a calibration is to estimate $\mathbf{B}^t$ for a data set where **R** and **C** are known.

$$\widehat{\mathbf{B}^t} = \widehat{\mathbf{R}^{-1}}\,\mathbf{C} \tag{5}$$

where $\widehat{\mathbf{B}^t}$ is the calibration coefficient vector and $\widehat{\mathbf{R}^{-1}}$ is an estimate of the inverse of **R**. Inversion of **R** can be accomplished if **R** is first decomposed into three matrices,

$$\mathbf{R} = \mathbf{U}\mathbf{V}\mathbf{W}^t \tag{6}$$

where columns of **U** and **W** are orthonormal. **U** and **W** are dimensioned $m \times r$ and $n \times r$, respectively. **V** is an $r \times r$ nonsingular matrix, and r is the rank of **R**. **U** and **W** form a coordinate system, and account for all the variance in the data, much like x, y, z spatial coordinates account for spatial variability in the placement of a physical object. **V** is the relative scaling between the various axes in the coordinate system. Since **U** and **W** are orthonormal, a pseudo-inverse of **R** can be obtained by

$$\widehat{\mathbf{R}^{-1}} = \mathbf{W}\mathbf{V}^{-1}\mathbf{U}^t. \tag{7}$$

Thus, a calibration coefficient vector, $\widehat{\mathbf{B}^t}$, is calculated by substituting Eq. 7 into Eq. 5.

The calibration step in PLS decomposes **R** into **U**, **V**, and **W** using an iterative approach. Details of the algorithm are given in several sources.[22-24] Stated simply, one factor (a column of **U** and **W** and a diagonal matrix element in **V**) is generated in each iteration. The algorithm orders the factors in such a way that each describes decreasing amounts of variance within the calibration spectra that are correlated to **C**. Therefore, the first factor encompasses the maximum nonrandom variance in the spectra that also correlates with myoglobin S. Factors generated late in the process describe smaller variances in the data that may be due to weak signals and random noise.

Since the factors are ordered according to the amount of variance in the data they describe, spectral signals can be segmented from noise to a certain degree. The calibration coefficient vector associated with the optimal number of factors, k^*, will yield the best possible estimation of myoglobin S by PLS. This $\widehat{\mathbf{B}^t}$ vector will include enough factors so that all or most of the spectral complexities relevant to S within the calibration data set will be modeled. However, it will also exclude the latter factors which describe spectral variations due to noise and other uncorrelated phenomena.

Cross-validation within a calibration set can be used to calculate PRediction Error Sum of Squares (PRESS) values. PRESS values are indications of how closely a model fits the calibration data, and are used to specify k^*. For a data set with m samples whose myoglobin S values are known, one sample is removed from the data set. With the use of PLS, the remaining $m - 1$ sample spectra are regressed against S to calculate a best fit model. This predictive model is then used to estimate S of the sample left out. The entire process is repeated, leaving out each sample spectrum in turn. After each factor is added to the model, a PRESS value is calculated by comparing each predicted S value with each known value, and then summing the squares of the prediction errors.

A model containing the number of factors that yields the minimum PRESS value, k_{min}, does not necessarily contain the optimal number of factors, k^*, since this model may overfit the data.[23] The best model includes the fewest number of factors such that the PRESS for that model is not significantly greater than the minimum PRESS value. As described by Haaland and Thomas,[23] an F-statistic, $F(k) = \text{PRESS}(k)/\text{PRESS}(k_{min})$, can be computed for each model. The optimal number of factors is the smallest k such that $F(k) < F_{0.25,\,m,m}$. The value 0.25 is obtained from the 75% confidence level for the F distribution, and the subsequent indices refer to the number of degrees of freedom. In this analysis, both are equal to the number of calibration spectra, m. PLS predictions of myoglobin S are obtained by a multiplication between absorbance spectra of test samples and the $\mathbf{B}^t$ vector associated with k^* factors.

Stagewise Multiple Linear Regression (SMLR). Calibration of stagewise regression obtains the vector of coefficients $\mathbf{B}^t$ when **R** and **C** are known, as described by Eq. 4. In SMLR, myoglobin S is regressed linearly against absorbances at each wavelength. The wavelength with the highest correlation coefficient is saved. The amount of variance in the spectra that this wavelength accounts for is then subtracted from the spectra. Upon the next iteration, myoglobin S is regressed linearly against the reduced data set, identifying the wavelength with the second-highest correlation to the constituent value. After each additional wavelength is added, $\mathbf{B}^t$ is obtained by solving

$$\mathbf{C} = \mathbf{1}b_0 + \mathbf{R}_s\mathbf{B}^t \tag{8}$$

where $\mathbf{R}_s$ is a matrix containing absorbances of the m sample spectra at the selected k wavelengths ($m \times k$), $\mathbf{B}^t$ comprises k calibration coefficients, **1** is a vector of ones, and b_0 is an intercept term.

As in the selection of factors in PLS, an optimal number of wavelengths for the calibration model, k^*, balances adequate modeling and overfitting. The optimal model

contains the minimum number of wavelengths such that adding another wavelength to the model does not significantly increase the prediction accuracy.[20] For objective determination of k^* from models produced with different k values, a partial F-test[25] is useful. Prediction of S from a test spectrum takes the form of

$$c = b_o + \sum_{j=1}^{k^*} r_j b_j \quad (9)$$

where c is the estimated myoglobin S value, b_o is the intercept, and r_j and b_j are the absorbance and calibration coefficient of the jth wavelength, respectively.

MATERIALS AND METHODS

Optical Spectrophotometer. Myoglobin fractional saturation measurements were made with a charge-coupled device (CCD) array spectrophotometer developed for this purpose. A fiber-optic bundle conveyed light from a quartz halogen source to the sample. In an effort to simulate the *in vivo* situation where myoglobin and hemoglobin fractional saturations and concentrations vary independently, spectra were acquired from myoglobin and hemoglobin solutions in separate cuvettes. The cuvette holder accommodated two abutting cuvettes in the light path. With this configuration, transmission spectra containing both myoglobin and hemoglobin can be acquired. A second fiber-optic bundle carried the transmitted light to the input slit of a monochromator, where the light was dispersed into a spectrum.

A CCD linear array detector (Texas Instruments, Dallas, TX) was affixed to the spectrograph output slit and collected the spectrum. The CCD array has been modified for visible spectroscopy as described by Sweedle *et al.*[26] Each of the 3456 CCD pixels is 10.7 × 10.7 μm. Groups of eight adjacent pixels were summed and averaged together. The resulting spectral sampling of the system was 0.8 nm/spectral point.

A Macintosh II (Apple Computer, Inc., Cupertino, CA), along with software that was written in Think C (Symantec Corp., Cupertino, CA), supplied clock signals to the CCD array. The protocol directed the serial digitization of the accumulated charge in the CCD pixels by a 100-kHz A/D board (National Instruments, Corp., Austin, TX). Typical myoglobin spectra had a signal-to-noise ratio of 500, with a maximal noise level of 0.001 absorbance units. Absorbances between 500 and 675 nm were retained and processed.

Myoglobin and Hemoglobin Solutions. Lyophilized horse heart myoglobin from Sigma Chemical Co. (St. Louis, MO) was used. When the crystals are brought into solution, the heme iron in the myoglobin is predominantly in the oxidized state, resulting in metmyoglobin (metMb). This oxidized form of myoglobin does not bind oxygen. The concentration of metMb is held very low in the body by the enzyme system metmyoglobin reductase.[1,3]

Standard myoglobin solutions consisted of a 10-mL mixture of 0.1 mM myoglobin, 0.5 mM NADH, and 10.0 mM Tris buffer (pH 7.5). A method developed by Kajita *et al.*[27] was used to reduce metMb solutions. Reduction is initiated aerobically by the addition of 100 μL of 25 μM phenazine methosulfate (PMS). At temperatures between 21 and 26°C, reduction was complete in thirty minutes. The presence of metMb could be detected by a broad peak at 640 nm, where MbO_2 does not absorb. Solutions remained reduced for sixty minutes to two hours before reverting to metMb.

Hemoglobin solutions were made from heparinized venous blood obtained from mongrel dogs. The blood was diluted with 0.05 M Tris buffer (pH = 7.4) until the heme concentration was roughly the same as that of the myoglobin. This concentration ratio was intended to mimic skeletal muscle at rest, where the heme concentrations of myoglobin and hemoglobin are on the same order (see Introduction). In addition, one hemoglobin solution was diluted to half the concentration of the others to provide a hemoglobin interferent at a different concentration. There was no evidence of methemoglobin in any of the spectra used for the data collection.

Gas Equilibration. After addition of PMS to the myoglobin mixture, approximately 0.75 mL of the solution was put into five glass cuvettes, each having a pathlength of 4 mm. The cuvettes were mounted on a gas equilibration rocker, equipped with gas humidifiers. Humidified, calibrated gases were transmitted to the cuvettes via stainless steel needle tubing while the cuvettes were tilted end to end at a rate of 1 Hz. At room temperatures between 21 and 26°C, myoglobin solutions attained equilibrium with the gases in 90 min to 3 ½ h. The barometric pressure was recorded during the experiment, and the room temperature was monitored every fifteen minutes. During the collection of a complete data set, the maximum variation in room temperature was 2°C.

Five tanks of calibrated gases were obtained from Airco Special Gases (Vancouver, WA). The oxygen contents of the gases were chosen to fully span the monotonically increasing portion of the myoglobin fractional saturation curve. They were 0%, 0.102%, 0.29%, 0.489%, and 0.648% O_2, with the balance being nitrogen. At 25°C and a barometric pressure of 760 Torr, the gas compositions correspond to oxygen tensions between 0 and 4.77 Torr. The vapor pressure of a humidified gas is dependent on both temperature and barometric pressure. Since temperature and pressure conditions varied between experiments, the same gas compositions yielded different oxygen tensions on different days. In addition to the calibrated gases, a myoglobin solution was equilibrated with room air to obtain a fully oxygenated myoglobin sample.

For a representation of the full range of hemoglobin oxygenation states that occurs *in vivo,* fully oxygenated (HbO_2), fully deoxygenated (Hb), and half-saturated hemoglobin solutions were made. They were obtained by equilibration with room air, N_2, and 2.97% O_2, respectively. In addition, the hemoglobin solution prepared at half-heme concentration was equilibrated with 2.97% O_2, resulting in an approximately half-saturated, half-concentrated solution.

Data Collection. Spectra of the five myoglobin solutions were collected at thirty-minute intervals after the addition of PMS to metMb. When the myoglobin spectra from consecutive time points were identical, it was determined that the solution had attained equilibrium with the gas. Upon equilibration, combination spectra with

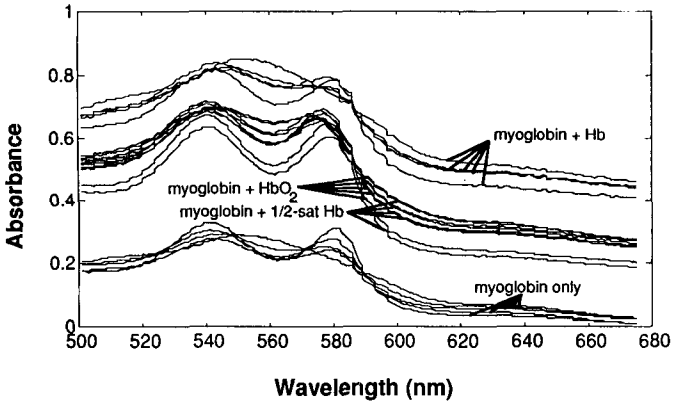

FIG. 2. Absorbance spectra of myoglobin/hemoglobin combinations. Spectra of myoglobin equilibrated with the five gases appear in the 0.2–0.3 absorbance range. All other spectra contain contributions from both myoglobin and hemoglobin. The group of spectra at the highest absorbance values are spectra of myoglobin and deoxygenated hemoglobin. Spectra of myoglobin plus oxygenated hemoglobin and myoglobin plus half-saturated hemoglobin appear in the 0.4–0.6 absorbance range, and are difficult to distinguish from one another when overlayed.

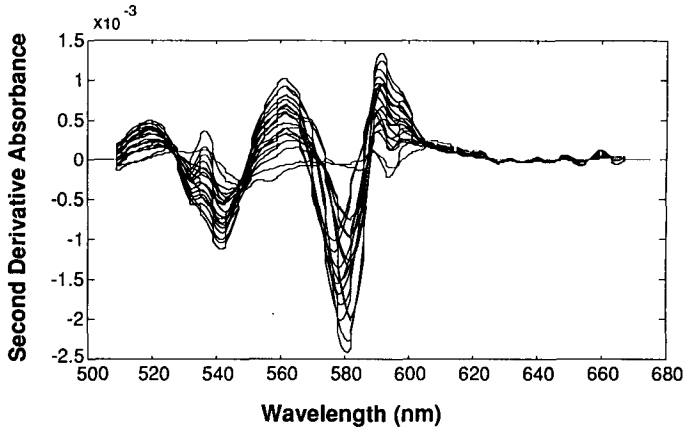

FIG. 3. Second derivatives of spectra in Fig. 2. The second-derivative processing removed most of the baseline variations in the raw spectra due to concentration differences and cuvette-to-cuvette variations. Removal of the baseline is best observed in the 620–675 range.

each myoglobin solution and the four hemoglobin solutions were obtained. Spectra of the hemoglobin solutions alone were also taken to monitor the stability of these samples.

Second-Derivative Preprocessing. In order to remove spectral baselines and dc offsets, the second derivative of each of the spectra was determined with the use of discrete differences.[28] A three-point averaging was used to compute the second derivative so that noise was reduced in the resultant spectra. Provided that the bandwidth of the background sloping peak is large in comparison to those of the absorbance peaks of interest, the second derivative resolves the sharp peaks and provides a near-zero baseline elsewhere.[28,29]

Multispectral Analysis. All data analysis was carried out in Matlab (The MathWorks, Inc., Natick, MA). The CLS routine, included in the Matlab package, restricted the least-squares coefficients to nonnegative values. A description of the algorithm is given by Lawson and Hanson.[21] The PLS and SMLR procedures were written on the basis of their descriptions by Haaland and Thomas[23] and Draper and Smith,[20] respectively.

In order to make a consistent comparison, the same data set was used for the calibration step in the three multispectral analysis techniques described. A second data set, referred to as the test set, was used for the prediction step.

Reference Myoglobin Oxygen Fractional Saturation Values. There are several reasons why myoglobin S values were used as the standard reference for myoglobin oxygenation, instead of pO_2 values within the solutions. As mentioned before, the absolute pO_2 of a humidified gas will depend on the temperature and barometric pressure. Second, myoglobin oxygen binding is temperature-dependent. Therefore, myoglobin at different temperatures but equilibrated with gas at the same pO_2 will have different myoglobin fractional saturations. Calibration is complicated further by the fact that myoglobin fractional saturation varies nonlinearly with pO_2. However, assuming that the total myoglobin concentration ($[MbO_2]$ + $[Mb]$) is constant, myoglobin spectral changes and $[MbO_2]$ vary linearly with S (see Eq. 1). Therefore, using S values permits a linear calibration of spectral changes over a range of pO_2 values.

The fractional saturation values shown on the x axis for data points in Figs. 4 through 7 were determined by measuring the spectra of myoglobin alone, without hemoglobin interference. Fractional saturation was determined from two reference myoglobin oxygenations: fully desaturated (0% O_2) and fully saturated (21% O_2 room air). The pure metMb spectrum was measured before PMS was added. A nonnegative least-squares estimation of the MbO_2, Mb, and metMb coefficients was made from the second derivatives of the spectral measurements. Individual fractional saturation points were calculated from Eq. 1 with the use of these coefficients. Due to the nature of Eq. 1 and the nonnegative constraint on the coefficients, all fractional saturations were bound between zero and one. Variations in estimated fractional saturation due to deviations in the pure spectra did not exceed 5% in either the calibration or prediction set. The average metMb contribution to the spectra was 8%.

RESULTS AND DISCUSSION

Figure 2 shows a set of myoglobin/hemoglobin combination spectra. Spectra of myoglobin solutions equilibrated with the five gases used have the lowest absorbances and appear in the 0.2–0.3 absorbance range. All other spectra contain contributions from hemoglobin as well. The additional presence of hemoglobin in the light path results in absorbance values two to three times higher than those observed in spectra of myoglobin solutions only.

A lack of coincidence at the two isosbestic points within this wavelength range (530 nm and 589 nm) is observed in the five spectra containing only myoglobin. Deviations in optical pathlength, cuvette wall thickness, and cuvette positioning within the holder are the most probable causes for different baseline offsets. A similar lack of coincidence at isosbestic points is often seen in spectra of myoglobin and hemoglobin acquired in tissue.[12] Varying spectral baselines can be attributed to wavelength-de-

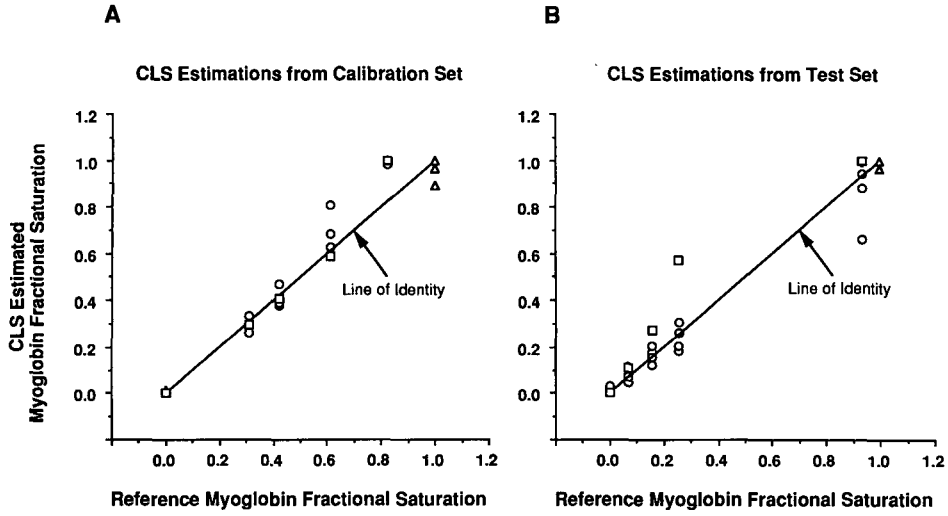

FIG. 4. CLS estimation results, with the line of identity. CLS prediction error increases greatly with myoglobin fractional saturation. (**A**) Myoglobin/hemoglobin spectra were regressed against pure spectra (MbO_2 Mb, metMb, HbO_2, and Hb) taken on the same day. (**B**) Myoglobin/hemoglobin spectra from the test data set were regressed against pure spectra from the calibration data set using CLS. Estimations from solutions containing only myoglobin are shown by squares, estimations from solutions that contain myoglobin and hemoglobin by circles. Triangles indicate deviations in myoglobin fractional saturation due to variations in pure MbO_2 and Mb spectra in the calibration data set (**A**) and in the test data set (**B**) and are partially obscured by other data symbols. The myoglobin fractional saturation values are bounded by zero and one, because of the nonnegative coefficients used in Eq. 1.

pendent scattering, variations in myoglobin and hemoglobin contents between different muscles, and temporal changes in blood flow and local myoglobin content.

Second derivatives of the spectra in Fig. 2 are shown in Fig. 3. Successful removal of the dc offsets between spectra is most apparent in the 620–675 nm range. Within the broad envelope of peaks in the 570–585 nm range, variations in peak wavelength and intensity can be seen. These reflect varying proportions of the α-peak contributions from HbO_2 (blue-shifted) and MbO_2 (red-shifted) in the solutions.

Reduction of baseline differences with the use of second derivatives enhances the relative magnitude of changes in the spectra, and the MbO_2 and Mb spectra become more distinguishable. The Mb spectrum deviates little from zero in the second derivative because the single Mb peak is very broad. In contrast, the negative-going α- and β-peaks in the second-derivative MbO_2 spectrum are very distinct.

CLS. CLS performance was first assessed on spectra within one data set. Second derivatives of mixed myoglobin/hemoglobin spectra in the calibration data set were placed in the response matrix, **R** (see Theory). The second derivatives of pure MbO_2, Mb, metMb, HbO_2, and Hb spectra taken during the same experiment comprised **K**. Figure 4A shows the estimated myoglobin S from this analysis. The line of identity is superimposed on the plot. The standard deviation of the residuals between reference and estimated S values, SE_{est}, was 0.083. The correlation coefficient, R^2, was 0.971 (Table I). Excluding metMb from the pure spectrum matrix **K** did not significantly change the accuracy of the estimations.

TABLE I. Statistical comparison between CLS, PLS, and SMLR modeling of the calibration data set (n = 20).

	CLS	PLS[a]	PLS[b]	SMLR
SE_{est}	0.083	0.018	0.018	0.018
R^2	0.971	0.996	0.996	0.996

[a] Calibration on full wavelength range.
[b] Calibration on 570–610 nm range only.

Estimates of fractional saturation based on spectra from solutions that contained only myoglobin are marked by squares in Fig. 4A. When only these data points are considered, the difference in the way the x-axis reference values and the y-axis estimated values were obtained was that HbO_2 and Hb pure spectra were included in the **K** matrix in the latter, but not in the former. Since no hemoglobin was present in these spectra, any deviations from the line of identity are a result of the algorithm incorrectly assigning nonzero coefficients to HbO_2 and/or Hb. Except for the most oxygenated solution, fractional saturation values in pure myoglobin solutions were accurately estimated. Most of the prediction error in the remaining solutions, marked by circles, is caused by inaccurate CLS determinations of [MbO_2] and [Mb] when hemoglobin is present. Triangle symbols in Fig. 4A show the variation in estimated S for representative pure MbO_2 and Mb spectra in the calibration set. The variation in the pure spectra is comparable to the variation of the CLS estimated values about the line of identity.

The most striking feature in Fig. 4A is the large increase in prediction error as S increases. A propagation of error calculation reveals the reason. An error of magnitude ϵ_1 in the coefficient for MbO_2 results in a percent error in S of $(\text{Mb}\ \epsilon_1)/[\text{MbO}_2(\text{MbO}_2 + \text{Mb} + \epsilon_1)]$. An error of magnitude ϵ_2 in the coefficient for Mb propagates to a percent error of $\epsilon_2/(\text{MbO}_2 + \text{Mb} + \epsilon_2)$ in S. If $\epsilon_1 = \epsilon_2$, these two expressions are different by a multiplicative constant of Mb/MbO_2. At high fractional saturation values, [Mb] is low and [MbO_2] is high, and this constant is small. Under these conditions, the percent error in S caused by an error in the MbO_2 coefficient is much less than that caused by the same amount of error in the Mb

TABLE II. Statistical comparison between CLS, PLS, and SMLR, indicating accuracy of estimations based on test data set (n = 25).

	CLS	PLS[a]	PLS[b]	SMLR
SE_{est}	0.094	0.070	0.041	0.052
R^2	0.962	0.965	0.994	0.987

[a] Calibration on full wavelength range.
[b] Calibration on 570–610 nm range only.

coefficient. Therefore, at high values of S, the percent error in S is more sensitive to errors in the Mb coefficient than in the coefficient for MbO_2. Unfortunately, since the Mb and Hb signals are highly overlapping, there is significant error in the CLS determination of the Mb coefficient.

When myoglobin/hemoglobin spectra from the test set were regressed against the pure spectra from the calibration set using CLS, myoglobin S estimates were not as accurate as those from the calibration set (see Fig. 4B). Table II shows that SE_{est} increased to 0.094, while R^2 decreased to 0.962, in comparison to the first case. The test data set was acquired at a different temperature from the calibration set, so that it contained spectra at different myoglobin and hemoglobin oxygenations. In addition, there were slightly different concentrations of myoglobin, metMb, and hemoglobin in the solutions. Instrumental noise also contributed to differences between these data sets. These differences between the calibration and test data sets were primarily responsible for the increase in errors of estimations from solutions containing only myoglobin (compare squares in Fig. 4A and 4B). Estimations based on spectra that also contained hemoglobin had large deviations because errors due to hemoglobin interference were compounded by differences between the data sets. As in Fig. 4A, the prediction error increases dramatically with S, a result of the poor CLS distinction between the spectra of Mb and Hb.

PLS. A PLS calibration by cross-validation was done on the myoglobin/hemoglobin spectra in the calibration data set. The optimal number of factors indicated by the PRESS values was five. Under ideal conditions, it was expected that only four factors would be required to optimally describe the calibration data set since there are four independent components (MbO_2, Mb, HbO_2, and Hb) causing variations in the observed spectra. The fifth factor was probably needed to account for nonlinearities in the spectral baselines that the second derivative did not remove.

PLS estimations obtained by multiplying the calibration coefficient vector (which encompasses the first five factors) and each spectrum in the calibration set are shown in Fig. 5A. Accuracy of the estimations is very good, with SE_{est} = 0.018 and R^2 = 0.996 (Table I). Five factors model the calibration set well. Unlike CLS, the estimations based on spectra of solutions containing both myoglobin and hemoglobin were just as accurate as those based on solutions containing only myoglobin. Also, the magnitude of prediction errors remained constant over the full range of S.

When estimations were made from the test set with the use of the same calibration coefficient vector as before, the magnitudes of the errors increased (Fig. 5B). As seen in Table II, SE_{est} = 0.070 and R^2 = 0.965. All estimations of S from solutions that contained only myoglobin (squares) were overestimated. Differences in myoglobin concentrations between data sets could account for this trend.

To compare the predictive capabilities of two analysis techniques objectively, one should make a comparison between the ratio $PRESS_{method\ 1}/PRESS_{method\ 2}$ and the $F_{\alpha,m,m}$ statistic.[23] The variable m represents the number of degrees of freedom, which for this comparison equals the number of spectra used in the analyses. The quantity $1 - \alpha$ defines the confidence level at which the null hypothesis can be rejected. In this case, the null hypothesis states that the errors of both methods have equal variances, or that neither method provides signif-

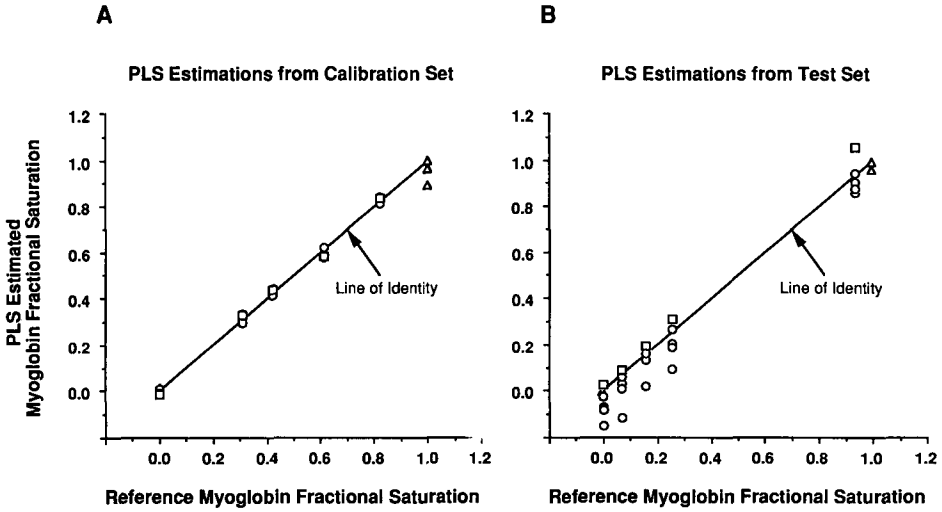

FIG. 5. PLS estimation results, with the line of identity. In contrast to CLS, the magnitude of prediction error is constant over the entire S range. (A) Estimations of myoglobin fractional saturation from the calibration data set. Five factors model the calibration set well. (B) PLS estimations resulting from a multiplication between the calibration coefficient vector and myoglobin/hemoglobin spectra in the test set. Myoglobin fractional saturation estimations based on solutions containing hemoglobin were consistently low. Estimations from solutions with only myoglobin are shown by squares, estimations from solutions that also contain hemoglobin by circles. Triangles indicate deviations in myoglobin fractional saturation due to variations in pure MbO_2 and Mb spectra in the calibration data set (A) and in the test data set (B).

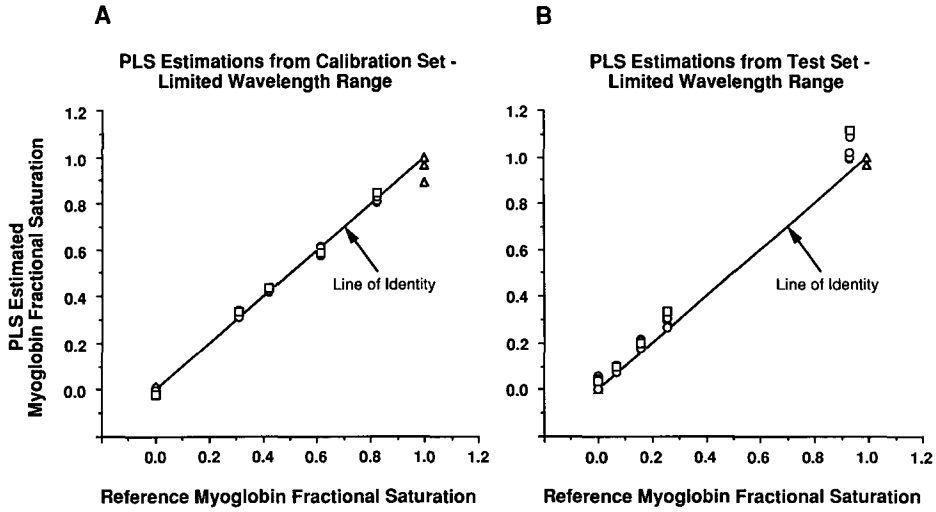

FIG. 6. PLS estimation results, with the line of identity. Estimations are significantly improved when the calibration is done on a limited wavelength range. (**A**) Estimations of myoglobin fractional saturation from the calibration data set. (**B**) PLS estimations resulting from a multiplication between the calibration coefficient vector and myoglobin/hemoglobin spectra in the test set. Estimations from solutions with only myoglobin are shown by squares, estimations from solutions that also contain hemoglobin by circles. Triangles indicate deviations in myoglobin fractional saturation due to variations in pure MbO_2 and Mb spectra in the calibration data set (**A**) and in the test data set (**B**).

icantly better estimations. When CLS and PLS estimations based on the calibration data set (data in Figs. 4A and 5A) are compared, the null hypothesis can be rejected at the 99.9% level. PLS clearly provides more accurate estimations in this case. However, when the test data set is estimated (Figs. 4B and 5B), one can reject the null hypothesis only at the 75% level. The magnitudes of error for PLS and CLS are not significantly different.

PLS results were improved by limiting the spectral region over which the calibration was made. Cross-validation calibrations were made on several spectral regions. Wavelengths between 570 and 610 nm yielded the best myoglobin S estimations. This wavelength range contained the α-peaks of myoglobin and hemoglobin, which are the most rapidly varying regions of the spectrum with respect to wavelength. Figure 6A shows that the calibration set is estimated accurately. The standard error of the estimate was 0.018, and the correlation coefficient was 0.996 (Table I). Estimations from the test data improved considerably over those from the full-spectrum analysis. Results yielded an SE_{est} of 0.041 and an R^2 of 0.994 (Fig. 6B, Table II). A PRESS comparison between these test set estimations and those from the CLS analysis reveals that the null hypothesis described above can be rejected at the 95% confidence level. Therefore, when the wavelength range is limited to include only peaks highly correlated to myoglobin S, PLS can provide more accurate estimates than CLS.

Stagewise MLR. SMLR was applied to second-derivative spectra of myoglobin/hemoglobin combinations from the calibration data set. The partial F-test confirmed that four wavelengths comprised the optimal calibration model. Results shown in Fig. 7A verify that the SMLR model estimated the calibration data set well. The standard error and correlation coefficients were indistinguishable from the PLS values in Table I ($SE_{est} = 0.018$, $R^2 = 0.996$). When fractional saturations were estimated from spectra in the test data set, SMLR had an SE_{est} of 0.052 and an R^2 of 0.987 (Fig. 7B, Table II).

A comparison of PRESS results from SMLR and PLS reveals that the prediction error variances were not significantly different when only the calibration data are considered (Figs. 5A, 6A, and 7A). Both of these methods modeled the calibration data set much better than CLS. Likewise, when the test data set was used for estimation (Figs. 6B and 7B), SMLR errors and errors from the PLS analysis using a limited wavelength range were not significantly different.

The success of this selected-wavelength method is due to the inclusion of only highly correlated data in the calibration. As few as four points out of the 220 in the full second-derivative spectrum can accurately model spectral changes caused by variations in myoglobin oxygenation. The wavelength most correlated with myoglobin S is located on the red side of the α-peak of MbO_2 at 583 nm. This wavelength contains information about the relative spectral contributions from MbO_2 and HbO_2. The remaining three wavelengths in the model are located at 637, 646, and 651 nm. Since there is little variation in peak height in this region (Fig. 3), these wavelengths account for baseline variations in the second-derivative spectra.

It should be noted that, when the second derivatives of the spectra are calculated, contributions from adjacent wavelengths are included in each second-derivative value. When SMLR was calibrated on raw spectra (no second-derivative preprocessing) from the calibration data set using four wavelengths, estimations from raw spectra in the test data set were extremely inaccurate. The standard error of the estimate was 0.413, with a correlation coefficient of 0.439. Differences in myoglobin and hemoglobin concentrations between data sets caused large baseline disparities, which led to an inadequate modeling of the test data set by the calibration step. These results indicate that acquisition of full spectra and subsequent

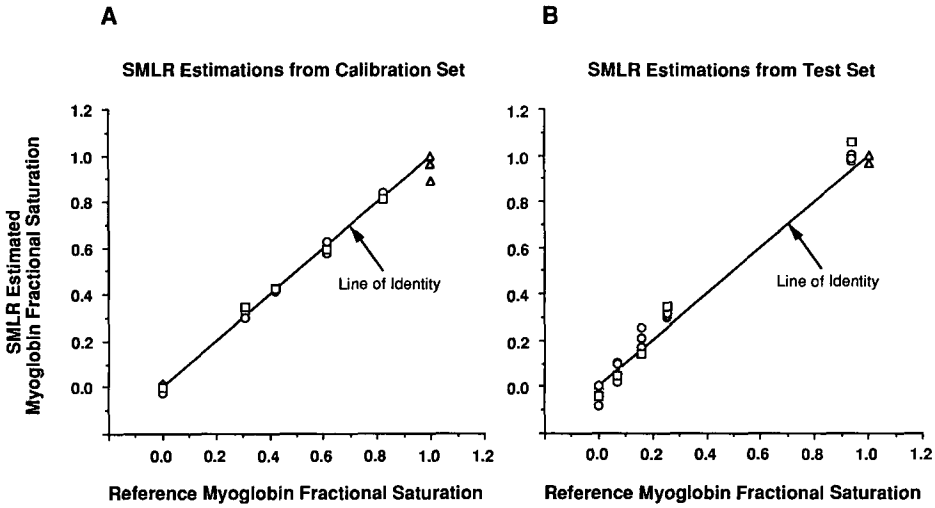

FIG. 7. SMLR estimation results, with the line of identity. SMLR yields more accurate results than CLS, and comparable results with the PLS limited-wavelength analysis. (**A**) Estimations of myoglobin fractional saturation from the calibration data set. Four wavelengths model the calibration data set well. (**B**) SMLR estimations, which are linear combinations of absorbances at wavelengths highly correlated to myoglobin fractional saturation. Estimations from solutions with only myoglobin are shown by squares, estimations from solutions that also contain hemoglobin by circles. Triangles indicate deviations in myoglobin fractional saturation due to variations in pure MbO_2 and Mb spectra in the calibration data set (**A**) and in the test data set (**B**).

second-derivative processing is needed for SMLR to yield accurate estimations of myoglobin S when baseline differences exist between calibration and test data sets.

CONCLUSIONS

Visible spectra of myoglobin and hemoglobin solutions with independently varying fractional saturations were analyzed with the use of classical least-squares, partial least-squares and stagewise multiple linear regression algorithms. Among the two full-spectrum algorithms, PLS analysis resulted in smaller estimation errors than CLS. The difference was statistically significant when the PLS calibration was done on a limited wavelength range that included only the sharp α-peaks of myoglobin and hemoglobin. PLS has a distinct advantage over CLS because the magnitude of PLS estimation errors did not vary over the range of myoglobin S values. An important feature of partial least-squares analysis is that residuals between the model and the spectra in the calibration set can be calculated for each wavelength. Large residuals would indicate improper modeling in a particular wavelength region.

Stagewise regression also provided accurate estimations of myoglobin oxygen fractional saturation. The estimation errors of this method were comparable to those from the PLS analysis using a limited wavelength range. SMLR was effective because it uses only spectral information that is highly correlated with S. This characteristic of the algorithm allows accurate estimations in situations where the peaks of interest are overlapping with interfering peaks. Also, accurate results can be obtained despite myoglobin and hemoglobin concentration differences between the calibration and test data sets.

It is important to note that second-derivative preprocessing was necessary for all three algorithms to estimate myoglobin fractional saturation accurately. Offsets and nonlinear baselines caused by differences in heme concentration, scattering, and instrumental noise were largely removed, allowing reasonable estimations from data sets taken under different conditions.

The use of PLS and SMLR brings the acquisition of quantitative, noninvasive myoglobin oxygen fractional saturation measurements *in vivo* a step closer. Several important basic physiology questions could be addressed with an intracellular oxygen sensor. For example, knowing the oxygenation level of tissue during ischemic, hypoxic, and recovery periods may help to elucidate the role of oxygen in the control of cellular respiration under these conditions.

ACKNOWLEDGMENTS

We are thankful to M. J. Kushmerick and E. O. Feigl for their constructive suggestions regarding the experimental design, for many stimulating discussions related to this paper, and for reviewing the manuscript. We thank C. H. Barlow and J. J. Kelly for their helpful advice regarding metmyoglobin reduction. We thank F. D. Smith for his technical expertise in constructing the gas equilibration rocker. This work was funded by a Biomedical Research Support Grant, NIH-RR07096.

1. B. A. Wittenberg and J. B. Wittenberg, Annu. Rev. Physiol. **51,** 857 (1989).
2. R. P. Cole, Science **216,** 523 (1982).
3. A. Rossi-Fanelli and E. Antonini, Arch. Biochem. Biophys. **77,** 478 (1958).
4. K. E. Conley, S. R. Kayar, K. Rosler, H. Hoppeler, E. R. Weibel, and C. R. Taylor, Resp. Physiol. **69,** 47 (1987).
5. L. J. Kagen, *Myoglobin* (Columbia University Press, New York, 1973).
6. D. L. Drabkin, J. Biol. Chem. **182,** 317 (1950).
7. G. A. Millikan, Physiol. Rev. **19,** 503 (1939).
8. M. Tamura and O. Hazeki, Annu. Rev. Physiol. **51,** 813 (1989).
9. B. Chance, S. Nioka, J. Kent, K. McCully, M. Fountain, R. Greenfeld, and G. Holtom, Anal. Biochem. **174,** 698 (1988).
10. N. B. Hampson and C. A. Piantadosi, J. Appl. Physiol. **64,** 2449 (1988).
11. C. A. Piantadosi, J. Crit. Care **4,** 308 (1989).
12. M. Tamura, N. Oshino, B. Chance, and I. A. Silver, Arch. Biochem. Biophys. **191,** 8 (1978).
13. L. Caspary, J. Hoffman, H. R. Ahmad, and D. W. Lubbers, Adv. Exp. Med. Biol. **191,** 263 (1985).

14. J. Hoffmann and D. W. Lubbers, Adv. Exp. Med. Biol. **200,** 125 (1986).
15. C. R. Honig, T. E. J. Gayeski, and R. J. Connett, Adv. Exp. Med. Biol. **180,** 651 (1984).
16. T. E. J. Gayeski, R. J. Connett, and C. R. Honig, Am. J. Physiol. **248,** H914 (1985).
17. C. H. Barlow, K. A. Kelly, and J. J. Kelly, Appl. Spectrosc. **46,** 758 (1992).
18. C. R. Honig, *Modern Cardiovascular Physiology* (Little, Brown and Company, Boston, 1981), 1st ed.
19. D. H. Burns, J. B. Callis, and G. D. Christian, Anal. Chem. **58,** 1415 (1986).
20. N. R. Draper and H. Smith, *Applied Regression Analysis* (John Wiley and Sons, New York, 1981), 2nd ed.
21. C. L. Lawson and R. J. Hanson, *Solving Least-Squares Problems* (Prentice-Hall, Englewood Cliffs, New Jersey, 1974).
22. A. Lorber, L. E. Wangen, and B. R. Kowalski, J. Chemom. **1,** 19 (1987).
23. D. M. Haaland and E. V. Thomas, Anal. Chem. **60,** 1193 (1988).
24. K. R. Beebe and B. R. Kowalski, Anal. Chem. **59,** 1007 (1987).
25. D. G. Kleinbaum, L. L. Kupper, and K. E. Muller, *Applied Regression Analysis and Other Multivariable Methods* (PWS-Kent Publishing Co., Boston, 1988), 2nd ed.
26. J. V. Sweedle, R. D. Jalkian, and M. B. Denton, Appl. Spectrosc. **43,** 953 (1989).
27. A. Kajita, K. Noguchi, and R. Shukuya, Biochem. Biophys. Res. Com. **39,** 1199 (1970).
28. F. Holler, D. H. Burns, and J. B. Callis, Appl. Spectrosc. **43,** 877 (1989).
29. W. L. Butler and D. W. Hopkins, Photochem. Photobiol. **12,** 439 (1970).