ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 278, No. 1, April, pp. 148–160, 1990

Factor Analysis of the Near-Ultraviolet Absorption Spectrum of Plastocyanin Using Bilinear, Trilinear, and Quadrilinear Models[1]

Stewart R. Durell, Choon-Hwan Lee,[2] Robert T. Ross,[3] and Elizabeth L. Gross[3]
Biophysics Program and Department of Biochemistry, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210

Received July 31, 1989, and in revised form December 4, 1989

Factor analysis was used to resolve the spectral components in the near-uv absorption spectrum of plastocyanin. The data set was absorption as a function of four variables: wavelength, species of plastocyanin, oxidation state of the copper center, and environmental pH. The data were fit with the traditional bilinear model, as well as with trilinear and quadrilinear models. Trilinear and quadrilinear models have the advantage that they uniquely define the components, avoiding the indeterminacy of bilinear models. Bilinear analysis using the absorption spectra of tyrosine and copper metallothionein as targets resulted in a two-component solution which was nearly identical to that obtained using trilinear and quadrilinear models, for which no targets are required. The two-component models separate the absorption into tyrosine and copper center components. The absorption of tyrosine is found to be pH dependent in reduced plastocyanin, and the absorption magnitude of the reduced copper center is the same in the four different plastocyanin species. Further resolution is provided by a three-component quadrilinear model. The results indicate that there are at least two different electronic transitions which cause the absorption of the reduced copper center and that one of them couples to a tyrosine residue. It is the absorption of this coupled tyrosine residue which is pH dependent. Correlation of the results with previous studies indicates that it is Tyr 83 which is the perturbed residue. The separation of the absorption of the copper center and Tyr 83 provides spectroscopic probes for the conformations of the north pole and east face reaction sites on the plastocyanin protein.

Plastocyanin (PC)[4] is a protein located in the thylakoid lumen of plant chloroplasts, where it mediates electron transport between cytochrome *f* and Photosystem I (1, 2). PC contains a single copper atom, located near the surface of the protein, which participates in the electron transport reactions (3). The crystal structure of poplar PC has been determined by X-ray analysis (4–6). The conformation of the copper center is affected by the reduction of the copper atom, and the conformation of the reduced copper center is affected by a change in environmental pH. In contrast, the conformation of the protein portion of the molecule is not substantially affected by the changes in oxidation state and pH. However, the X-ray results do not rule out the possibility that other localized areas of the protein undergo small conformational changes in response to the large change in pH of the thylakoid lumen under different levels of illumination. Changes of this sort may influence the electron transporting activity of PC *in vivo*.

Two sites on the PC molecule have been identified for the binding of electron transporting reactants: These are where the copper center is exposed to the surface of the molecule (designated the north pole) and a conserved patch of carboxyl residues surrounding Tyr 83 (designated the east face) (4, 6). The reduced copper center and Tyr 83 chromophores are ideally located to be intrinsic spectroscopic probes of the local conformations of these two reactant binding sites. Thus, it is very desirable to be able to resolve the absorption of these chromophores in the near-uv absorption spectrum of PC.

Dr. Gross' group has been using near-uv (250–320 nm) absorption and circular dichroic spectroscopy to investigate the effects of pH and oxidation state on the conformations of different plant species of PC in solution (7–9). These studies have shown that the absorption of PC

[1] This research was supported by Grant 1 R01 GM38300 from the National Institutes of Health, and by grants of computing resources from the Ohio State University Instruction and Research Computer Center and from the Ohio Supercomputer Center.

[2] Present address: Max-Planck-Institut fur Strahlenchemie; Stiftstrasse 34-36; 4330 Mulheim a.d.Ruhr, FRG.

[3] To whom correspondence should be addressed.

[4] Abbreviations used: MT, metallothionein; PC, plastocyanin; RMS, root-mean-square; SVD, singular value decomposition.

0003-9861/90 $3.00

increases throughout the near-uv region of the spectrum upon reduction of the protein. This increase in absorption is pH dependent, being greater at pH 7.0 than at pH 5.0. The difference spectrum for the pH-dependent change resembles the spectrum of a tyrosine residue, but its magnitude is greater than that expected for an environmentally induced change in tyrosine extinction (9). The absorption change has a cooperative pH titration curve with a value of $n = 2$ (9).

The chromophores in the PC molecule which absorb in the near-uv region include the tyrosine residues (265–290 nm), the phenylalanine residues (250–265 nm), and the copper center in the reduced oxidation state (10). There are no tryptophan residues in higher plant PCs (1, 6). The absorption of tyrosine and phenylalanine amino acid residues in proteins is well characterized (11). In contrast, neither the band shape nor the types of electronic transitions responsible for the absorption of the reduced copper center are known (possibilities include Cu(I) → ligand charge transfer transitions and Cu(I) 3d → 4s Rydberg transitions) (10).

In this paper we demonstrate the use of bilinear, trilinear, and quadrilinear factor analysis to separate the components in the near-uv absorption spectrum of PC. The data are the absorption of PC as a function of four variables: wavelength, plant species, oxidation state, and environmental pH. Bilinear factor analysis is a traditional method for determining the independent components in a data set. There are many examples of bilinear factor analysis of spectroscopic measurements in the chemistry literature (12, 13), but only a few pertaining to the absorption of proteins (14–16).

A disadvantage of a bilinear model is that it does not uniquely define the responses of the components to the independent variables. This well-known problem is referred to as the "fundamental indeterminacy" or "rotation problem" (17). In the applications to protein absorption cited above, the components are determined by assuming a particular form for the response of the components to one of the independent variables. The ability to identify the components with bilinear factor analysis is dependent on the particular application, and by how much is previously known about the components. We present the bilinear factor analysis of the absorption spectra of PC partly to demonstrate the uncertainty in determining the components.

Unlike the bilinear model, the trilinear and quadrilinear models uniquely define the responses of the components to noninteracting independent variables. These models are said to have the "intrinsic axis property" (17). Therefore, trilinear and quadrilinear factor analyses provide a self-contained prediction for the components. Trilinear and quadrilinear factor analyses are applicable to any system where the response variable is a linear function of each of three or four noninteracting independent variables. Part of the goal of this paper is to determine whether the plant species, oxidation state, and environmental pH variables are noninteracting and thus suitable for trilinear and quadrilinear analysis.

The use of trilinear and quadrilinear models in factor analysis is still new (18, 19), especially for the analysis of spectroscopic data. There are a few examples of trilinear factor analysis of fluorescence measurements (20–22). To our knowledge, this is the first published application of trilinear and quadrilinear factor analysis of absorption spectroscopy. Due to the novelty of this approach, we are presenting the derivation of the different models as they apply to absorption measurements. In addition, a brief description of the algorithms developed by Dr. Ross' group for bilinear, trilinear, and quadrilinear factor analysis is provided.

THEORY

The absorption of light is commonly modeled as a bilinear product of two independent variables: the extinction coefficient and the concentration of a chromophore. For a sample with F different chromophores, Beer's law can be expressed as

$$a_i = D \sum_{f=1}^{F} \epsilon_{if} c_f, \qquad [1]$$

where a_i is the absorption of the sample at wavelength i, ϵ_{if} is the extinction coefficient of the f chromophore at wavelength i, c_f is the concentration of the f chromophore, and D is the pathlength of the light beam through the sample. The pathlength D will be omitted from subsequent equations.

In a protein the extinction coefficients and spectral forms of the chromophores may be affected by interactions with their local environments. In turn, these interactions may be affected by conformational changes of the protein, or by the direct influence of changes in the chemical condition of the protein and/or the surrounding solvent. In the models discussed in this paper, a chromophore with multiple spectral forms is treated as if it consisted of multiple chromophores with concentrations dependent on the chemical conditions. If the effects of different chemical-condition variables do not interact, then the chromophore concentrations can be separated into the product of terms representing the effects of each chemical-condition variable (i.e., a separation of variables). Taking into account the variation of one, two, and three noninteracting chemical-condition variables, the concentration term in Eq. [1] can be expressed as

$$c_{jf} = x_{jf} \qquad [2a]$$

or

$$c_{jkf} = x_{jf} y_{kf} \qquad [2b]$$

or

$$c_{jklf} = x_{jf} y_{kf} z_{lf}. \quad [2c]$$

Substituting Eq. [2] into Eq. [1] forms bilinear, trilinear, and quadrilinear models for the absorption measured as a function of two, three, and four independent variables, as indicated by i, j, k, and l:

$$a_{if} = \sum_{f=1}^{F} \epsilon_{if} x_{jf} \quad [3a]$$

$$a_{ijk} = \sum_{f=1}^{F} \epsilon_{if} x_{jf} y_{kf} \quad [3b]$$

$$a_{ijkl} = \sum_{f=1}^{F} \epsilon_{if} x_{jf} y_{kf} z_{lf}. \quad [3c]$$

The model of Eq. [3b] was proposed by Harshman (23) and by Carroll and Chang (24), and is referred to as a PARAFAC (parallel factors) or CANDECOMP (canonical decomposition) model in the psychometric literature (17–19, 25). The models of Eqs. [3a] and [3c] can be thought of as two- and four-variable equivalents, respectively, of the PARAFAC–CANDECOMP model.

Up to this point, the absorption has been described as the sum of contributions of all of the chromophores. However, some of the chromophores may be indistinguishable in terms of their spectra and/or the effects of the chemical-condition variables on their concentrations. To be distinguishable the absorption of the chromophores must respond uniquely to at least two of the independent variables (26). The components which are definable in the models of Eq. [3] differ in respect to at least two of the independent variables, and may be the superpositions of indistinguishable chromophores. Thus, the number of independent components in the data may be the same as the number of chromophores in the protein, but can be less.

Equation [3a] written in matrix notation becomes

$$A_{2\text{-way}} = EX^t, \quad [4]$$

where E is an I by F matrix, X is a J by F matrix, I and J are the total number of different absorption wavelengths and chemical conditions, F is redefined as the number of independent components, and t represents the matrix transpose operation. The column vectors of E are the wavelength-dependent extinction coefficients of the components (i.e., the spectra of the components) and the column vectors of X are the chemical-condition-dependent concentrations of the components.

The fundamental indeterminacy of bilinear models (17, 26) can be seen by the fact that $A_{2\text{-way}}$ is unaffected by the transformation of E and X by any nonsingular F by F matrix T and its inverse T^{-1}, respectively,

$$A_{2\text{-way}} = (ET)(T^{-1}X^t) = E'X'', \quad [5]$$

where E' and X' are the transformed matrices of E and X. The transformation of E and X affects both the shape of the spectra and the responses to the chemical condition of the components in the model.

Equations [3b] and [3c] can be expressed as the sum of the Kronecker products of the component vectors (21),

$$A_{3\text{-way}} = \sum_{f=1}^{F} E_f \otimes X_f \otimes Y_f \quad [6a]$$

and

$$A_{4\text{-way}} = \sum_{f=1}^{F} E_f \otimes X_f \otimes Y_f \otimes Z_f, \quad [6b]$$

where the subscripted f indicates the column vectors of the I, J, K, and L by F matrices E, X, Y, and Z, respectively, and $\otimes$ denotes the Kronecker product. The intrinsic axis property of these models is seen by the fact that the only possible transformations which do not affect $A_{3\text{-way}}$ and $A_{4\text{-way}}$ are permutation and scalar multiplication of the component vectors (17, 26). These transformations would affect the order and lengths of the component vectors, but not their orientations in euclidean space (17, 26). Thus, the spectra and chemical-condition-dependent responses of the real components in the data are defined by the trilinear and quadrilinear models.

EXPERIMENTAL PROCEDURES

The data used for the analysis in this paper were compiled from near-uv absorption spectra (250–320 nm) of PC previously presented by our group. The absorption spectra of poplar, spinach, and lettuce PC are from Draheim *et al.* (8), and the spectrum of parsley PC is from Durell *et al.* (9). To prepare the samples, PC was isolated according to standard methods (27, 28) and then further purified by a Pharmacia FPLC using a mono Q HR 5/5 anion-exchange column.

For each plant species of PC, the spectra of four different sample preparations were used. These were the spectra of 100% oxidized PC at pH 5.0 and 7.0, and 100% reduced PC at pH 5.0 and 7.0. Thus, the data were the absorption of PC as a function of four independent variables: the absorption wavelength, the plant species of PC, the oxidation state of the copper center, and the pH of the solution. The data were normalized so that the absorption could be expressed as the millimolar extinction coefficients ($mM^{-1}\ cm^{-1}$) of PC. The maximum error in the data was estimated to be 0.3 $mM^{-1}\ cm^{-1}$ by analysis of the variation in absorption of independently prepared PC samples (8).

For bilinear, trilinear, and quadrilinear factor analysis the variables were combined to express the absorption data as 2-way, 3-way, and 4-way arrays. A 2-way array corresponds to a rectangle, with two orthogonal coordinates. A 3-way array corresponds to a rectangular box, with three orthogonal coordinates. A 4-way array corresponds to a 4-dimensional analog of a rectangular box, with four orthogonal coordinates.

For bilinear analysis the wavelength variable was assigned to one coordinate of the 2-way array and the combination of the other three variables assigned to the other coordinate. This second coordinate has 16 values, which represent all the possible combinations of one of the four plant species of PC in one of the two oxidation states and at one of the two environmental pH values. Since this combined variable re-

fers only to chemical aspects of PC it is referred to as the chemical-condition variable.

For trilinear analysis the wavelength variable, the plant species variable, and the combination of the oxidation state and pH variables were assigned to the three coordinates of the 3-way data array. The combined oxidation state and pH coordinate has four values, which represent all the possible combinations of PC in one of the two oxidation states and at one of the two environmental pH values. For quadrilinear analysis the wavelength, plant species, oxidation state, and pH variables were each assigned to a separate coordinate of the 4-way data array.

All procedures were coded in Fortran 77. The programs FA2 and FA3 executed the bilinear analysis on an IBM 3081-D mainframe computer. IMSL subroutines (International Mathematical and Statistical Libraries, Houston, TX) were used for the singular value decomposition (SVD) and other matrix algebra operations. The programs PARAFAC3 and PARAFAC4 (22) executed the trilinear and quadrilinear analyses on a CRAY X-MP/24 supercomputer.

Bilinear analysis. The procedure for bilinear analysis has two main parts: derivation of a basis vector set of the absorption data, and transformation of the basis set to find the linear combinations which represent the components. The first part is accomplished by the closed-form SVD operation (29, 30). The resulting basis vectors are the same as the eigenvectors which are described in Principal Component Analysis (31).

Several methods have been proposed to determine the number of components in a bilinear model (12, 14, 32, 33). These methods exploit the property of the SVD operation to separate most of the real spectral information from the noise. When the basis vector pairs are sorted into descending order of the value of their associated singular values, each successive vector pair has less information. The real absorption components are generally contained in a subset of the leading basis vector pairs, and the remaining vectors pairs represent only noise.

The minimum number of leading basis vector pairs needed to satisfactorily model the original data is taken to be the number of real absorption components. Trial models of the data are formulated by multiplying increasing numbers of the leading basis vector pairs together. The quality of each trial model is judged from the residuals. The criteria for a satisfactory model are that the residuals appear random and that the residuals are smaller than the experimental error. The root-mean-square (RMS) statistic for each set of residuals is calculated using Eq. [7] for the number of degrees of freedom,

$$df(F)_{2\text{-way}} = (I - F)(J - F), \qquad [7]$$

where I and J are the number of rows and columns of the data array and F is the number of components. Use of least-squares fitting and of the RMS statistic is strictly appropriate only when the error in the data is normally distributed, of uniform variance, and uncorrelated. While this may not be true for the absorption data, these methods are commonly used in nonlinear least-squares parameter estimation (34).

Since both the real absorption component vectors and the SVD basis vectors are basis sets of the absorption data, the real absorption component vectors can be formed by linear combinations of the SVD basis vectors; see Eq. [5]. This procedure is referred to as a basis set transformation. Due to the fundamental indeterminacy of bilinear models, the correct transformation matrix, T, cannot be determined without including additional information about the real components. The criteria and constraints used to select the correct transformation matrix are dependent on the type of data being studied, and on how much is already known about the components.

One constraint, which was used in all cases, was that the predicted extinction coefficients and concentrations must be nonnegative. A second constraint, which was also used in all cases, was that the column vectors of the transformation matrix have a constant euclidean length. This normalization fixes a parameter in each column vector of T. In some instances, one or more entire column vectors of T were fixed by requiring the spectrum of a component to be as close as possible to the "target" spectrum of a known or suspected chromophore (12, 16, 35). Other researchers have used targets for the response of one or more components to changes in environmental conditions such as redox potential or pH (14). We used the spectrum of tyrosine in water (36, 37) as one target and the spectrum of the copper–protein metallothionein from *Neurospora* (38, 39) as another target. The tyrosine spectrum was shifted in wavelength to produce the best fit to the basis vectors.

The normalization and target constraints fix $F^2 - (F - 1)(F - n)$ elements of T, where n is the number of targets used. A nonlinear function minimizing routine (40, 41) was used to determine values for the remaining elements of T which comply with the nonnegativity constraint.

Trilinear and quadrilinear analyses. Unlike the bilinear model, the component vectors in the trilinear and quadrilinear models are uniquely defined (see Theory). Therefore, the component vectors were determined by finding the least-squares fit of each model to the data. This is expressed mathematically as the solution of Eq. [8]:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(a_{ijk} - \sum_{f=1}^{F}\hat{\epsilon}_{if}\hat{x}_{jf}\hat{y}_{kf})^2 = \text{minimum} \qquad [8a]$$

and

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L}(a_{ijkl} - \sum_{f=1}^{F}\hat{\epsilon}_{if}\hat{x}_{jf}\hat{y}_{kf}\hat{z}_{lf})^2 = \text{minimum} \qquad [8b]$$

Equation [8] was derived from Eqs. [3b] and [3c] and was solved by first taking the partial derivative with respect to each parameter (i.e., all of the ϵ_{if}, x_{jf}, y_{kf}, and z_{lf}). This set of nonlinear equations was then solved by an iterative minimization procedure described by Lee (22). For the cases where negative components resulted, the equations were solved using a nonnegativity constraint.

To determine the number of real components in the data, the quality of trial trilinear and quadrilinear models formulated with increasing numbers of F were examined in the same manner as for the bilinear model. Equations [9a] and [9b] were used to determine the number of degrees of freedom for the trilinear and quadrilinear models.

$$df(F)_{3\text{-way}} = IJK - F(I + J + K - 2) \qquad [9a]$$

$$df(F)_{4\text{-way}} = IJKL - F(I + J + K + L - 3) \qquad [9b]$$

RESULTS

Residual Analysis

Figure 1 displays the RMS residuals for the bilinear, trilinear, and quadrilinear models with one to five components. All of the one- and two-component models have an RMS residual larger than the maximum error in the PC absorption spectra (0.3 mM^{-1} cm^{-1}). The RMS residuals for the trilinear and quadrilinear models are not significantly affected by the application of the nonnegativity constraint. To a first approximation, the curves are biphasic, with cusps occurring for the models with three components. For models with the same number of components, the quality of the fit improves in the order of the quadrilinear, trilinear, and bilinear models.

Figure 2 displays the original-minus-fitted values, 2-way residual matrix of the three-component bilinear model. The residual matrices of the three-component trilinear and quadrilinear models, also arranged in 2-way arrays, are qualitatively similar. The change from two to

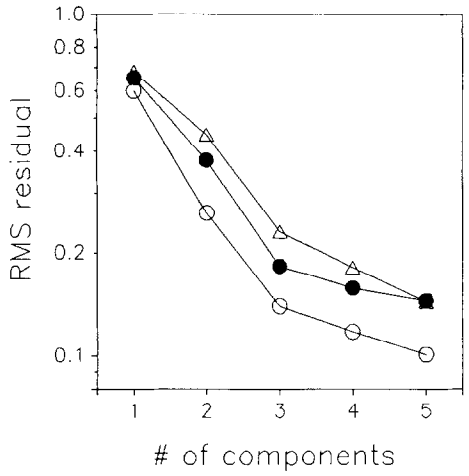

FIG. 1. Semi-logarithm plot of the RMS residual vs the number of components for the bilinear (○), trilinear (●), and quadrilinear (△) models. The trilinear and quadrilinear models were derived with the nonnegativity constraint.

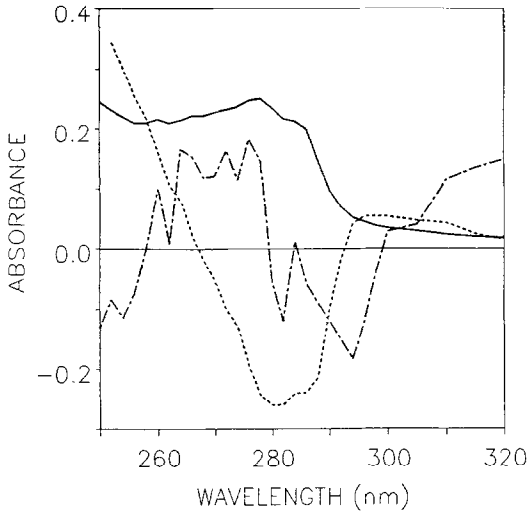

FIG. 3. The first (—), second (- - -) and third (– – –) wavelength-dependent basis vectors of the singular value decomposition of the 2-way absorption data array. The basis vectors are normalized to have a euclidean length of 1.

three components in the bilinear, trilinear, and quadrilinear models results in randomization of the residuals in the 250–296 nm wavelength region. However, a systematic deviation is still seen in the residuals in the 296–320 nm region. It should be noted, though, that the magnitude of the absorption of PC in this latter region of the spectrum is relatively small (see Fig. 6), and the magnitudes of the residuals are generally less than the maximum error in the spectra.

Bilinear Analysis

Figure 3 displays the first three SVD wavelength basis vectors of the 2-way absorption data array. The vectors are normalized to a euclidean length of 1. The singular values corresponding to these basis vectors are 124.90, 11.39, and 4.47, and the sum of all 16 singular values is 148.53. The first wavelength basis vector is the average of the 16 PC absorption spectra, and depicts recognizable features of the data. Evident in this vector is the tyrosine absorption band with the characteristic peak at 278 nm and shoulder at 284 nm (11). The small absorbance above 290 nm is not due to tyrosine, but is attributed to absorption of the copper center (8). The magnitude in the 250–265 nm region is due to the combination of absorption from the copper center, the tyrosine residues, and the phenylalanine residues. Absorption from the $n \rightarrow \pi^*$ transitions of the peptide bonds may also contribute to this region. The fine structure in this region is due to the absorption of the phenylalanine residues (8). The second basis vector has a negative peak at 280 nm and a shoulder at 286 nm, both of which are slightly red-shifted from the corresponding structures in the first basis vector. The third basis vector is less regular than the others, with oscillations occurring over relatively short wavelength intervals.

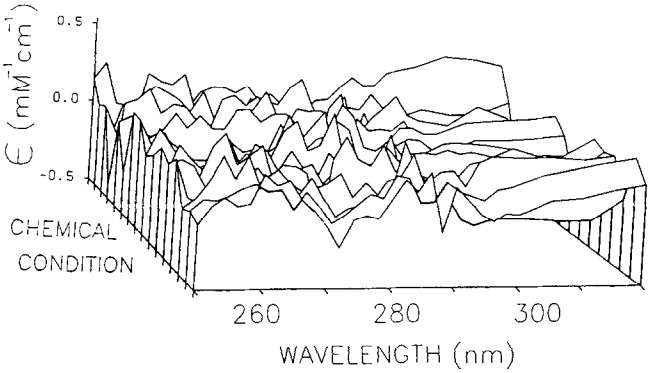

FIG. 2. Residual matrix (original − predicted spectra) for the three-component bilinear model.

Figure 4 displays the least-squares fit of linear combinations of the first two and three wavelength basis vectors to the tyrosine and *Neurospora* MT absorption targets. For both targets, inclusion of the third basis vector only slightly improves the quality of fit. The tyrosine target spectrum has been red-shifted 2.0 nm from the spectrum of tyrosine in water to have a wavelength maximum at 278 nm. This shift results in a 60% reduction of the RMS residuals for the two basis vector fit. The fitted tyrosine spectra show the characteristic tyrosine absorption peak and shoulder at 278 and 284 nm, respectively. The basis vectors reproduce the general form of the *Neurospora* MT target spectrum (Fig. 4B); however, there are significant deviations in the fine structure of these spectra. The fitted spectra do not have the peak of

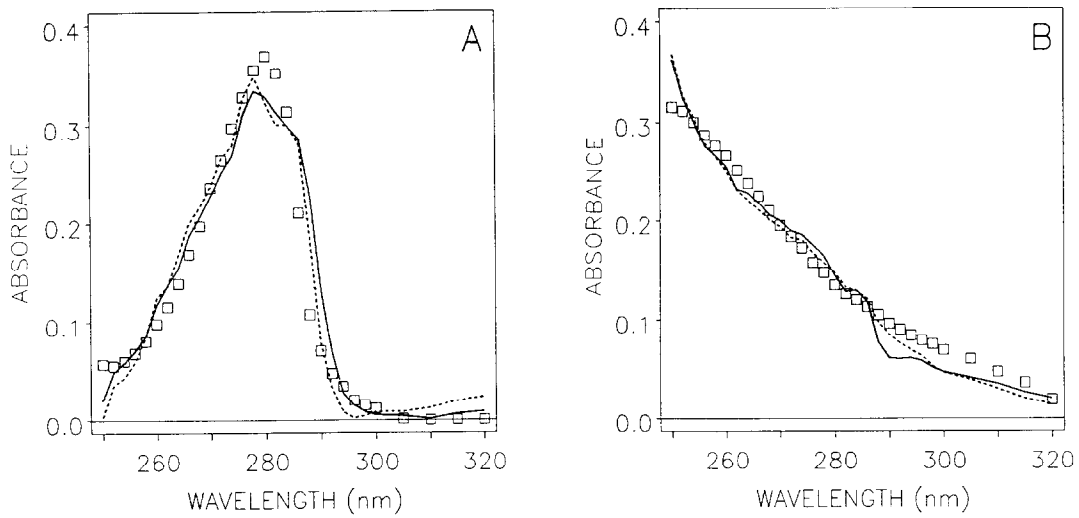

FIG. 4. Least-squares fits of linear combinations of the first two (—) and three (- - -) wavelength-dependent basis vectors to the (A) tyrosine and (B) *Neurospora* MT absorption targets (□). The spectra are normalized to have a euclidean length of 1.

the target spectrum at 250 nm, but instead have small peaks at 276 and 284 nm.

The attempt to determine the wavelength and chemical-condition-dependent component vectors by linear combinations of the first two and three SVD basis vector pairs using only the normalization and nonnegativity constraints results in a broad range of possible solutions (not shown). The variation is too large to derive any conclusions about the components. The additional use of one of the two absorption target criteria still does not provide for an interpretable range of solutions. For example, Fig. 5 displays the range of possible values for the nonfixed, wavelength-dependent component vector of a two basis vector pair transformation using the tyrosine target, normalization, and the nonnegativity constraints. In this case, only a single element of the transformation matrix is not fixed. When both the tyrosine and the MT targets are used to fix the first two columns of the transformation matrix, there is no nonnegative solution for the third component of a three-component bilinear model.

To provide a bilinear model for analysis, a two-component bilinear model is formed by using both the tyrosine and the *Neurospora* MT absorption targets. Since all of the elements of the transformation matrix are defined by the targets, this model is referred to as the "fixed bilinear model." The wavelength-dependent component vectors of the fixed bilinear model are the same as the two SVD basis vector fits to the targets shown in Fig. 4. Table I shows the predicted extinction coefficients at selected wavelengths for the two components of the fixed bilinear model in the 16 PC absorption spectra. The extinction of the tyrosine component varies among the four PC species according to the number of tyrosine residues in each (two in poplar PC, three in spinach and lettuce PC, and four in parsley PC). This can be most easily seen in the ⟨TYR⟩ column of Table I, in which the average tyrosine extinction coefficient for a particular oxidation state and pH shows little variation for the different PC species. For each PC species, the extinction coefficient of the tyrosine component is significantly greater in the reduced state at pH 7.0 than for the other three combinations of oxidation state and pH, for which there are smaller variations. The MT component is characterized by a greater magnitude in reduced PC than in oxidized PC, with smaller variations in each oxidation state due to changes of pH and PC species. Even in the oxidized state, however, the MT component contributes substan-

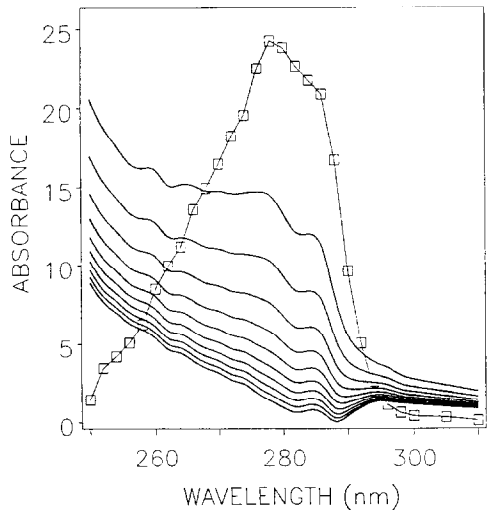

FIG. 5. The range of possible spectra for the second component of a two-component bilinear model which satisfies the nonnegativity criterion and fixes the spectrum of the first component (□) with the tyrosine absorption target. The spectra have an arbitrary euclidean length.

TABLE I

Predictions of the Fixed Bilinear Model Extinction Coefficients (mM^{-1} cm^{-1}) of the Tyrosine (TYR) and MT Components

	278 nm			260 nm	
PC species	TYR	⟨TYR⟩[a]	MT	TYR	MT
Poplar					
ox5	2.7	1.3	2.1	0.9	3.2
ox7	2.4	1.2	1.7	0.8	2.6
red5	2.6	1.3	4.6	0.9	7.1
red7	3.4	1.7	4.5	1.2	6.9
Spinach					
ox5	4.3	1.4	1.8	1.5	2.8
ox7	4.6	1.5	1.4	1.6	2.2
red5	4.2	1.4	4.1	1.5	6.3
red7	5.1	1.7	4.3	1.8	6.6
Lettuce					
ox5	4.0	1.4	1.9	1.4	2.9
ox7	3.9	1.3	1.7	1.4	2.6
red5	4.1	1.4	4.5	1.4	6.9
red7	4.5	1.5	4.1	1.6	6.3
Parsley					
ox5	6.1	1.5	1.6	2.2	2.5
ox7	5.7	1.4	1.8	2.0	2.7
red5	5.7	1.4	4.6	2.0	7.1
red7	7.7	1.9	5.2	2.7	8.1

[a] The ⟨TYR⟩ label denotes the average extinction coefficients of the tyrosine residues determined by dividing the value for TYR by the number of tyrosine residues in that PC species (two in poplar PC, three in spinach and lettuce PC, and four in parsley PC).

tially to the absorption of PC in the main tyrosine absorption region (265–290 nm). Using the prediction for oxidized spinach PC at pH 7.0 as an example, the MT component contributes 24% to the net absorption at 278 nm.

Figure 6 displays the predicted effects of the oxidation state and pH variables on the two components in the spectrum of spinach PC. At pH 5.0 the increase in extinction of the PC spectrum upon reduction is due solely to the increase of the MT component: the reduced minus oxidized difference extinction coefficients at 278 nm of the tyrosine and MT components are −0.1 and 2.3 mM^{-1} cm^{-1}, respectively. At pH 7.0 the tyrosine component contributes 15% to the increase in the PC spectrum upon reduction: The reduced minus oxidized difference extinction coefficients at 278 nm of the tyrosine and MT components are 0.5 and 2.9 mM^{-1}, respectively. A similar pattern occurs for the other species. In the spectra of poplar, lettuce, and parsley PC at pH 7.0, the tyrosine component contributes 27, 20, and 37%, respectively, to the total increase of extinction at 278 nm upon reduction. In the 250–265 nm region of the PC spectrum the increase of extinction is due primarily to the increase of the MT component at both pH 5.0 and pH 7.0.

Trilinear and Quadrilinear Analyses

Two-component models. Without the nonnegativity constraint, the spectra of the tyrosine components in the two-component trilinear and quadrilinear models were negative in the 250–255 nm wavelength region. With this constraint, the two components of these models are nearly identical to the tyrosine and MT components of the fixed bilinear model. The extinction coefficients of the components in the constrained trilinear and quadrilinear models are similar to those of the fixed bilinear model, and also respond similarly to the changes of the PC species, oxidation state, and pH variables (data not shown).

Three-component models. The nonnegativity constraint was also required for the three-component trilinear model. The spectra of the components of the trilinear and quadrilinear models are similar to each other. However, in the constrained trilinear model the predicted magnitude of the MT component is zero in the spectrum of parsley PC. Since this is considered to be an invalid result, the three-component trilinear model is not considered further.

Figure 7 displays the wavelength-dependent component vectors of the three-component quadrilinear model. The components of this model are designated as the tyrosine, MT, and 3rd components based on correspondence to the two-component models. The spectrum of the tyrosine component is nearly identical to the spectrum of the tyrosine components in the two-component models. The spectrum of the MT component has the same shoulders at 278 and 284 nm as in the two-component models, but drops sharply at 292 nm to form a second, broad band in the 294–320 nm region. The spectrum of the 3rd component appears to be a mixture of the tyrosine and MT components, showing a sharp decrease in the 250–260 nm region (similar to the MT component spectrum) and a peak and shoulder at 280 and 286 nm (which are red-shifted 2.0 nm from the tyrosine component spectrum). The peaks of this component correspond to the minima in the second SVD basis vector, which is used for bilinear analysis.

Table II shows the predicted extinction coefficients at selected wavelengths of the three components of the three-component quadrilinear model. The extinction coefficients of both the MT and the 3rd components are considerably greater for reduced PC than for oxidized PC. In each oxidation state the magnitude of the MT component is greater at pH 5.0 than at pH 7.0, and the magnitude of the 3rd component is greater at pH 7.0 than at pH 5.0. The effects of oxidation state and pH changes on the magnitude of the tyrosine component are significantly different from that observed for the fixed bilinear model. In particular, the magnitude of the tyrosine component is larger in the spectrum of oxidized PC than in that of reduced PC, and in each oxidation state,

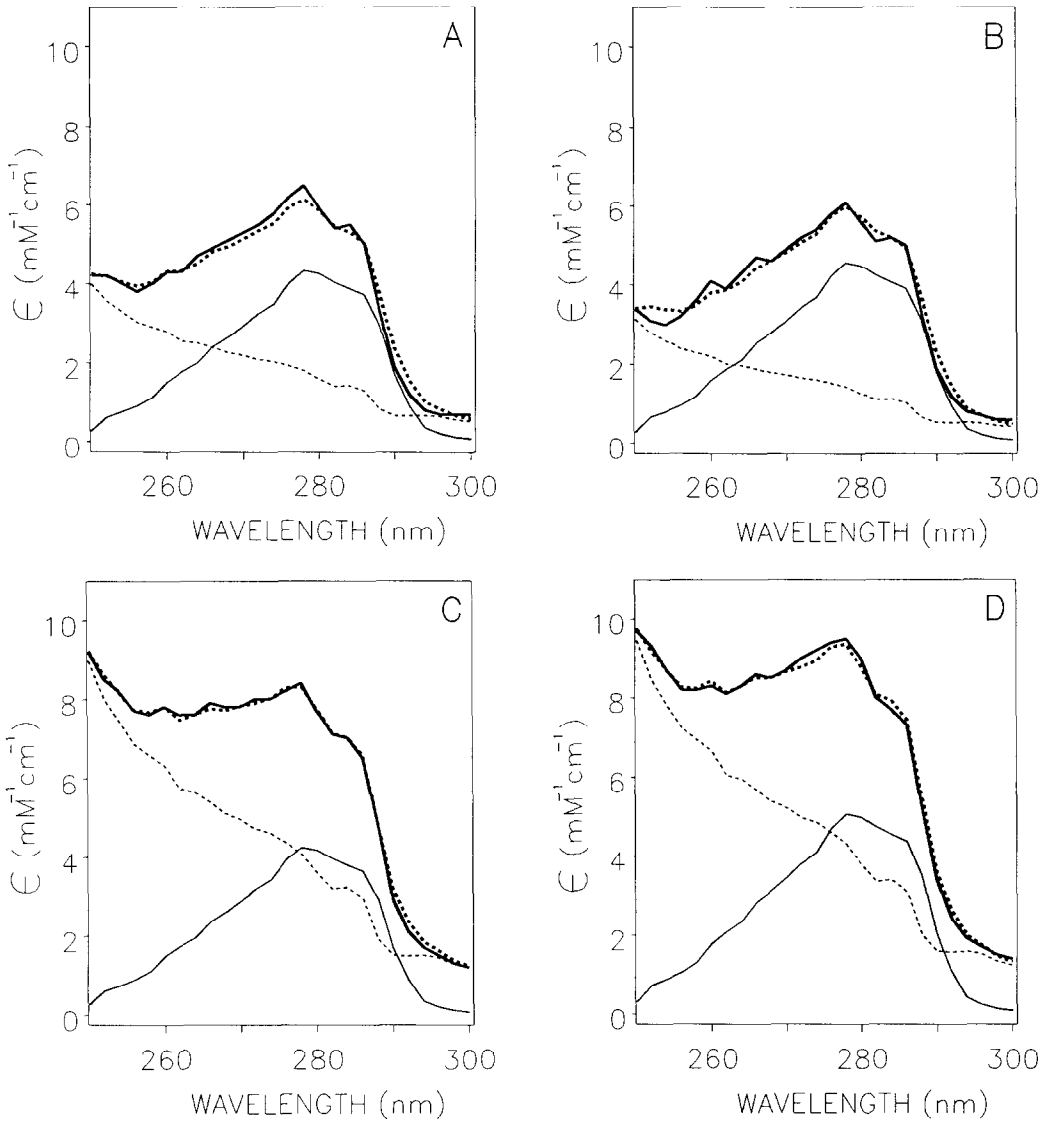

FIG. 6. The original (——), predicted (· · ·), tyrosine component (—), and MT component (- - -) spectra of the fixed bilinear model of the absorption of spinach PC. (A) Oxidized PC at pH 5.0, (B) oxidized PC at pH 7.0, (C) reduced PC at pH 5.0, and (D) reduced PC at pH 7.0.

larger at pH 5.0 than at pH 7.0. However, the response of the sum of the tyrosine and 3rd components is similar to that of the tyrosine component in the fixed bilinear model.

Figure 8 displays the predicted effects of the oxidation state and pH variables on the three components in the spectrum of spinach PC. At both pH 5.0 and pH 7.0, the increase in extinction of the PC spectrum upon reduction is due to the combined increase of the MT and 3rd components. The reduced minus oxidized difference extinction coefficients at 278 nm of the MT and 3rd components are 1.6 and 2.4 mM^{-1} cm^{-1} at pH 5.0 and 1.2 and 3.7 mM^{-1} cm^{-1} at pH 7.0, respectively. Reduction of PC results in a decrease of extinction of the tyrosine component. The magnitude of this decrease is the same at pH 5.0 as at pH 7.0. In the spectra of poplar, spinach, lettuce, and parsley PC at pH 7.0, the reduced minus oxidized difference extinction coefficients at 278 nm of the tyrosine component are −0.8, −1.7, −1.5, and −2.2 mM^{-1} cm^{-1}, respectively.

DISCUSSION

Residual Analysis

Two main results indicate that there are three components in the data: (a) Three components are required to reduce the RMS residuals below the maximum magnitude of error in the absorption spectra; (b) the rate of RMS residual reduction with the incorporation of greater numbers of components is biphasic, and decreases for more than three components in the model (Fig. 1). The fact that this same behavior occurs in the

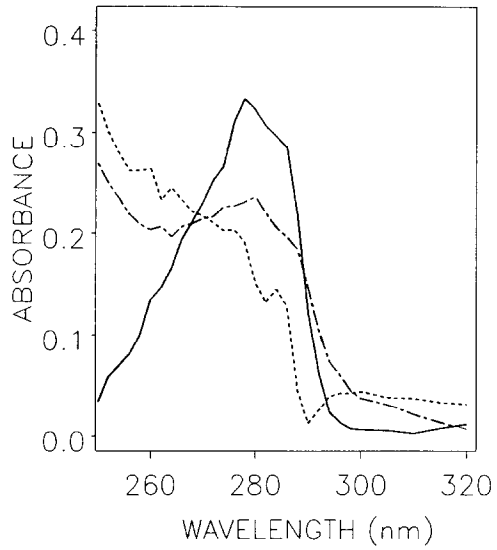

FIG. 7. The wavelength-dependent parameter vectors of the tyrosine (—), MT (- - -), and 3rd (– – –) components of the three-component quadrilinear model. The vectors are normalized to have a euclidean length of 1.

bilinear, trilinear, and quadrilinear models is further confirmation. The first, rapid phase of the RMS residual reduction is interpreted as fitting both the real components and the noise in the data; the second, slow phase is interpreted as fitting only the noise.

The observation that the nonnegativity constraint changes the spectra and extinction coefficients of the components of the trilinear and quadrilinear models without significantly altering the RMS residuals indicates that the components are not completely defined. This is due to error in the data and interaction among the effects of the variables. It is clear that the absorption wavelength variable (i.e., the setting of the monochrometer) does not interact with the chemical-condition variables. For models with the same number of components, the progressive increase of the RMS residuals for the bilinear, trilinear, and quadrilinear models indicates that the PC species, oxidation state, and pH variables do interact. However, the facts that the increase of the RMS residual is relatively small and that similar components occur in the three types of models indicate that these variables are sufficiently noninteractive to validate the use of the trilinear and quadrilinear models.

Absorption Component Analysis

Two-component models. Due to the fundamental indeterminacy, supplemental information must be used in bilinear analysis to determine the components. The accuracy of the predicted components depends on the choice of constraints. Since real absorption components are always positive, the nonnegativity constraint must not falsely bias the results. If there is an absorption region where only a single component absorbs, then the nonnegativity criterion alone will fully define all of the components (12). However, the components in the near-uv spectrum of PC are heavily superimposed and absorption targets are required.

The tyrosine target was selected because it is known that tyrosine absorption is a major component in the 265–290 nm wavelength region of the PC spectrum (7). The *Neurospora* MT target (38, 39) was selected to approximate the absorption component due to the copper center in the spectrum of reduced PC. This species of MT contains six reduced copper atoms bound by seven cysteine residues. The coordination geometry of the copper atoms is not known, but in yeast MT the copper atoms have a tetrahedral coordination (42). The near-uv absorption of *Neurospora* MT is due solely to electronic transitions involving the copper atoms, since the protein is devoid of aromatic amino acid residues. These transitions may be analogous to transitions in the copper center of PC. This suggests that charge transfer transitions between the copper atom and the sulfur atoms of the Cys 84 and Met 92 ligands are responsible for the copper center absorption in reduced PC (9). The appropriateness of these targets for a two-component model, and thus the accuracy of the predicted components of the fixed bilinear model, is indicated by the fact that nearly identical components occur in the two-component trilinear and quadrilinear models, for which outside criteria are not used.

The small deviations between the targets and the fitted spectra (Fig. 4) provide some insight into the chromophores in PC. The fact that the absorption band of the fitted tyrosine component is broader than the absorption band of the target suggests that the spectra of the different tyrosine residues are slightly shifted due to environmental perturbations. The deviation of the spectrum of the MT component from the 250-nm peak in the MT target may be because this band is blue-shifted in the spectrum of PC. The possible variability of the wavelength maximum of this transition is indicated by the fact that it is shifted to 280 nm in the spectrum of yeast MT (43). The shoulder peaks at 278 and 284 nm in the MT component spectrum are probably due to a small amount of tyrosine absorption, and not a predicted feature of the absorption of the reduced copper center.

The fixed bilinear model is used to analyze the tyrosine and MT components of the two-component models because it is free of the error caused by the interaction of the chemical-condition variables. The tyrosine component is interpreted as the superimposed absorption of the tyrosine residues in PC. This is supported by three results: (a) The spectrum of the tyrosine component is very close to that of a tyrosine residue; (b) the magnitude of the tyrosine component corresponds to the number of tyrosine residues in each PC species; (c) the average

TABLE II

Predictions of the Three-Component Quadrilinear Model Extinction Coefficients ($mM^{-1}\ cm^{-1}$) of the Tyrosine (TYR), MT, and 3rd Components

	278 nm						260 nm		
PC species	TYR	⟨TYR⟩[a]	MT	3rd	$\overset{\text{TYR}}{\underset{+3rd}{}}$	$\left\langle\overset{\text{TYR}}{\underset{+3rd}{}}\right\rangle$	TYR	MT	3rd
Poplar									
ox5	2.1	1.0	2.0	0.6	2.7	1.4	0.8	2.8	0.5
ox7	1.9	0.9	1.5	0.9	2.8	1.4	0.7	2.1	0.8
red5	1.1	0.6	4.1	2.5	3.6	1.8	0.4	5.6	2.2
red7	1.0	0.5	3.0	3.9	4.9	2.5	0.4	4.2	3.4
Spinach									
ox5	4.1	1.4	1.5	0.7	4.8	1.6	1.6	2.1	0.6
ox7	3.7	1.2	1.1	1.1	4.8	1.6	1.4	1.6	1.0
red5	2.2	0.7	3.1	3.2	5.4	1.8	0.9	4.3	2.8
red7	2.0	0.7	2.3	4.9	6.9	2.3	0.8	3.2	4.3
Lettuce									
ox5	3.6	1.2	2.0	0.6	4.2	1.4	1.4	2.7	0.5
ox7	3.3	1.1	1.5	0.9	4.2	1.4	1.3	2.0	0.7
red5	2.0	0.7	4.0	2.4	4.4	1.5	0.8	5.6	2.1
red7	1.8	0.6	3.0	3.7	5.5	1.8	0.7	4.1	3.2
Parsley									
ox5	5.3	1.3	1.0	1.3	6.6	1.7	2.1	1.4	1.1
ox7	4.8	1.2	0.8	1.9	6.7	1.7	1.9	1.1	1.7
red5	2.9	0.7	2.1	5.4	8.3	2.1	1.2	2.9	4.7
red7	2.6	0.7	1.6	8.3	10.9	2.7	1.0	2.1	7.3

[a] The ⟨TYR⟩ label is the same as in Table I. $\overset{\text{TYR}}{\underset{+3rd}{}}$ denotes the sum of the TYR tyrosine and 3rd components, and $\left\langle\overset{\text{TYR}}{\underset{+3rd}{}}\right\rangle$ denotes the average of this value determined analogously to ⟨TYR⟩.

extinction coefficients at 278 nm of the tyrosine residues (see Table I) correspond to the extinction coefficients of tyrosine in water and in ethanol (1.4 and 1.8 $mM^{-1}\ cm^{-1}$, respectively) (37). An important prediction of the two-component models is that the net tyrosine absorption is pH dependent when the copper center is reduced. This is in accord with an earlier study of the absorption spectrum of PC using different methods (9). The change in absorption suggests that there is a pH-dependent conformational change at the site of at least one of the tyrosine residues that operates only in the reduced form of PC.

In oxidized PC the MT component is interpreted as the superimposed absorption of the phenylalanine residues and the tail ends of the peptide bond chromophores. The increase in magnitude of the MT component upon reduction is interpreted as due to the absorption of the reduced copper center. The fact that this change in magnitude is nearly the same in the four PC species indicates that the structure of the copper center is also nearly the same in these four PC species.

Three-component model. The three-component quadrilinear model further resolves the absorption of the tyrosine residues and the copper center. The interpretation of the three components is as follows. The tyrosine component, which has a peak at 278 nm with a shoulder at 284 nm, is due to the absorption of unperturbed tyrosine residues. The MT component is due to the absorption of the phenylalanine, peptide bond, and copper center chromophores (as in the two-component models). The 3rd component is due to the absorption of a second copper center electronic transition and a perturbed tyrosine residue.

The interpretation of the 3rd component is based on the fact that the spectrum appears to be a superposition of the MT and tyrosine components. The tyrosine part of this component is interpreted as being perturbed because the peaks are red-shifted 2 nm from a normal tyrosine spectrum. It is known that changes in environmental polarity and hydrogen bonding can shift the absorption spectrum of tyrosine by as much as 4 nm (11, 36, 44). One possible explanation for the 3rd component is that it is the absorption of coupled electronic transitions belonging to the copper center and a tyrosine residue. Electronic coupling would also explain the small red-shift of the tyrosine absorption maxima. This leads to the interpretation that there are at least two separate electronic transitions which cause the absorption of the reduced copper center: one that belongs to the MT component and one that couples to a tyrosine residue. The

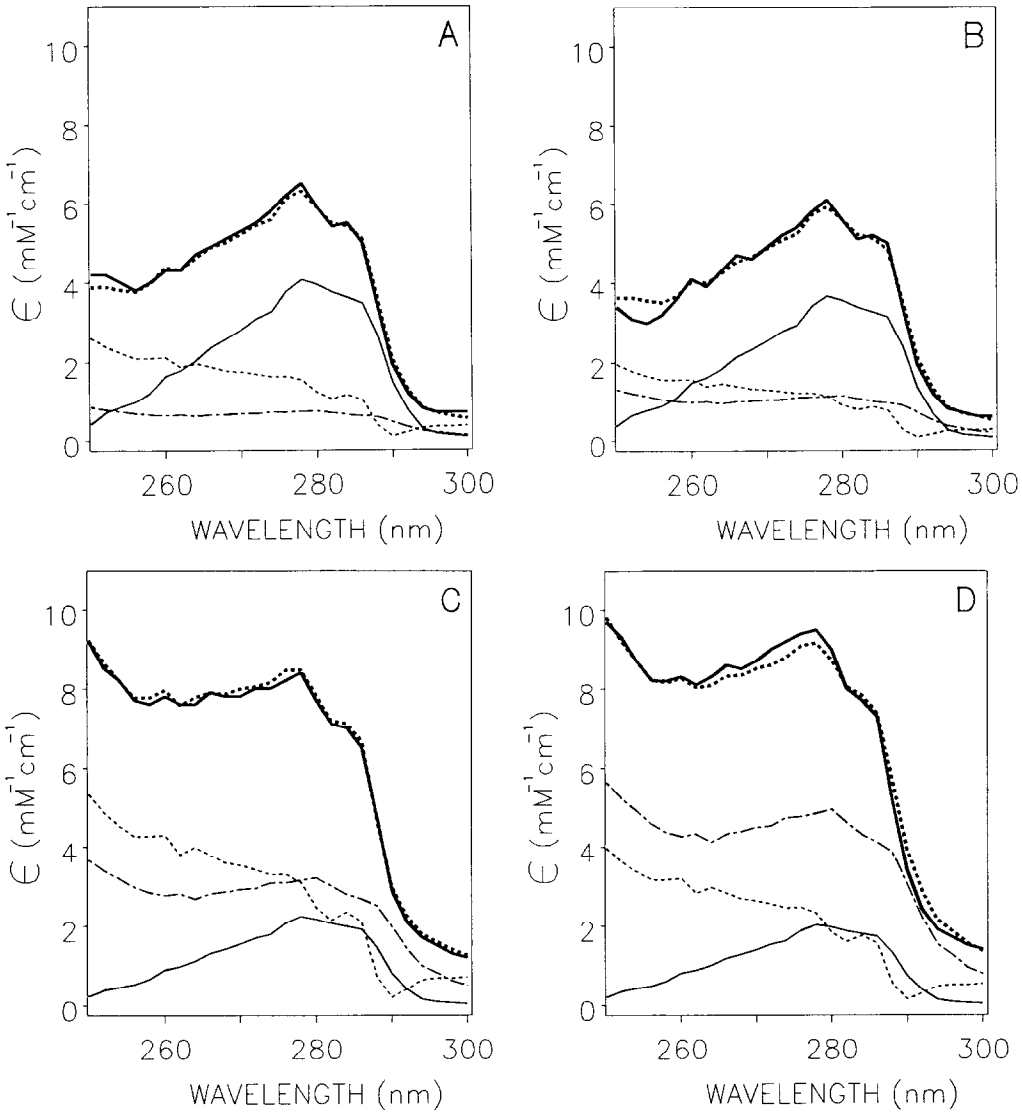

FIG. 8. The original (——), predicted (· · ·), tyrosine component (—), MT component (- - -), and 3rd component (– – –) spectra of the three-component quadrilinear model of the absorption of spinach PC. (A) Oxidized PC at pH 5.0, (B) oxidized PC at pH 7.0, (C) reduced PC at pH 5.0, and (D) reduced PC at pH 7.0.

coupled transition hypothesis would also explain the oxidation state-dependent behavior of the tyrosine component. The coupling of a tyrosine residue would decrease the number of unperturbed tyrosine residues, and thus decrease the magnitude of the tyrosine component in the reduced form of PC. In conflict with this hypothesis is the fact that the 3rd component is also present in the spectrum of oxidized PC. However, the magnitude is considerably less in this oxidation state, and may be due to the error introduced by the interaction of the chemical-condition variables.

The results indicate that it is only the perturbed residue in the 3rd component which is responsible for the pH dependence of the tyrosine absorption in reduced PC (discussed above for the two-component models). This may be due to the pH-dependent change in conformation of the reduced copper center (4, 5) and/or a change in the environment of the perturbed tyrosine residue (11, 44). It has been noted previously that the pH-dependent change in tyrosine absorption in reduced PC is larger than would be expected from model studies of environmental effects on tyrosine absorption (9, 10). However, this phenomenon can be explained by an added pH-dependent change in the absorption of the copper center transition and/or by the exchange of oscillator strength between the two coupled transitions.

The 3rd component is observed in all four of the PC species, including poplar, which only has two tyrosine residues (numbers 80 and 83). Tyr 83 is more likely to be the coupled residue, because it is located substantially

closer to the copper center (10 vs 20 Å) at the east face binding site. Due to sequence homology, this would also be the perturbed residue in the other PC species. The substantially larger magnitude of the 3rd component in the spectrum of reduced parsley PC (see Table II) indicates that Tyr 62 may also be involved in the coupled complex. Previously it has been described that the tyrosine absorption in reduced PC titrates according to the coupled pH-dependent activity of electron transport through the east face site (9). This supports the assignment of Tyr 83 as the perturbed residue, because it is located next to the titratable acidic residues at the east face site.

CONCLUSIONS

Both a bilinear model and the newer trilinear and quadrilinear models provide useful insight into the near-uv absorption spectrum of PC.

Bilinear models of absorption have the advantage of using two variables that are always noninteracting: wavelength and chemical condition. Their disadvantage of needing substantial supplementary information to determine the components was overcome by the successful use of tyrosine and metallothionein targets to define a two-component model.

Although trilinear and quadrilinear models do not require supplementary information, they have the disadvantage of requiring that the different chemical variables not interact. That some interaction does occur is indicated by the increase in the residual error in the trilinear and quadrilinear models compared to the bilinear model.

The similarity of the two-component spectra from the fixed bilinear model and from the trilinear and quadrilinear models helps validate the results from each.

One result of the two-component models is that the copper center has the same conformation in the four different PC species. Another result, in correspondence with previous findings, is that the tyrosine absorption in reduced PC is pH dependent. This indicates the presence of conformational changes around one or more of the tyrosine residues.

The three-component model indicates that there are at least two electronic transitions which cause the absorption of the reduced copper center. The results support the hypothesis that one of these transitions is electronically coupled to Tyr 83. In turn, this coupled entity is responsible for the pH dependence of the tyrosine absorption observed in reduced PC. The resolution of the absorption of the copper center and Tyr 83 chromophores from the superimposed near-uv spectrum of PC provides a useful probe of the north pole and east face reaction sites.

ACKNOWLEDGMENTS

We thank Dr. Sue Leurgans for her assistance in understanding multilinear models and related topics and Dr. James Draheim for the absorption spectra of PC.

REFERENCES

1. Boulter, D., Haslett, B. G., Peacock, D., Ramshaw, J. A. M., and Scawan, M. D. (1977) *Int. Rev. Biochem.* **13,** 1–40.
2. Haehnel, W., Propper, A., and Krause, H. (1980) *Biochim. Biophys. Acta* **593,** 384–399.
3. Freeman, H. C. (1981) *Coord. Chem.* **21,** 29–51.
4. Guss, J. M., and Freeman, H. C. (1983) *J. Mol. Biol.* **169,** 521–563.
5. Guss, J. M., Harrowell, P. R., Murata, M., Norris, V. A., and Freeman, H. C. (1986) *J. Mol. Biol.* **192,** 361–387.
6. Sykes, A. G. (1985) *Chem. Soc. Rev.* **14,** 283–315.
7. Draheim, J. E., Anderson, G. P., Duane, J. W., and Gross, E. L. (1986) *Biophys. J.* **49,** 891–900.
8. Draheim, J. E., Anderson, G. P., Pan, R. L., Rellick, L. M., Duane, J. W., and Gross, E. L. (1985) *Arch. Biochem. Biophys.* **237,** 110–117.
9. Durell, S. R., Draheim, J. E., and Gross, E. L. (1988) *Arch. Biochem. Biophys.* **267,** 217–227.
10. Tamilarasan, R., and McMillin, D. R. (1986) *Inorg. Chem.* **25,** 2037–2040.
11. Donovan, J. W. (1969) *in* Physical Principles and Techniques of Protein Chemistry (Leach, S. J., Ed.), Part A, pp. 101–170, Academic Press, New York.
12. Malinowski, E. R., and Howery, D. G. (1980) Factor Analysis in Chemistry, Wiley-Interscience, New York.
13. Ramos, L. S., Beebe, K. R., Carey, W. P., Sanchez, M. E., Erickson, B. C., Wilson, B. E., Wangen, L. E., and Kowalski, B. R. (1986) *Anal. Chem.* **58,** 294R–315R.
14. Shrager, R. I., and Hendler, R. W. (1982) *Anal. Chem.* **54,** 1147–1152.
15. Hofrichter, J., Sommer, J. H., Henry, E. R., and Eaton, W. A. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2235–2239.
16. Halaka, F. G., Babcock, G. T., and Dye, J. L. (1985) *Biophys. J.* **48,** 209–219.
17. Kruskal, J. B. (1983) *Proc. Symp. Appl. Math.* **28,** 75–104.
18. Kroonenberg, P. M. (1983) Three-Mode Principal Component Analysis: Theory and Applications, DSWO Press, Leiden.
19. Kroonenberg, P. M. (1983) *Brit. J. Math. Stat. Psych.* **36,** 81–113.
20. Appellof, C. J., and Davidson, E. R. (1981) *Anal. Chem.* **53,** 2053–2056.
21. Callis, J. B. (1984) *in* Ultrahigh Resolution Chromotography, American Chemical Society Symposium Series No. 250 (Ahuja, S., Ed.), pp. 171–198, American Chemical Society.
22. Lee, C.-H. (1988) Ph.D. Thesis, The Ohio State University.
23. Harshman, R. A. (1970) *UCLA Work. Pap. Phonetics* **16,** 1–84.
24. Carroll, J. D., and Chang, J. J. (1970) *Psychometrika* **35,** 283–319.
25. Law, H. G., Snyder, Jr., C. W., Hattie, J. A., and McDonald, R. P. (1984) Research Methods for Multimode Data Analysis, Praeger, New York.
26. Kruskal, J. B. (1977) *Linear Alg. Appl.* **18,** 95–138.
27. Davis, D. J., and San Pietro, A. (1979) *Anal. Biochem.* **95,** 254–259.
28. Graziani, M. T., Agro, A. F., Rotilio, G., Barra, D., and Mondovi, B. (1974) *Biochemistry* **13,** 804–809.

29. Marchiarullo, M. A., and Ross, R. T. (1985) *Biochim. Biophys. Acta* **807,** 52–63.

30. Golub, G. H., and Reinsch, C. (1970) *Numer. Math.* **14,** 403–420. [Reprinted *in* Handbook for Automatic Computation, II, Linear Algebra, Springer-Verlag, New York.]

31. Cochran, R. N., and Horne, F. H. (1977) *Anal. Chem.* **49,** 846–853.

32. Warner, I. M., Christian, G. D., Davidson, E. R., and Callis, J. B. (1977) *Anal. Chem.* **49,** 564–573.

33. Lin, C.-H., and Liu, S.-H. (1978) *J. Chin. Chem. Soc.* **25,** 167–177.

34. Bates, D. M., and Watts, D. G. (1988) Non-linear Regression Analysis And Its Applications, Wiley, New York.

35. Cochran, R. N., Horne, F. H., Dye, J. L., Ceraso, J., and Suelter, C. H. (1980) *J. Phys. Chem.* **84,** 2567–2575.

36. Bailey, J. E., Beaven, G. H., Chignell, D. A., and Gratzer, W. B. (1968) *Eur. J. Biochem.* **7,** 5–14.

37. Gratzer, W. B. (1968) *in* Handbook of Biochemistry (Sober, H. A., Ed.), p. B-18, The Chemical Rubber Co., Cleveland.

38. Beltramini, M., and Lerch, K. (1981) *FEBS Lett.* **127,** 201–203.

39. Beltramini, M., and Lerch, K. (1986) *Environ. Health Perspect.* **65,** 21–27.

40. Nelder, J. A., and Mead, R. (1964) *Comput. J.* **7,** 308.

41. Box, M. J. (1965) *Comput. J.* **8,** 42.

42. Weser, U., and Hartmann, H.-J. (1985) *in* Copper Proteins and Copper Enzymes, (Lontie, R. Ed.), Vol. III, pp. 151–173, CRC Press, Boca Raton.

43. Rupp, H., Cammack, R., Hartmann, H.-J., and Ulrich, W. (1979) *Biochim. Biophys. Acta* **578,** 462–475.

44. Strickland, E. H., Wilchek, M., Horwitz, J., and Billups, C. (1972) *J. Biol. Chem.* **247**(2), 572–580.