Metadata of the chapter that will be visualized online

Chapter Title	Quantitative Fluorescence Spectral Analysis of Protein Denaturation	
Copyright Year	2013	
Copyright Holder	Springer Science+Business Media, LLC	
Corresponding Author	Family Name	**Stokkum**
	Particle	**van**
	Given Name	**Ivo H.M.**
	Suffix	
	Division	Department of Physics and Astronomy
	Organization	Faculty of Sciences, VU University (VU) Amsterdam
	Address	De Boelelaan 1081, 1081 HV, Amsterdam, The Netherlands
Author	Family Name	**Laptenok**
	Particle	
	Given Name	**Sergey P.**
	Suffix	
Abstract	This chapter describes a procedure of global analysis of the steady-state spectra measured with different concentrations of the denaturant to quantitatively study protein denaturation. With the help of physicochemical models, relevant spectral parameters that characterize the folding intermediate and thermodynamic parameters that describe a three-state model N ⇔ I ⇔ U can be estimated.	
Key words (separated by "-")	Global analysis - Protein denaturation - Singular value decomposition - Steady-state fluorescence	

Chapter 3

Quantitative Fluorescence Spectral Analysis of Protein Denaturation

Ivo H.M. van Stokkum and Sergey P. Laptenok

Abstract

This chapter describes a procedure of global analysis of the steady-state spectra measured with different concentrations of the denaturant to quantitatively study protein denaturation. With the help of physicochemical models, relevant spectral parameters that characterize the folding intermediate and thermodynamic parameters that describe a three-state model N ⇔ I ⇔ U can be estimated.

Key words Global analysis, Protein denaturation, Singular value decomposition, Steady-state fluorescence

1 Introduction

During denaturant-induced equilibrium (un)folding of a particular protein, a molten globule-like folding intermediate is formed [1]. Here we describe how the steady-state fluorescence spectrum monitored as a function of the denaturant concentration can be used to infer the properties of the folding intermediate with the help of global analysis [2].

2 Materials

Materials, Protein Expression, and Purification. All chemicals used were of the highest purity available. The concentration guanidine hydrochloride (GuHCl) was determined by measuring the refractive index of the sample used, as described previously [3]. A variant of apoflavodoxin from *Azotobacter vinelandii*, which contains two tryptophan residues (i.e., W74-W128-F167 (WWF)), was obtained and purified as described [3]. In all experiments protein concentration was 4 μM in 100-mM potassium pyrophosphate buffer, pH = 6.0. Temperature was set to 25 °C.

Yves Engelborghs and Antonie J.W.G. Visser (eds.), *Fluorescence Spectroscopy and Microscopy: Methods and Protocols*, Methods in Molecular Biology, vol. 1076, DOI 10.1007/978-1-62703-649-8_3,

Steady-State Fluorescence Spectra. Steady-state fluorescence spectra were obtained with a Fluorolog 3.2.2 spectrofluorometer (Horiba, Jobin Yvon, Optilas, Alphen aan den Rijn, the Netherlands), as described previously [3]. The excitation wavelength was 300 nm, excitation and emission slit widths were 2 nm, and emission spectra were recorded between 305 and 400 nm with 1-nm steps. All spectra were corrected for wavelength-dependent instrumental response characteristics. Background fluorescence emission was measured under the same circumstances, except that now no protein is present in the samples, and was subsequently subtracted from the corresponding fluorescence spectra of samples with protein.

3 Methods

3.1 Determination of the Number of Components Contributing to the Steady-State Fluorescence Spectra

The steady-state spectra measured at n_d denaturant concentrations can be collated in a matrix Ψ where a column of the $n_\lambda \times n_d$ matrix Ψ contains a spectrum measured at n_λ wavelengths at a particular denaturant concentration, whereas a row contains the emission measured at a particular wavelength at n_d denaturant concentrations. The rank of this matrix Ψ can be estimated with the help of the singular value decomposition (SVD) [4–11] (*see* **Note 1**).

Figure 1 depicts the results from the singular value decomposition (SVD) analysis of the data matrix Ψ obtained from denaturant-induced unfolding of WWF apoflavodoxin. The scree plot of the singular values (Fig. 1b) shows a kink after $i = 3$ indicating the presence of at least three significant components [12]. The first three LSVs and RSVs (black, red, and blue in Fig. 1c, d) show clear structure. The fourth LSV and RSV (light green) are noise-like traces. In conclusion, SVD indicates that three species are present in the data matrix. There are several methods to resolve these species; most well known are soft modeling (e.g., Multivariate Curve Resolution (MCR) [13–15]) and hard modeling with the help of models that incorporate physicochemical knowledge and aim for the estimation of parameters that are physicochemically interpretable. The latter approach is termed global analysis [11, 16]. Advantages of hard over soft modeling have been demonstrated [13, 17]. A prerequisite for global analysis (*see* **Note 2**) is the availability of suitable physicochemical models.

3.2 Global Analysis of the Steady-State Fluorescence Spectra with the Help of a Spectral Model

At each GuHCl denaturant concentration, the observed emission spectrum $\psi(\lambda)$ is described as a linear combination of spectra arising from native protein (n), folding intermediate (i), and unfolded protein (u):

$$\psi(\lambda) = c_n f_n(\lambda) + c_i f_i(\lambda) + c_u f_u(\lambda) \quad (1)$$

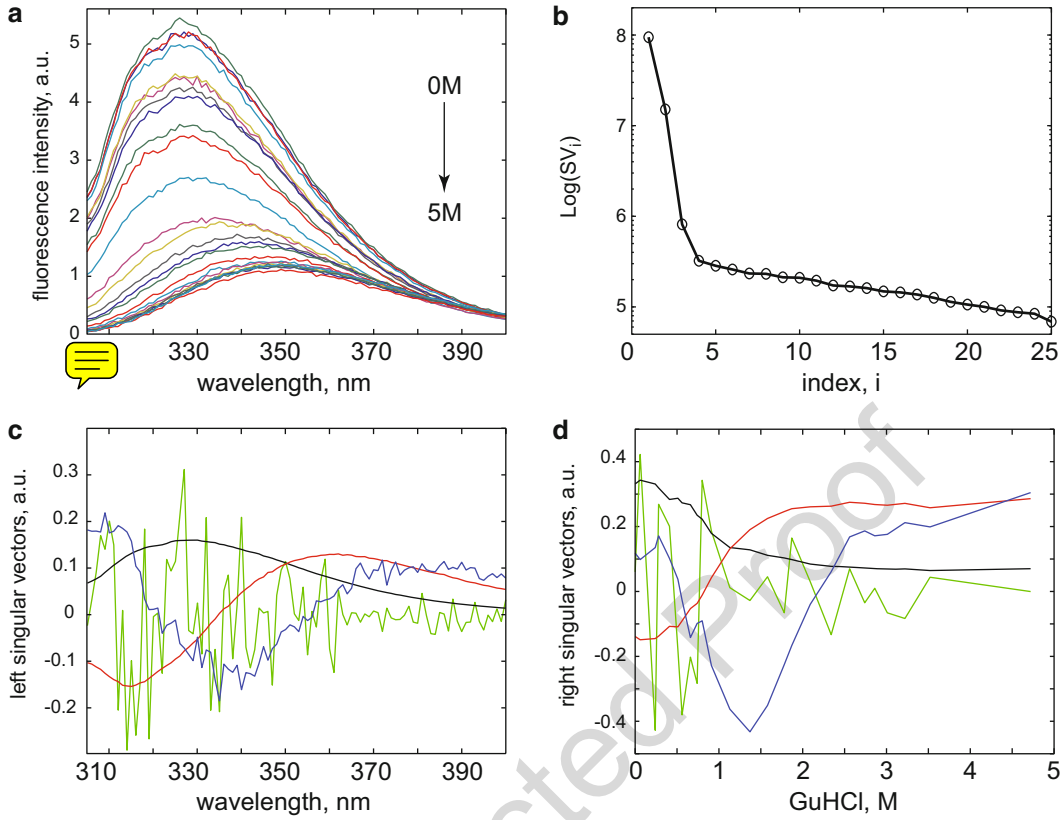

Fig. 1 Results from the singular value decomposition (SVD) analysis of the data matrix collated from steady-state fluorescence data obtained from denaturant-induced unfolding of WWF apoflavodoxin. (**a**) Denaturation trajectory of steady-state fluorescence spectra obtained at increasing concentrations of denaturant. (**b**) Scree plot of the singular values shows a kink after $i = 3$ indicating the presence of at least three significant components. (**c**) The first four left singular vectors (LSVs) (colored *black*, *red*, *blue*, and *light green*, respectively). (**d**) The accompanying first four right singular vectors (RSVs)

Steady-state fluorescence spectra obtained of protein at 0 and 4.72 M GuHCl are used as reference spectra that characterize the native and unfolded protein, respectively. The steady-state fluorescence spectrum of the folding intermediate was modeled as a skewed Gaussian in the energy domain ($\bar{\nu} = 1/\lambda$) and is described by three parameters: peak location $\bar{\nu}_{\max}$, width $\Delta\bar{\nu}$, and skewness b [16, 18]:

$$f_i(\bar{\nu}) = \bar{\nu}^5 \exp(-\ln(2)\{\ln(1 + 2b(\bar{\nu} - \bar{\nu}_{\max})/\Delta\bar{\nu})/b\}^2) \qquad (2)$$

where the parameter $\bar{\nu}_{\max}$ is the Franck-Condon wave number of maximum emission. The full width at half maximum (FWHM) is given by $\Delta\bar{\nu}_{1/2} = \Delta\bar{\nu}\sinh(b)/b$ (*see* **Note 3**).

Now all spectra are globally analyzed as a linear combination of spectra arising from native protein, folding intermediate, and unfolded protein. The three parameters that describe the shape of

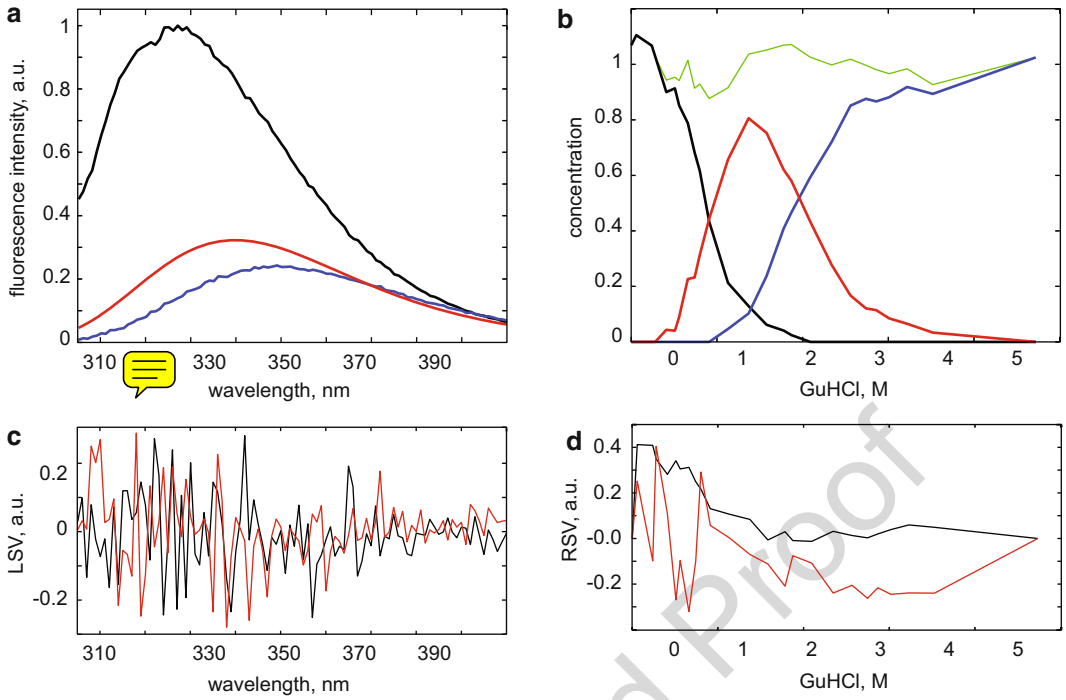

Fig. 2 Global analysis of steady-state fluorescence data matrix with the help of a spectral model. (**a**) Steady-state fluorescence spectra of WWF apoflavodoxin in 0 M GuHCl (native protein, *black*) and in 4.72 M GuHCl (unfolded protein, *blue*), respectively. The steady-state fluorescence spectrum of the folding intermediate (*red*) is modeled as a skewed Gaussian and is estimated from the global analysis of all unfolding data. (**b**) Estimated concentrations of the different folding species (colored *black*, *red*, and *blue*) as a function of denaturant concentration. The sum of the three concentrations is shown as a thin *light green line*. (**c**, **d**) Results from the singular value decomposition of the residual matrix. (**c**) The first [illegible]. (**d**) The first [illegible]

the spectrum of the folding intermediate and the concentrations of each folding species are the unknown parameters that need to be estimated from a global fit of all data. The $n_\lambda \times n_d$ matrix Ψ can be written as a matrix product.

$$\Psi = F(\bar{\nu}_{max}, \Delta\bar{\nu}, b)C^T \quad (3)$$

where the F matrix contains three columns $\left[f_n f_i(\bar{\nu}_{max}, \Delta\bar{\nu}, b) f_u\right]$ and n_λ rows and the C matrix contains three columns $\left[c_n c_i c_u\right]$ and n_d rows. The rank of the C, F, or Ψ matrix is three, consistent with the SVD. The concentration parameters were constrained to be nonnegative (*see* **Note 4**). The estimated spectrum of the folding intermediate is depicted in red in Fig. 2a. Note that is blue shifted relative to that of the unfolded protein (blue). The estimated spectral parameters are $\bar{\nu}_{max} = 29,290 \pm 40/\text{cm}$, $\Delta\bar{\nu} = 4,830 \pm 30$, and $b = -0.209 \pm 0.006$. The root mean square error of the fit was 0.41 % of the maximum of the data. The matrix of residuals resulting from the global analysis can best be diagnosed with the help of its SVD. Shortcomings of the model used show up as trends in the most

important left or right singular vectors. No such trends are present in the first (black) and second (red) LSV or RSV (Fig. 2c, d). Therefore, the fit can be accepted.

In this way the product $c_i f_i$ can be estimated, and thus, we can determine the shape of c_i as a function of the GuHCl denaturant concentration. In order to estimate the concentration c_i relative to the other concentrations, we use the constraint that the sum $c_n + c_i + c_u$ should be close to one at all GuHCl concentrations. This is estimated by means of a subsequent linear regression. All estimated concentrations are depicted in Fig. 2b, as well as this sum $c_n + c_i + c_u$ (depicted in light green). The small deviations of this sum from one are considered acceptable.

3.3 Global Analysis of the Fractions of the Different Folding Species with the Help of a Thermodynamic Model

As shown by Bollen et al. [1], denaturant-induced equilibrium unfolding of apoflavodoxin is described by a three-state model: N ⇔ I ⇔ U, in which N represents native, U represents unfolded molecules, and I is a folding intermediate. Consequently, the two corresponding equilibrium constants (i.e., K_{IN} and K_{UI}) and associated free energy differences (i.e., ΔG^0_{IN} and ΔG^0_{UN}) are

$$\begin{aligned} K_{IN} &= \frac{[N]}{[I]} = exp\left[-(\Delta G^0_{IN} + m_{IN} \times [D])/0.59\right] \\ K_{UI} &= \frac{[I]}{[U]} = exp\left[-(\Delta G^0_{UI} + m_{UI} \times [D])/0.59\right] \end{aligned} \tag{4}$$

where m_{IN} and m_{UI} describe the denaturant dependence of ΔG^0_{IN} and ΔG^0_{UI}. The number 0.59 in this equation equals the gas constant R times temperature T (=298 K) and is in kcal/mol. The fractional populations of each folding state (c_U, c_I, c_N) follow from

$$\begin{aligned} c_U &= \frac{1}{1 + K_{UI} + K_{IN} \times K_{UI}} \\ c_I &= \frac{K_{UI}}{1 + K_{UI} + K_{IN} \times K_{UI}} \\ c_N &= \frac{K_{IN} \times K_{UI}}{1 + K_{UI} + K_{IN} \times K_{UI}} \end{aligned} \tag{5}$$

Global analysis yielded the fractional populations (Fig. 2b) of the folding species at a particular denaturant concentration, and these fractions were subsequently used to estimate ΔG^0_{IN}, m_{IN}, ΔG^0_{UI}, and m_{UI}.

A least absolute values (LAV, *see* **Note 5**) approach was used during global analysis, because this approach is more robust against outliers than the least squares (LS) method (*see* **Note 6**). The LAV criterion that is minimized as a function of the thermodynamic parameters is

$$\min\left(\sum_i \left|c^i_{U^*} - c^i_U\right| + \left|c^i_{I^*} - c^i_I\right| + \left|c^i_{N^*} - c^i_N\right|\right) \tag{6}$$

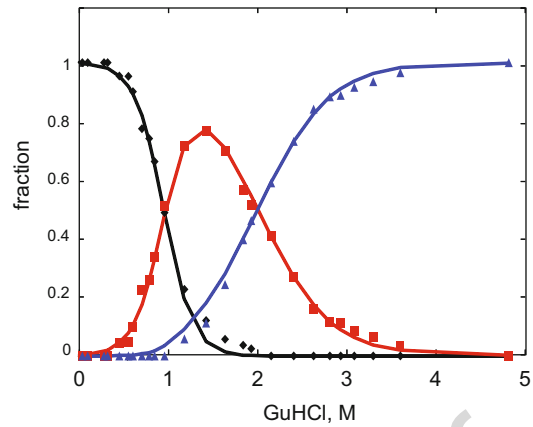

Fig. 3 Global analysis of the fractions of the different folding species with the help of a thermodynamic model. Symbols indicate relative concentrations of the three species (colored *black*, *red*, and *blue*) as a function of the denaturant concentration estimated in Fig. 2b (*diamonds*: native; *squares*: intermediate; *triangles*: unfolded). *Solid lines* depict global LAV fit

t.1 **Table 1**
Thermodynamic parameters estimated from GuHCl-induced equilibrium unfolding of WWF apoflavodoxin (*see* Note 8)

Criterion	ΔG^0_{UI} (kcal/mol)	m_{UI} (kcal/mol M^{-1})	ΔG^0_{IN} (kcal/mol)	m_{IN} (kcal/mol M^{-1})	ΔG^0_{UN} (kcal/mol)	m_{UN} (kcal/mol M^{-1})
LAV	−2.9	1.5	−3.3	3.6	−6.2	5.1
LS	−3.0	1.5	−3.1	3.4	−6.1	4.9

where c_U, c_I, and c_N are calculated with Eq. 5 and c_{U*}, c_{I*}, and c_{N*} are the *normalized* concentrations (*see* **Note** 7) from the global analysis (Fig. 2b), with i the summation index, which corresponds to the different concentrations of denaturant used. The results from the global analysis of the fractions of the different folding species, with the help of a thermodynamic model, are depicted in Fig. 3. The fit is considered satisfactory. The thermodynamic parameters estimated with LS or LAV listed in Table 1 are well interpretable [2]. The difference in the estimated values between the LAV and LS criteria is within 10 % relative precision and thus not significant. This is no surprise, since there are no large outliers here. However, in general it is advisable to use the LAV method [19, 20] when available.

4 Notes

1. The singular value decomposition (SVD) is a model-free matrix factorization technique, which decomposes the data into a sum of orthonormal vector products scaled by singular values. Here, the left singular vector (LSV) represents spectral dimension and the right singular vector (RSV) represent denaturant concentration dimension. The contribution to the data is the product of the nth left singular vector and right singular vector scaled by the nth singular value. The singular vectors are ordered based on their contribution to the data as represented by the magnitude of the singular values as shown in the scree plot. The ordinate of this scree plot is logarithmic. Ideally, the transition between data and noise appears as a kink in the scree plot. The technique can be used to explore the number of independent components in the data matrix, which is an important aspect of defining an initial model.
2. Public domain software for global analysis is available [21, 22].
3. Note that the expression for the skewed Gaussian contains a term $\ln(1 + 2b(\bar{\nu} - \bar{\nu}_{\max})/\Delta\bar{\nu})/b$ for which a limit exists when skewness parameter b approaches zero. Since $\lim_{b\to 0} \ln(1 + bx)/b = x$ the expression simplifies to $f_i(\bar{\nu}) = \bar{\nu}^5 \exp\left(-\ln(2)\{2(\bar{\nu} - \bar{\nu}_{\max})/\Delta\bar{\nu})\}^2\right)$ which is a normal Gaussian with FWHM $\Delta\bar{\nu}$. In practical computations with nonzero b, the argument of the natural logarithm has to be tested first. If it is positive, the amplitude $f_i(\bar{\nu})$ can be computed, else $f_i(\bar{\nu}) \equiv 0$.

   For the actual computation of $f_i(\bar{\nu})$ at a particular wavelength λ, one substitutes $\bar{\nu} = 1/\lambda$ in Eq. 2. The conversion from wavelength to wave number [23], $f(\bar{\nu}) = \lambda^2 f(\lambda)$, is already taken into account in Eq. 2 [11].
4. There are several ways to ensure nonnegativity of the concentration parameters. Firstly, one can use unconstrained least squares and estimate the nonlinear spectral shape parameters (peak location $\bar{\nu}_{\max}$, width $\Delta\bar{\nu}$, and skewness b) and the matrix of conditionally linear parameters C with the help of the variable projection algorithm [16, 24]. When some of the estimated concentration parameters become negative, they can be constrained to zero, which means that at that denaturant concentration a certain component does not contribute. After imposing the constraint, the data have to be refitted. This process can be automated with the help of the nonnegative least squares algorithm [25] in combination with the variable projection algorithm [26].
5. Least absolute values minimization can most easily be done using the Excel Solver function or with dedicated algorithms [27].

6. Outliers are often present with this small number (typically 15–25 denaturant concentrations) of data points. A disadvantage of LAV analysis is that it does not report standard errors. An estimate of the relative standard errors is 10 %.
7. The small deviations of the sum $c_n + c_i + c_u$ from 1 (indicated in light green in Fig. 2b) that were present are removed when dividing by this sum. The normalized concentrations are defined as $c_{N^*} = c_n/(c_n + c_i + c_u)$, $c_{I^*} = c_i/(c_n + c_i + c_u)$, and $c_{U^*} = c_u/(c_n + c_i + c_u)$ at each denaturant concentration.
8. Note that ΔG^0_{UN} and m_{UN} which describe the thermodynamic stability against unfolding have been computed using $\Delta G^0_{UN} = \Delta G^0_{UI} + \Delta G^0_{IN}$ and $m_{UN} = m_{UI} + m_{IN}$.

References

1. Bollen YJM, Sanchez IE, van Mierlo CPM (2004) Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. Biochemistry 43 (32):10475–10489. doi:10.1021/bi049545m
2. Laptenok SP, Visser NV, Engel R et al (2011) A general approach for detecting folding intermediates from steady-state and time-resolved fluorescence of single-tryptophan-containing proteins. Biochemistry 50(17):3441–3450. doi:10.1021/bi101965d
3. Visser NV, Westphal AH, van Hoek A et al (2008) Tryptophan-tryptophan energy migration as a tool to follow apoflavodoxin folding. Biophys J 95(5):2462–2469. doi:10.1529/biophysj.108.132001
4. Golub GH (1996) Matrix computations. The Johns Hopkins University Press, Baltimore
5. Henry ER (1997) The use of matrix methods in the modeling of spectroscopic data sets. Biophys J 72(2):652–673
6. Henry ER, Hofrichter J (1992) Singular value decomposition—application to analysis of experimental-data. Methods Enzymol 210: 129–192
7. Hendler RW, Shrager RI (1994) Deconvolutions based on singular-value decomposition and the pseudoinverse—a guide for beginners. J Biochem Biophys Methods 28(1):1–33. doi:10.1016/0165-022x(94)90061-2
8. Shrager RI (1986) Chemical-transitions measured by spectra and resolved using singular value decomposition. Chemom Intell Lab Syst 1(1):59–70. doi:10.1016/0169-7439(86)80026-0
9. Shrager RI, Hendler RW (1982) Titration of individual components in a mixture with resolution of difference spectra, pKs, and redox transitions. Anal Chem 54(7):1147–1152. doi:10.1021/ac00244a031
10. Satzger H, Zinth W (2003) Visualization of transient absorption dynamics—towards a qualitative view of complex reaction kinetics. Chem Phys 295(3):287–295. doi:10.1016/j.chemphys.2003.08.012
11. van Stokkum IHM, Larsen DS, van Grondelle R (2004) Global and target analysis of time-resolved spectra. Biochim Biophys Acta 1657:82–104
12. Catell RB (1966) The scree test for the number of factors. Multivar Behav Res 1:245–276
13. de Juan A, Maeder M, Martinez M, Tauler R (2000) Combining hard- and soft-modelling to solve kinetic problems. Chemom Intell Lab Syst 54(2):123–141. doi:10.1016/s0169-7439(00)00112-x
14. de Juan A, Tauler R (2003) Chemometrics applied to unravel multicomponent processes and mixtures—revisiting latest trends in multivariate resolution. Anal Chim Acta 500(1–2): 195–210. doi:10.1016/s0003-2670(03)00724-4
15. de Juan A, Tauler R (2006) Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. Crit Rev Anal Chem 36(3–4):163–176. doi:10.1080/10408340600970005
16. van Stokkum IHM, Linsdell H, Hadden JM et al (1995) Temperature-induced changes in protein structures studied by fourier transform infrared spectroscopy and global analysis. Biochemistry 34(33):10508–10518
17. van Stokkum IHM, Mullen KM, Mihaleva VV (2009) Global analysis of multiple gas chromatography-mass spectrometry (GC/MS) data sets: a method for resolution of co-eluting

components with comparison to MCR-ALS. Chemom Intell Lab Syst 95(2):150–163. doi:10.1016/j.chemolab.2008.10.004

18. van Stokkum IHM, Larsen DS, van Grondelle R (2004) Global and target analysis of time-resolved spectra. Biochim Biophys Acta Bioenerg 1657(2–3):82–104. doi:10.1016/j.bbabio.2004.04.011
19. Vandenbos A (1988) Nonlinear least-absolute-values and minimax model-fitting. Automatica 24(6):803–808. doi:10.1016/0005-1098(88)90056-8
20. Seber GAF, Wild CJ (1989) Nonlinear regression. Wiley, New York
21. Mullen KM, van Stokkum IHM (2007) TIMP: an R package for modeling multi-way spectroscopic measurements. J Stat Softw 18(3)
22. Snellenburg JJ, Laptenok SP, Seger R et al (2012) Glotaran: a java-based graphical user interface for the R-package TIMP. J Stat Softw 49:1–22
23. Lakowicz JR (2006) Principles of fluorescence spectroscopy, 3rd edn. Springer, New York
24. Golub GH, LeVeque RJ (1979) Extensions and uses of the variable projection algorithm for solving nonlinear least squares problems. In: Proceeding of the 1979 Army Numerical Analysis and Computer Conferences, ARO Report 79–3
25. Lawson CL, Hanson RJ (1974) Solving least squares problems. Prentice Hall, Englewood Cliffs
26. Mullen KM, van Stokkum IHM (2009) The variable projection algorithm in time-resolved spectroscopy, microscopy and mass spectrometry applications. Numer Algorithms 51(3):319–340. doi:10.1007/s11075-008-9235-2
27. Schlossmacher EJ (1973) Iterative technique for absolute deviations curve fitting. J Am Stat Assoc 68(344):857–859. doi:10.2307/2284512