Journal of Pharmaceutical Sciences 112 (2023) 3240–3247

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Lessons Learned

Correcting Ultraviolet-Visible Spectra for Baseline Artifacts

Andrew J. Basalla[a,b], Brent S. Kendrick[b,*]

[a] KBI Biopharma, Inc., Louisville, CO, USA
[b] First Principles Biopharma, LLC, Louisville, CO, USA

ARTICLE INFO

Article history:
Received 16 May 2023
Revised 16 August 2023
Accepted 16 August 2023
Available online 23 August 2023

Keywords:
Rayleigh light scattering
Mie light scattering
Ultraviolet (UV) spectroscopy
Correction
Curve fitting
Biopharmaceutical characterization

ABSTRACT

Rayleigh and Mie light scattering from particulates, soluble protein aggregates, or large proteins can lead to inaccuracy of concentration measurements using ultraviolet (UV) spectroscopy and Beer's Law. While a number of light scattering correction equations have been proposed in the literature, they can also lead to incorrect values if samples vary in particulate and/or soluble aggregate levels or depart in other ways from which the equations were developed. We propose a curve-fitting baseline subtraction approach based on fundamental Rayleigh and Mie scattering equations which also factors in instrument baseline artifacts. We validated this Rayleigh-Mie correction against a wide variety of positive and negative controls, including protein size standards, protein aggregates induced by forced degradation, lentivirus and polystyrene nanospheres.

Introduction

The use of ultraviolet (UV) spectroscopy in combination with the Beer-Lambert law is the de-facto standard for accurate measurements of protein concentrations due to the method's low sample requirement, wide availability, and fast results. An accurate protein concentration measurement is critical not just to establish the protein content in a sample for dosing, but feeds into several other analytical methodologies including creation of accurate potency standards for biological assays, determination of protein-ligand binding constant and stoichiometry.[1] Furthermore, determination of the protein extinction coefficient itself, either by amino acid analysis, nitrogen content, or the Edelhoch method are all critically dependent on measuring the UV absorbance without light scattering artifacts. The accuracy of UV-based concentration measurements is compromised by many factors including errors in the extinction coefficient value (often prone to ±15% error) and a variety of baseline artifacts.[1] The simplest, and perhaps most common, technique to correct for baseline artifacts is to simply subtract the absorbance value of the baseline at 350 nm (e.g. a region typically devoid of significant chromophore absorbance). This type of baseline correction is widely used to account for instrument drift, electronic noise, and other systematic errors. However, this can be a severely inaccurate approach since baseline artifacts often exponentially increase with decreasing wavelengths due to Rayleigh and Mie light scattering from the protein, protein aggregates, and both protein and non-protein particulates.[3] These light scattering artifacts inflate the absorbance measurement above the true value and often in inconsistent amounts due to the intrinsic heterogeneity of protein aggregates and particles.[4] Therefore, being able to correct an absorbance spectra measurement for the effect of light scattering in a way that can adapt to sample heterogeneity is critical for accurate protein concentration determination.

The idea of correcting an absorbance measurement for light scattering interference is not novel and, as such, a variety of methods for correcting for light scattering have been published previously. A majority of these methods include quick approximation methods similar to those published by Mach et al. and Pace et al. which both utilize a function at two discreet wavelengths outside of where the protein chromophores would absorb.[2,5] The light scattering correction method published by Mach et al.[6] is provided in Eq. (1),

$$OD(\lambda) = 10^{(m+1)* \log(OD_{320}) - m* \log(OD_{350})} \quad (1)$$

where $m = 64.32 - 25.67 * \log\lambda$, λ is a given wavelength, OD_{320} and OD_{350} is the absorbance at 320 and 350 nm, respectively. Another example is the light scattering correction method published by Porterfield et al.[7], provided in Eq. (2), and is based primarily on an assumption of pure Rayleigh scattering:

Abbreviations: UV, ultraviolet; BSA, bovine serum albumin; IgG, immunoglobulin gamma; IgA, immunoglobulin A; IgM, immunoglobulin m; AAV, adeno associated virus; VPE, variable pathlength extension.

* Corresponding author.
E-mail address: brent.kendrick@fp-biopharma.com (B.S. Kendrick).

https://doi.org/10.1016/j.xphs.2023.08.015

$$A_{corrected,l} = A_\lambda - \left(\frac{A_{340} - A_{360}}{(340\ nm)^{-4} - (360\ nm)^{-4}}\right)\lambda^{-4} - A_{340} + \left(\frac{A_{340} - A_{360}}{(340\ nm)^{-4} - (360\ nm)^{-4}}\right)340nm^{-4} \quad (2)$$

where λ is a given wavelength, $A_{corrected,l}$ and A_λ are the corrected and native absorbance values, respectively, at the given wavelength, and A_{340} and A_{360} are absorbance values at 340 and 360 nm, respectively.

Before one applies the empirical Eqs. (1) or (2) it is important to ascertain under which conditions the equations are valid. In the Mach et. al. study, the only proteins with significant light scattering were the α-A_2 and α-B_2 crystallins which had < 8% absorbance contribution due to scattering, and indeed the authors concluded that their equation was valid provided the scattering levels do not exceed 10%.[2] Similarly, the Porterfield and Zlotnick study was specifically carried out on Brome Mosaic Virus (ca. 3.7 MDa) and Hepatitis B Virus (ca. 4.6 MDa) with similar molecular masses. They noted that the Rayleigh approximation did not perfectly account for the light scattering contribution. In our experience, protein, viral and virus-like particle therapeutics and vaccines now comprise a very wide range of molecular weights, in some cases exceeding 10 MDa for certain modalities. Variable distributions of sizes, aggregate content and aggregate size distributions in samples can have a dramatic effect on the underlying exponential baseline functions. A variety of degradation and formulation conditions can create large aggregates and particulates that result scattering levels that exceed the 10% value specified in the Mach and Middaugh study. One of the aims of our investigation was to assess a variety of sample types to establish if and when Eqs. (1) and (2) may not be valid, and to provide a methodology that can be adapted to a wide range of sample types and scattering conditions.

Leach et al introduced a first-principles method derived from Rayleigh and Mie light scattering theory over 60 years ago,[8] (Eq. 3),

$$OD(\lambda) = k\lambda^{(-n)} \quad (3)$$

where λ is a given wavelength, and k and n are parameters fit from the model. This method fits the entire chromophore-free baseline to an exponential. The resulting baseline is subsequently subtracted from the original trace to generate a trace that is essentially free of light scattering interference. However, in order to perform the fit, the authors utilized a log-log linearization method, likely due to a lack of computation tools at the time. As linearization is known to increase errors, this method can now be readily replaced with a much more accurate curve-fitting approach using modern computer algorithms to minimize errors. Other than replacing the fitting technique, the only significant shortfall we found in the Leach and Scheraga equation is the lack of a parameter to correct for instrument artifacts and/or buffer mismatch. Therefore, we investigated a modification of the Leach and Scheraga method to include a wavelength-independent correction factor (i.e. variable "b" in Eq. 4). This factor accounts for artifacts such as instrument drift (i.e. "vertical displacement" as described in Mach and Middaugh).[6] As such, this method is called the Rayleigh-Mie-Offset (RMO) method.

$$OD(\lambda) = k\lambda^{(-n)} + b \quad (4)$$

where λ is a given wavelength, and k, n, and b are parameters fit from the model and n is restricted to be $1 \leq n \leq 4$.

Materials and Methods

Materials

The proteins used for analysis included bovine serum albumin (BSA), lysozyme, thyroglobulin, immunoglobulin gamma (IgG), Immunoglobulin A (IgA), and Immunoglobulin M (IgM) and were all obtained from Sigma-Aldrich. All protein samples were clarified of visible particles by filtration through a 0.2 μm syringe filter (Fisherbrand, p/n: 09-719C). The adeno associated virus (AAV) was obtained from ViroVek and the modified Lentivirus was obtained from a proprietary source. Polystyrene beads of 20, 40, 100, 200, 400, and 800 nm were all obtained from ThermoFisher.

Absorbance Data Collection

Each protein was analyzed using a full wavelength absorbance scan on three separate spectrophotometers, a Beckman Coulter DU800, an Agilent Cary90 with a CTech Solo variable pathlength extension (VPE) attachment, and a Thermofisher Nanodrop One C. Where feasible, the tungsten lamp on the spectrophotometer was left off to eliminate the discontinuity artifact observed at the crossover point between the visible and ultraviolet lamps.[6] Measurements on the DU800 used a 60 μL, 3 mm pathlength cuvette and were scanned from 200 – 800 nm. Measurements on the SoloVPE used plastic disposable sample vessels and fibrettes. The pathlength was set to match the 3 mm pathlength of the DU800 and scanned from 200 – 800 nm. Measurements on the Nanodrop used the adjustable height table and scanned from 190 – 850 nm. We utilized the chromophore-free spectral range of 320 – 600 nm for all baseline curve-fitting with the exception of thyroglobulin, where we used 350 – 600 nm due to the chromophore at ca. 320 nm.

Sample Degradation

In order to induce visible particles in the protein samples, they were degraded via a Hulamixer inverting shaker. A 1 mg/mL preparation of BSA, IgG, and thyroglobulin were set on the Hulamixer set to 100 rpm for 74 hours until visible particles were observed.

Data Processing

A custom python script with a graphical user interface was developed to fit the non-chromophore containing baseline absorbance data against Eqs. 1 - 4. Once the baseline was fit, it was subtracted from the original spectra creating the corrected spectra. The python script was optimized for accuracy, speed and precision by comparing different fitting algorithms and minimizers found in open-source python libraries. The optimized curve-fitting algorithm for Equation 4 was packaged into a simple web-based user interface (available at: https://fp-biopharma.com) using FP-Bio UV-Vis App Version: v3.0.6.

Results and Discussion

A variety of sample types were utilized to enable assessment of sample physical properties on underlying spectral baselines. The samples were chosen specifically to include a wide range of molecular weights, diameters, aggregate/particle content and size distributions and biologic drug modalities. The selected samples should theoretically span a range encompassing pure Rayleigh scattering on the low end of the size range and a mixture of Rayleigh-Mie scattering on the high end, based on the well-established approximation that Rayleigh scattering predominantly occurs with sample diameters of approximately 1/10 of the incident wavelength[9] Specifically, at wavelengths typical for protein measurements (e.g. 280 nm), molecules less than about 20 nm should be practically pure Rayleigh scatterers and display a UV-Vis baseline with high curvature (proportional to λ^{-4} based on light scattering theory), while those above 20 nm should fall into the Mie scattering regime and display a curvature between λ^{-4} and λ^{-1}.

Table 1
Theoretical Rayleigh and Mie visible light scattering regions for a given molecular weight or hydrodynamic diameter.

	Sample type	Hydrodynamic Diameter (nm)	Molecular Weight (kDa)
Rayleigh	Lysozyme	3.6	14
	BSA	7	66
	IgG	11	150
	IgA	13	320
	Thyroglobulin	17	660
	IgM	25	900
Mie	AAV	30	3700
	Lentivirus	110	7000

We selected 6 model proteins, an AAV and a Lentivirus sample which span a range of 3.6 – 110 nm hydrodynamic diameter and 14 – 7000 kDa molecular weight (Table 1). Quantifying the soluble aggregate levels and particle distributions was outside the scope of this study since it would be impractical to account for a potential continuum of all high molecular weight species, soluble aggregates, subvisible and visible particle sizes and concentrations. For example, since light scattering intensity increases with molecular weight, either a small number of large particles or a large number of small particles can dominate the overall light scattering contribution in a sample. Therefore, our selection of samples and degradation conditions was driven by the goal of having a diverse sample set containing different monomer sizes and visibly different particle levels, and ideally, a diverse range of impacts on the underlying UV-Vis baseline due to light scattering.

Impact of Monomer Size

The unstressed samples were clarified by filtration through a 0.2 μm syringe filter and were free of visible particulates. While size exclusion chromotography (SEC) data was not available for every sample, we did ascertain that the lysozyme and IgG samples were greater than 90% monomer. The lentivirus sample had > 50% high molecular weight species eluting between the SEC void volume and monomer elution position (data not shown) and therefore should be expected to scatter light as a molecule with a diameter > 110 nm or molecular weight > 7000 kDa. Our experiments with these model samples demonstrate a good correlation between theoretical and measured light scattering responses. As observed in Fig. 1, and more clearly in Fig. 2A which has been normalized in the y-axis for the chromophore-free region, the baselines of proteins that are expected to be pure Rayleigh scatterers show an almost perfect λ^{-4} response, confirming that Rayleigh scattering is dominant in the size range below 20 nm (Lysozyme, BSA and IgG are not shown for clarity due to having a similar profile as the IgA trace). However, as the protein size gets larger, Mie scattering begins to occur and the protein's baseline curvature decreases, resembling the λ^{-3} trace.

Using the RMO method and original Leach and Scheraga equation, light scattering effects of each of the protein standards were corrected (least-squares fitting parameter results provided in Table 2) and compared in Table 3 (values with baselines corrected by subtraction at 350 nm, the Mach et. al. and the Porterfield et. al. methods not shown for clarity). Comparing the uncorrected and corrected absorbance values using either correction method shows in most cases, the uncorrected values are notably different from the corrected values. Visual inspection of the corresponding traces (for example Fig. 3 A and B) shows that the RMO corrected spectra consistently had a baseline at zero absorbance in the non-chromophore regions, as well as the least amount of light scattering induced offset at 280 nm relative to the Leach and Scheraga corrected spectra. This effect is observed quantitatively for virtually every sample in Table 3 where the RMO corrected spectra had the largest percent drop in absorbance at 280 nm relative to the Leach and Scheraga method (likely due to the Leach and Scheraga method not accounting for vertical baseline offsets). When translating to protein concentration, using uncorrected absorbance values would result in incorrect concentration values that are up to 30% off from the true value. Relating this to manufacturing of dosing of patients, an error of this magnitude could

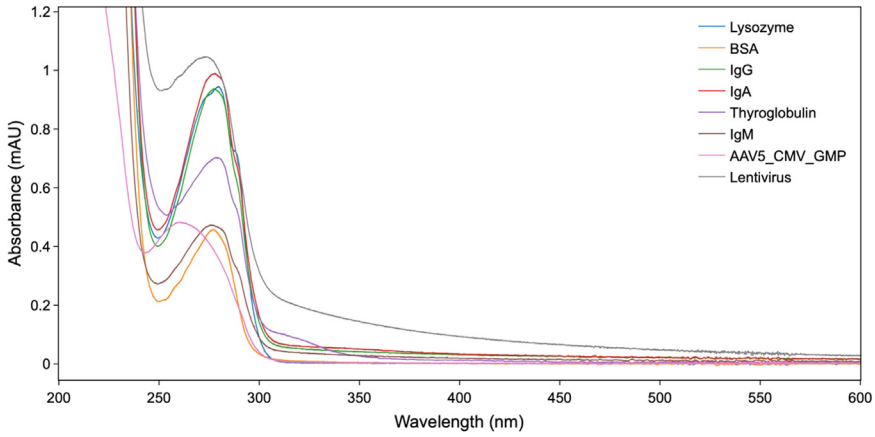

Figure 1. Full wavelength absorbance spectra of samples.

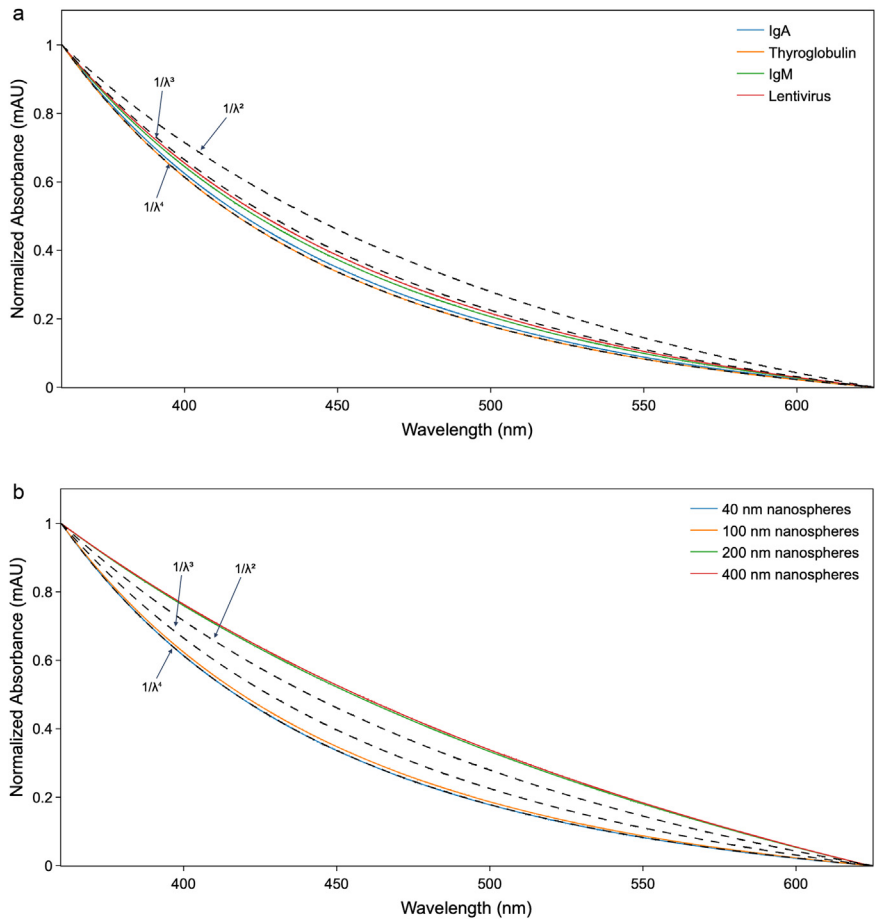

Figure 2. Wavelength dependency of the exponential baselines of a: proteins and b: polystyrene nanospheres. (Lysozyme, BSA and IgG not shown for clarity, and have the same profile as IgA).

Table 2
Parameter results from least-squares regressions.

Sample	RMO			Leach and Scheraga	
	k	n	b	k	n
Lysozyme	2.08×10^{7}	4.00	1.05×10^{-3}	2.46	1.20
BSA	6.04×10^{7}	4.00	-8.00×10^{-4}	3.26×10^{-17}	7.82
IgG	2.69×10^{5}	2.72	9.45×10^{-3}	625	1.65
IgA	1.56×10^{8}	3.78	1.19×10^{-2}	5.31×10^{3}	1.99
Thyroglobulin	2.72×10^{8}	4.00	1.58×10^{-3}	3.83×10^{6}	3.26
IgM	9.70×10^{6}	3.39	4.57×10^{-3}	1.53×10^{4}	2.26
AAV	2.27×10^{6}	4.00	2.04×10^{-3}	1.39×10^{-3}	-6.57×10^{-2}
Lentivirus	1.57×10^{7}	3.29	1.67×10^{-3}	5.26×10^{6}	3.10
Degraded BSA	5.24	1.00	2.16×10^{-3}	18.3	1.24
Degraded IgG	2.72×10^{8}	4.00	8.22×10^{-2}	0.428	0.256
Degraded Thyroglobulin	3.89×10^{6}	3.19	6.96×10^{-3}	4.47×10^{3}	2.01

be detrimental to patient safety and lead to greatly underdosing the patient or greatly under delivering in a production batch.

Impact of Protein Aggregate Impurities

Protein aggregates and particles are typical impurities in many protein samples. Therefore, a set of protein samples were agitated to induce aggregation and particle formation. The absorbance profile of each sample was corrected with the method of Leach and Scheraga and the RMO method. An aliquot of the degraded samples was filtered through a 0.2 μm syringe filter to provide a spectral profile free of large aggregates and enable assessment of the effectiveness of the Leach and Scheraga and RMO baseline correction methods. (Fig. 3). The visible particles in the degraded samples demonstrate varying degrees of baseline offset and curvature, which substantiates the necessity of using an adaptive exponential light scattering correction function. When fitting the degraded samples to either the Leach and Scheraga or RMO equation, the resulting exponential term is between 1 and 4, demonstrating that pure Rayleigh light scattering correction is not sufficient and must accommodate Mie scattering for accurate baseline correction in the presence of aggregated protein.

The IgG sample provides a good illustration of the need to incorporate an offset parameter to the baseline fitting algorithm (parameter b in Equation 4). This sample had a large positive baseline offset

Table 3
Absorbance at 280nm values for unstressed samples. Baselines uncorrected, corrected by the Leach and Scheraga method and RMO method. R^2 quality of baseline fit shown for the curve-fitting methods.

Protein	Absorbance at 280 nm			Percent difference from uncorrected		R^2 from fit	
	Uncorrected	Leach and Scheraga corrected	RMO corrected	Leach and Scheraga	RMO	Leach and Scheraga	RMO
Lysozyme	0.944	0.941	0.940	0.30%	0.47%	0.069	0.074
BSA	0.439	0.438	0.430	0.12%	2.06%	0.226	0.374
IgG	0.925	0.868	0.858	6.14%	7.30%	0.971	0.977
IgA	0.978	0.907	0.879	7.25%	10.15%	0.967	0.981
Thyroglobulin	0.701	0.661	0.655	5.64%	6.54%	0.910	0.922
IgM	0.464	0.420	0.409	9.50%	11.78%	0.942	0.948
AAV	0.355	0.353	0.353	0.57%	0.68%	0.0003	0.001
Lentivirus	0.995	0.718	0.698	27.76%	29.81%	0.997	0.997

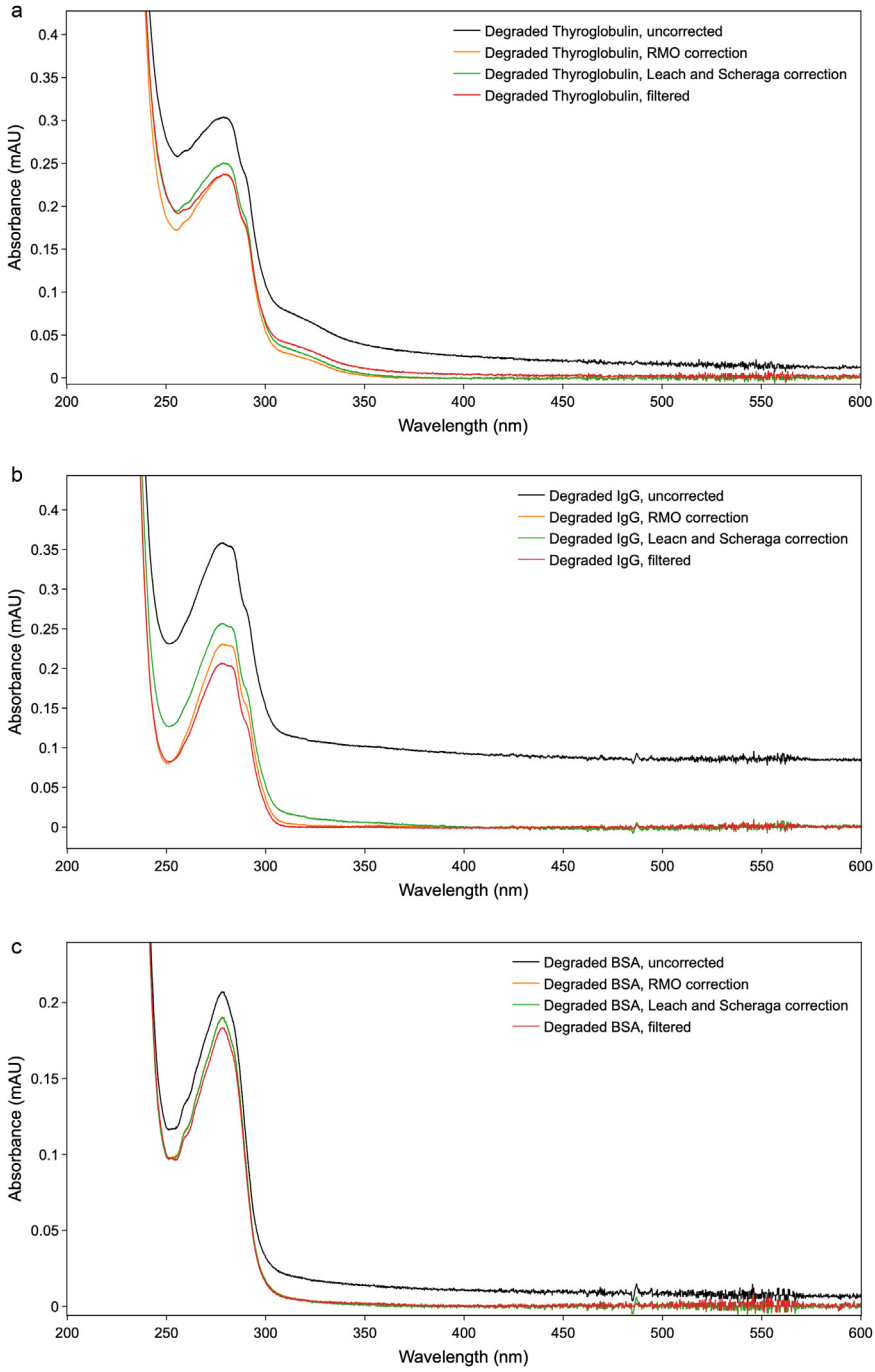

Figure 3. Absorbance spectra of degraded, filtered, and light scattering corrected data using the RMO and Leach and Scheraga methods for samples of A) Thyroglobulin; B) IgG; C) BSA.

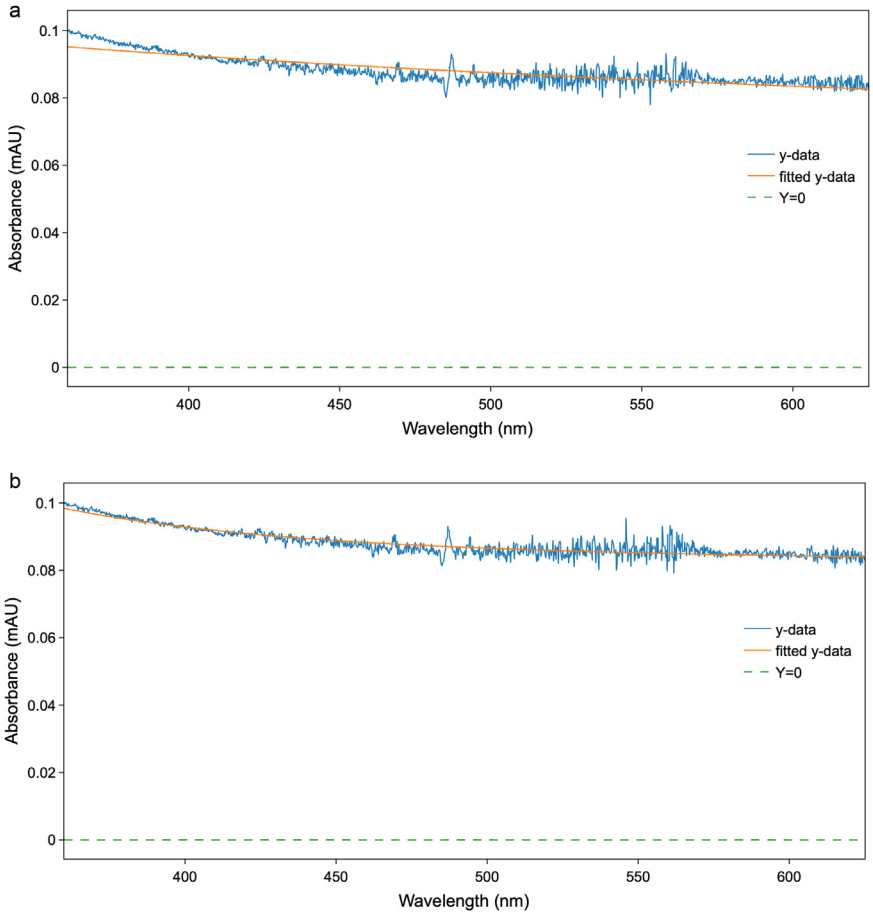

Figure 4. Difference of fits of the model on the degraded IgG baselines, (a) Leach and Scheraga (R^2 = 0.72), (b) RMO (R^2 = 0.84).

due to significant turbidity. This resulted in a poor fit (R^2 = 0.72) when fitting the spectrum to the Leach and Scheraga function (Fig. 4a) since, without a wavelength independent parameter, the least-squares best fit underestimates the exponential parameter (as observed by the low degree of curvature in the fitted line). The RMO method has an improved fit, with an R^2 = 0.84, Fig. 4b, visually accommodating both the wavelength independent offset and the curvature. The use of an extra parameter in curve fitting has been previously shown to account for any instrument or measurement artifacts and, in this case, accounts for any large turbidity offsets.[10]

Quantitative results of using the Leach and Scheraga and RMO methods for the degraded samples are presented in Table 4. Like the results for the unstressed samples, the RMO method appeared to result in a more accurate light scattering artifact correction compared to the original method and had equal or better R^2 values for the fit quality.

Impact of Subvisible and Visible Particles

Polystyrene nanosphere beads ranging from 20 nm to 800 nm were analyzed to observe the effect of particles' presence on Rayleigh and Mie scattering. Based on the profile of the exponential curve in the baseline spectra of these large beads in Fig. 2b, it is observed that as the particle size approaches the wavelength of light, the curves become less exponential. This is indicative of the expected increasing effect of the Mie scattering (which is proportional to particle diameter[11]) and decreasing effect of the Rayleigh scattering. Specifically, the 200 nm, 400 nm, and 800 nm particles' baselines have a smaller exponential factor than the λ^{-2} trace (the 800 nm trace is not shown for clarity due to having a similar profile as the 400 nm trace.) For the 20 nm, 40 nm, and 100 nm particles, there is still a dominant Rayleigh scattering effect even at these larger sizes (20 nm trace not shown due to overlaying directly on the 40 nm trace.)[11] Mie scattering is not dominant until particle size is at or above 200 nm. Since proteins,

Table 4
Absorbance at 280nm values for degraded proteins. Baselines uncorrected, corrected by the Leach and Scheraga method and RMO method. R^2 quality of baseline fit shown for the curve-fitting methods.

Protein	Absorbance at 280 nm			Percent difference from uncorrected		R^2 from fit	
	Uncorrected	Leach and Scheraga corrected	RMO corrected	Leach and Scheraga	RMO	Leach and Scheraga	RMO
Degraded BSA	0.203	0.186	0.187	8.37%	8.14%	0.639	0.640
Degraded IgG	0.356	0.254	0.229	28.52%	35.54%	0.722	0.838
Degraded Thyroglobulin	0.303	0.250	0.237	17.57%	21.78%	0.947	0.953

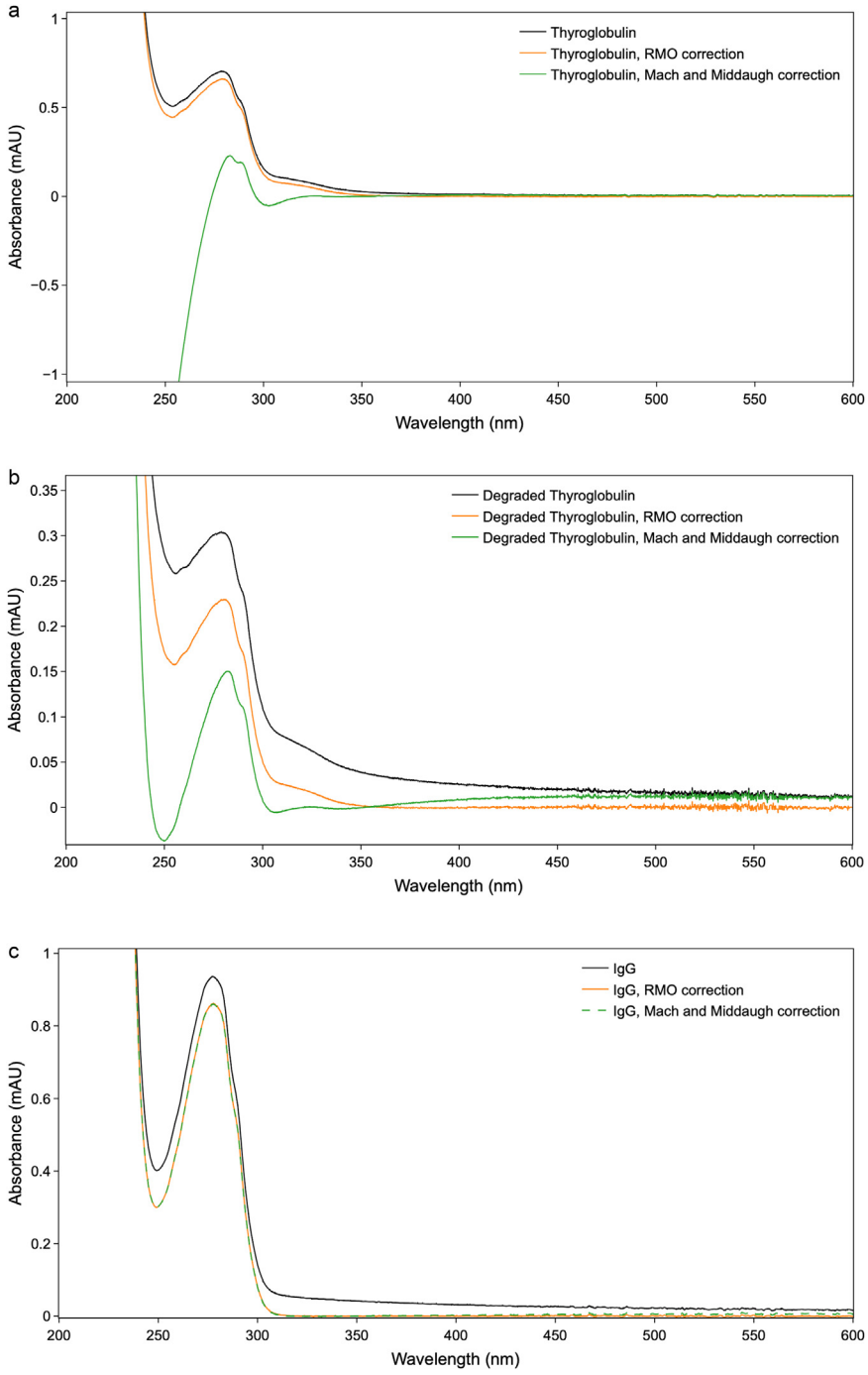

Figure 5. Uncorrected and corrected wavelength scans using the Mach and Middaugh method and RMO method for (a) thyroglobulin; (b) degraded thyroglobulin; (c) IgG.

protein aggregates and biological modalities can have a wide array of particle size distributions well above 200 nm in diameter, a general light scattering correction model should allow an exponential term to vary between 1 and 4 to account for varying levels of Mie scattering. Furthermore, the data also suggests that a Rayleigh light scattering correction using a pure l^4 correction or a simple A340 subtraction correction would not be accurate in biological samples with large amounts of aggregates/particles or an intrinsic monomer size greater than 200 nm.

Comparison of Spectrophotometers

We utilized the DU800 spectrophotometer for all data collection above. In addition, we selected a subset of samples (i.e. all samples shown in Fig. 1) to assess spectral profiles collected on the SoloVPE and Nanodrop 1C spectrophotometers. The DU800 was superior to the other spectrophotometers in consistency, noise, and accuracy (data not shown). The use of the same cuvette for both blanking and sample reading, leaving the visible lamp turned off during analysis

and truncating the spectra to not include the noise in the higher wavelength region[6] led to the most precise readings with lowest noise in the DU800. While the Nanodrop has the smallest volume requirement of only 2 μL (and accordingly a very small pathlength), it is better suited for quick absorbance measurements than full wavelength scans due to a higher level of noise present in the scans. The SoloVPE's main function is changing pathlengths to generate a slope and calculate concentration based on the slope. We experienced slight inconsistencies with SoloVPE spectral intensities when performing absorbance scans, likely due to manufacturing differences of the disposable cuvettes. Since light scattering is directly proportional to pathlength (much like Beers Law), any reduction in pathlength to minimize light scattering will proportionally decrease the chromophore signal, obviating any benefits with a smaller pathlength.

When is a Curve-Fitting Approach Like the RMO Method Necessary?

Utilizing a simple subtraction of the baseline value at 350 nm introduces accuracy error due to the significant baseline curvature observed in all samples (Fig. 2A). Recognition of this issue over the last few decades has led to the development of empirical equations and first-principles approaches to improve accuracy. Whether this error is significant for a given biologic should be determined for each sample type. Utilizing empirical equations such as those provided by Mach and Middaugh, Porterfield et al. and others offers simplicity and speed over curve-fitting approaches (e.g. Leach and Scheraga and the RMO method). However, there are pitfalls that exist where blind application of empirical equations can lead to a highly inaccurate A280 results. For example, the thyroglobulin protein has a chromophore with an absorbance at ca. 320 nm. Utilizing the Mach and Middaugh equation (Eq. 1), which was derived on samples with an exponential baseline rise, is not valid when an additional spectral peak is present in the baseline region between 320 − 350 nm. When we plot the thyroglobulin spectra that has been corrected using the Mach and Middaugh equation, indeed we observed the resulting spectra to quickly fall into negative absorbance at wavelengths lower than 300 nm (Fig. 5a and Fig. 5b). Utilizing the Porterfield equation results in a similar inaccuracy (data not shown). This demonstrates that when significant absorbance due to chromophores and/or aggregates/particles exist in the region used for scattering corrections, utilizing empirical equations can lead to highly inaccurate results. Therefore, before using a quick approximation method such as the Mach and Middaugh or Porterfield equations, we recommend first visually inspecting the baselines of representative samples. If there is an unusually large baseline offset (i.e. greater than approximately 0.1 AU) or a noticeably different curvature / shape of the sample's baseline relative to a standard protein such as BSA, then it may be prudent to validate one of the above empirical equations against the RMO method for each sample type to ensure the accuracy of the corrected value. For example, as demonstrated in Fig. 5c, the Mach and Middaugh correction method matches the RMO method for the particular IgG sample type tested. If the manufacturer of that IgG molecule desired to utilize the simpler Mach and Middaugh method for routine testing, they could first similarly verify the method performance for a variety of relevant IgG sample types against the RMO method.

Conclusion

Leach and Scheraga published a curve-fitting approach for light scattering correction over 50 years ago. Their method, modified to include an additional baseline offset parameter (e.g. the RMO method presented here) provides an accurate way of correcting for light scattering artifacts based on first principles of Rayleigh and Mie scattering effects. Modern curve-fitting software offers an accessible way to apply the RMO method. Scientists developing absorbance-based concentration methods (or utilizing absorbance spectra for determining extinction coefficients by the method of Edelhoch, AAA, etc.) may want to consider the following workflow. First, visually evaluate the baselines of their UV spectra with a set of negative and positive controls that represent likely sample types for which their method may be applied (e.g. employ samples that are unstressed vs. stressed, filtered vs. unfiltered, upstream vs. downstream purified, etc.). If significant baseline offsets, curvature or chromophore peaks exist, employ the RMO method to baseline correct the spectra. Compare the RMO method to the appropriate empirical equation-based method (e.g. a simple A350 subtraction, the Mach and Middaugh equation, etc.). Implement the equation-based method if it appears to be valid for the range of expected sample types. If the equation-based method is not valid for the sample types, then implement a curve-fitting algorithm such as the RMO method. Following this approach should ensure UV-absorbance based methods are developed with appropriate accuracy, speed and precision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Maity H, Wei A, Chen E, Haidar JN, Srivastava A, Goldstein J. Comparison of predicted extinction coefficients of monoclonal antibodies with experimental values as measured by the Edelhoch method. *Int J Biol Macromol*. 2015;77:260–265. https://doi.org/10.1016/j.ijbiomac.2015.03.027.
2. Mach H, Middaugh CR, Lewis RV. Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. *Anal Biochem*. 1992;200(1):74–80. https://doi.org/10.1016/0003-2697(92)90279-G.
3. Rayleigh Lord. XXXIV. On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky. *Lond Edinb Dublin Philos Mag J Sci*. 1899;47(287):375–384. https://doi.org/10.1080/14786449908621276.
4. Pignataro MF, Herrera MG, Dodero VI. Evaluation of peptide/protein self-assembly and aggregation by spectroscopic methods. *Molecules*. 2020;25(20):4854. https://doi.org/10.3390/molecules25204854.
5. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci Publ Protein Soc*. 1995;4(11):2411–2423.
6. Mach H, Middaugh CR. Ultraviolet spectroscopy as a tool in therapeutic protein development. *J Pharm Sci*. 2011;100(4):1214–1227. https://doi.org/10.1002/jps.22385.
7. Porterfield JZ, Zlotnick A. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology*. 2010;407(2):281–288. https://doi.org/10.1016/j.virol.2010.08.015.
8. Leach SJ, Scheraga HA. Effect of light scattering on ultraviolet difference spectra [1]. *J Am Chem Soc*. 1960;82(18):4790–4792. https://doi.org/10.1021/ja01503a008.
9. Lockwood D. Rayleigh and mie scattering. In: 2016:1097-1107. https://doi.org/10.1007/978-1-4419-8071-7_218.
10. Cascone C, Murphy KR, Markensten H, et al. AbspectroscoPY, a Python toolbox for absorbance-based sensor data in water quality monitoring. *Environ Sci Water Res Technol*. 2022;8(4):836–848. https://doi.org/10.1039/D1EW00416F.
11. Acharya R. Interaction of waves with medium. *Satellite Signal Propagation, Impairments and Mitigation*. Elsevier; 2017:57–86. https://doi.org/10.1016/B978-0-12-809732-8.00003-X.