


Performance Comparison of Spectral Distance Calculation Methods

Taiji Oyama¹ , Satoko Suzuki¹, Yasuo Horiguchi¹, Ai Yamane¹, Kenichi Akao¹, Koushi Nagamori¹, and Kouhei Tsumoto^{2,3}

Applied Spectroscopy
2022, Vol. 76(12) 1482–1493
© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00037028221121687

journals.sagepub.com/home/asp



Abstract

Circular dichroism (CD) spectroscopy is a widely used technique for assessing the higher-order structure (HOS) of biopharmaceuticals, including antibody drugs. Since the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use established quality control guidelines, objective evaluation of spectral similarity has been required in order to assess structural comparability. Several spectral distance quantification methods and weighting functions to increase sensitivity have been proposed, but not many reports have compared their performance for CD spectra. We constructed comparison sets that combine actual spectra and simulated noise and performed a comprehensive performance evaluation of each spectral distance calculation method and weighting function under conditions that consider spectral noise and fluctuations from pipetting errors. The results showed that using the Euclidean distance or Manhattan distance with Savitzky–Golay noise reduction is effective for spectral distance assessment. For the weighting function, it is preferable to combine the spectral intensity weighting function and the noise weighting function. In addition, the introduction of the external stimulus weighting function should be considered to improve the sensitivity. It is crucial to select the weighting function based on the balance between spectral changes and noise distributions for robust, sensitive antibody HOS similarity assessment.

Keywords

Circular dichroism, CD, spectroscopy, protein structures, biopharmaceutical characterization, antibody drugs, biosimilar

Date received: 2 May 2022; accepted: 2 August 2022

Introduction

Circular dichroism (CD) spectroscopy is a long-established technique for obtaining information on the higher-order structure (HOS) of biomolecules such as proteins and nucleic acids. Although CD spectroscopy does not have as high a resolution as X-ray crystallography or nuclear magnetic resonance (NMR), it is widely used for structural analysis of biomolecules because it is easy to measure samples in solution. Its applications range from secondary structure analysis, evaluation of thermodynamic properties, interaction analysis, and HOS similarity comparison.^{1–5}

Various methods have been developed for estimating the fraction of secondary structure using CD spectra.^{6–7} In recent years, algorithms have been developed that can accurately estimate the secondary structure in β -sheet-rich proteins, which was difficult in the past, so increasing the practicality of CD spectroscopy for β -sheet-rich antibody drugs.^{8–10} While structure estimation method using CD spectra has been established, the assessment of structural similarities (differences) using CD spectra has generally been performed visually and subjectively.

In the past decade, the application of CD spectroscopy to the analysis of antibody drugs has increased with the acceleration of the development, manufacture, and launch of antibody drugs and biosimilars.^{11–14} The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) has established quality guidelines, and ICH-Q6B defines CD spectroscopy as one of the methods for determining the HOS of a biopharmaceutical drug substance, drug product, or intermediate. In addition, ICH-Q5E provides guidelines for the objective evaluation of changes in the HOS of biopharmaceuticals due to changes in the manufacturing process. Based on this background, there is

¹JASCO Corporation, Tokyo, Japan

²The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

³Department of Bioengineering, School of Engineering, The University of Tokyo, Tokyo, Japan

Corresponding author:

Taiji Oyama, JASCO Corporation, 2967 5 Ishikawamachi, Hachioji, Tokyo 192-8537, Japan.

Email: taiji.oyama@jasco.co.jp

a need for an objective and accurate method for evaluating CD spectral similarities (differences) in the manufacturing process for biopharmaceuticals.

Some methods have been developed to quantify the differences in spectra (Table S1, Supplemental Material). Dinh et al. developed a method that uses the spectral intensity as a weighting function when calculating the Euclidean distance (weighted spectral distance or WSD).¹⁵ The method of weighting using noise spectra and the process for applying weighting functions not only to Euclidean distance but also to other distance calculation methods were developed.¹⁶

Performance comparisons have been conducted for some spectral distance calculation methods. Teska et al. compared the performance of the Euclidean distance, the correlation coefficient, the area of overlap (AOO), and the derivative correlation algorithm (DCA) for CD and Fourier transform infrared (FT-IR) spectra, which are numerical mixtures of natural and denatured IgG spectra, and provided guidelines for selecting the appropriate method.¹⁷ Kendrik et al. compared the performance of the WSD, AOO, and DCA methods by using CD, IR, and microfluidics modulation spectra of IgG spiked with impurities, and the limit-of-quantification (LOQ) method presented in ICH-Q2 (R1); they reported the effectiveness of the WSD for microfluidics modulation spectra.¹⁸ Jones reported that imperfect spectra, including noise, baseline offset, wavelength shift, and incorrect spectral intensity, affect distance calculations such as correlation coefficient, DCA, AOO, and WSD.¹⁹

In addition to the above methods, other methods for quantifying spectral differences have been reported, including Manhattan distance, modified area of overlap (MAO), and derivative correlation coefficient.^{16,20,21} For Euclidean and Manhattan distances, spectral normalization methods have been proposed preprocessing.²²

Each method can also be combined with a weighting function, such as the spectral intensity. Comprehensive knowledge of the performance of these spectral distance calculation methods in combination with weighting functions will improve the sensitivity and robustness of comparability assessments of the HOS of biopharmaceuticals using CD spectra and will guide method selection. We constructed two comparison sets, one using Herceptin, a widely used antibody drug, and the other using variable domain of heavy chain of heavy chain antibody (VHH), which is expected to be a next-generation antibody drug. The performance of each combination of spectral distance calculation methods and weighting functions was comprehensively evaluated in consideration of spectral noise and sample preparation errors.

Materials and Methods

Sample Preparation and Circular Dichroism Spectrum Measurement

Herceptin was purchased from Roche, human IgG was purchased from Sigma-Aldrich, and two different sequences of

VHH1 and VHH2 were provided by Dr. Akikazu Murakami (RePHAGEN Corporation). Herceptin and human IgG were dissolved in Milli-Q water to final concentrations of 0.80 mg/mL, 0.81 mg/mL (near-ultraviolet, or near-UV), 0.16 mg/mL, and 0.11 mg/mL (far-UV), respectively. VHH1 and VHH2 were dissolved in a 20 mM PBS solution. The CD spectra of each sample in the near- and far-UV regions were measured using a J-1500 CD Spectrometer (JASCO Corporation) under the conditions shown in Table S2 (Supplemental Material).

Calculations of Spectral Distance

The various spectral distance calculation methods are described in Eqs. 1 through 7, where R_i is the spectrum of the reference and U_i is the spectrum of the comparison sample. U_i and R_i spectra are normalized by the L2 norm ($\sqrt{\sum U_i^2}$ and $\sqrt{\sum R_i^2}$) in Eq. 2 or normalized by the L1 norm ($\sum |U_i|$ and $\sum |R_i|$) in Eq. 4. In this normalization process, the intensity change of the whole spectrum is canceled out. Similarly, the correlation coefficient, AOO, MAO, derivative correlation coefficient, and DCA do not consider changes in the intensity of the whole spectrum. The AOO method calculates the area where two normalized absolute spectra overlap.²³ Regions where the signs of the original spectra are reversed are not included in the calculation. The MAO method was developed to improve the dynamic range of the AOO method, and uses the square instead of the absolute value of the spectrum.²¹ The derivative correlation coefficient is calculated using the first derivative spectra of the reference and comparison samples, R'_i and U'_i .¹⁷ DCA is calculated by applying the calculation process shown in Eq. 7 to the derivative correlation coefficient.

Euclidean Distance

$$E = \sqrt{\frac{1}{n} \sum_{i=1}^n (U_i - R_i)^2} \quad (1)$$

Normalized Euclidean Distance

$$NE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{U_i}{\sqrt{\sum U_i^2}} - \frac{R_i}{\sqrt{\sum R_i^2}} \right)^2} \quad (2)$$

Manhattan Distance

$$M = \frac{1}{n} \sum_{i=1}^n |U_i - R_i| \quad (3)$$

Normalized Manhattan Distance

$$NM = \frac{1}{n} \sum_{i=1}^n \left| \frac{U_i}{\sum |U_i|} - \frac{R_i}{\sum |R_i|} \right| \quad (4)$$

Correlation Coefficient

$$R = \frac{\sum_{i=1}^n (U_i - \bar{U})(R_i - \bar{R})}{\sqrt{\sum_{i=1}^n (U_i - \bar{U})^2} \sqrt{\sum_{i=1}^n (R_i - \bar{R})^2}} \quad (5)$$

Derivative Correlation Coefficient

$$p = \sqrt{\frac{\left(\sum_{i=1}^n U_i R_i \right)^2}{\sum_{i=1}^n U_i^2 \sum_{i=1}^n R_i^2}} \quad (6)$$

Derivative Correlation Algorithm

$$DCA = \frac{(p^{21} + p)}{2} \quad (7)$$

Weighting Functions

Examples of weighted Euclidean distance and weighted correlation coefficients are shown in Eqs. 8 and 9, where ω_i is the weighting function. For the Euclidean distance, the weighting function is introduced as the coefficient of the square of the residuals of the spectra of the reference and comparison samples. For the correlation coefficients, the weighting functions are introduced as coefficients of the product of the deviation of the reference and the deviation of the comparison sample, the product of the deviation of the reference, and the deviation of the comparison sample. For AOO, the weighting functions are introduced as coefficients of the spectra of the reference and comparison samples, respectively, before area normalization.

The spectral intensity weighting function ($\omega_{spec,i}$) is obtained by taking the absolute value of the spectrum of the reference and normalizing it by the mean value, as shown in Eq. 10. The noise weighting function ($\omega_{spec,i}$), shown in Eq. 11, is obtained by taking the standard deviation of the noise ($\sigma_{sample,i}$) using the high-tension voltage (HT) spectrum obtained with the CD spectrum of the comparison sample and Eq. 12. V_i is the HT value of each wavelength and k is the number of scan.¹⁶ It is also useful to apply an external stimulus to the reference and investigate the wavelength range where the spectral change could occur. Use this as an external stimulus weighting function ($\omega_{ext,i}$). $\omega_{ext,i}$ is obtained by normalizing the absolute value of the subtraction spectrum (ϕ_i) between the reference and the spectrum of the reference

under different conditions of temperature, pH, concentration, additive concentration, and impurity concentration by the average value, as shown in Eq. 13. In this study, we employed an evaluation system in which each impurity (IgG and VHH2) is numerically spiked into the reference spectra of Herceptin and VHH1, respectively. We considered the spectral changes due to numerical impurity spikes as spectral changes due to an external stimulus. External stimulus weighting functions were obtained based on the difference spectra of 100% Herceptin and 100% IgG, and the difference spectra of 100% VHH1 and 100% VHH2. The external stimulus weighting functions for the distance calculation method with spectral normalization were obtained from the difference spectra of 100% Herceptin and 100% IgG and 100% VHH1 and 100% VHH2 after normalization. The L1 norm ($\sum |U_i|$ and $\sum |R_i|$) was used for the normalized Manhattan distance, correlation coefficient, AOO, and MAO, and the L2 norm ($\sqrt{\sum U_i^2}$ and $\sqrt{\sum R_i^2}$) was used for the normalized Euclidean distance. In other words, the external stimulus weighting function that we applied in this study weights the regions where we know that the spectra clearly change.

$$I = \sqrt{\frac{1}{n} \sum_{i=\lambda_1}^{\lambda_n} \{(U_i - R_i)^2 \omega_i\}} \quad (8)$$

$$I = \frac{\sum_{i=1}^n \{(U_i - \bar{U})(R_i - \bar{R}) \omega_i\}}{\sqrt{\sum_{i=1}^n \{(U_i - \bar{U})^2 \omega_i\}} \sqrt{\sum_{i=1}^n \{(R_i - \bar{R})^2 \omega_i\}}} \quad (9)$$

$$\omega_{spec,i} = \frac{|R_i|}{(1/n) \sum_{i=\lambda_1}^{\lambda_n} |R_i|} \quad (10)$$

$$\omega_{noise,i} = \frac{(1/n) \sum_{i=\lambda_1}^{\lambda_n} |\sigma_{sample,i}|}{|\sigma_{sample,i}|} \quad (11)$$

$$\log_{10} \left(\sqrt{k} \sigma_{sample,i} \right) = 1.3313 \times (\log_{10}(V_i))^2 - 3.5163 \times \log_{10}(V_i) - 0.137 \quad (12)$$

$$\omega_{ext,i} = \frac{|\phi_i|}{(1/n) \sum_{i=\lambda_1}^{\lambda_n} |\phi_i|} \quad (13)$$

Creating a Spiked MRE Spectrum Set with Random Noise

The process of generating spectra with wavelength-dependent noise is shown in Figure S1 (Supplemental Material), using the

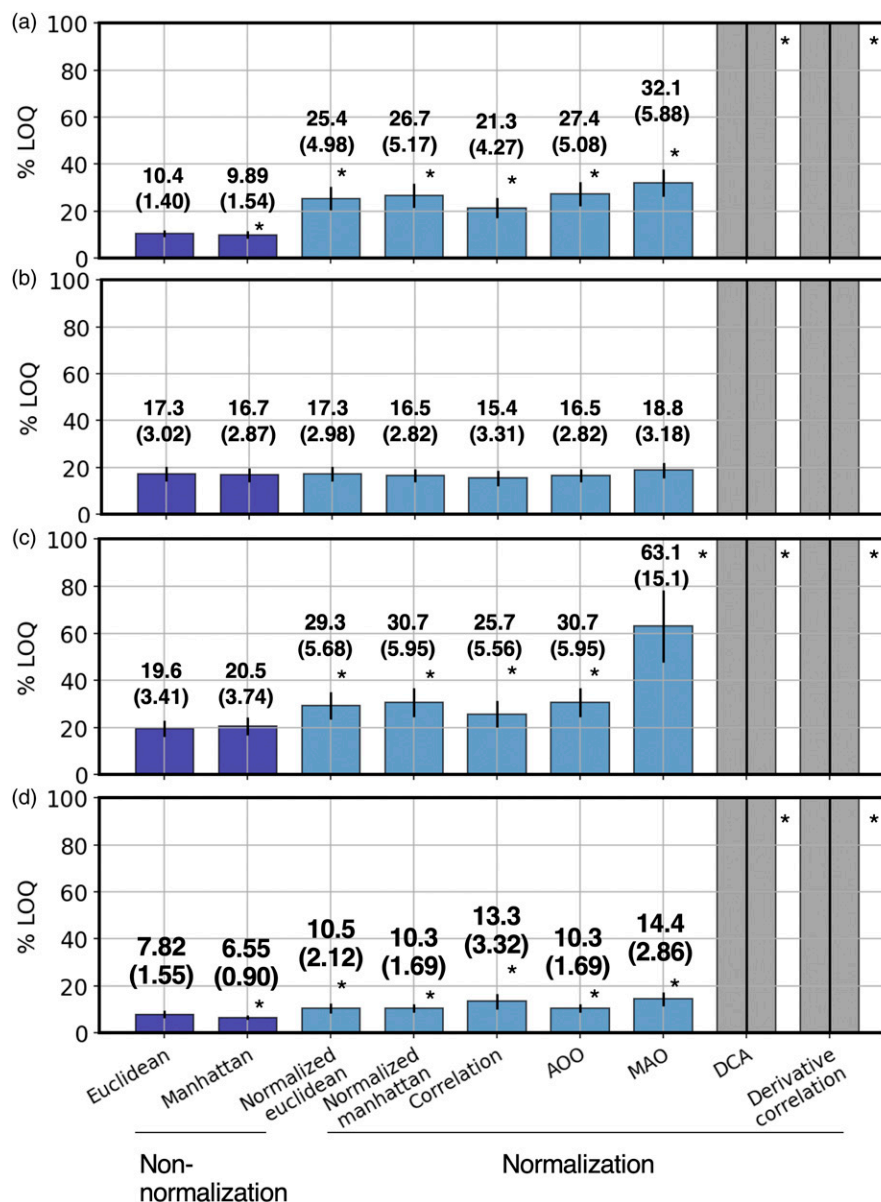


Figure 1. Limit-of-quantification comparison of various distance calculation methods. The LOQ values were calculated using the following comparison sets generated under the condition $k = 9$: (a) Herceptin/IgG near-UV, (b) Herceptin/IgG far-UV, (c) VHH1/VHH2 near-UV, and (d) VHH1/VHH2 far-UV. Each bar graph and error bar show the mean and standard deviation of the LOQ calculated after 100 independent trials of adding noise. The upper number in the bar graph indicates the mean value (standard deviation). Two-sample t -tests were performed for the LOQ for each distance calculation method and the Euclidean distance, and results showing $p < 0.05$ are marked with *.

comparison set of Herceptin and IgG as examples. A low-pass filter was applied to each original spectrum using the FFT filter function of Spectra Manager v.2.0 (JASCO Corporation) under the conditions shown in Table S3 (Supplemental Material). Representative examples of spectra before and after the application of the low-pass filter are shown in Figure S2 (Supplemental Material). The buffer spectrum was subtracted and converted to the mean residue ellipticity (MRE), which was used as the original spectrum. The spectra of IgG were numerically spiked with the spectra of Herceptin at 2%

intervals up to 40% (near-UV) and 1.5% intervals up to 30% (far-UV) to obtain a set of 20 spiked MRE spectra each. Similarly, VHH2 was spiked with VHH1 at 2% intervals up to 40% (near-UV) and at 1% intervals up to 20% (far-UV) to obtain a set of spiked MRE spectra.

The HT spectra corresponding to these sets of MRE spectra were also obtained by numerical spiking. Since HT spectra are not linear with respect to concentration, the HT spectra were converted to absorbance spectra using the HTOD conversion function of Spectrum Manager Ver.

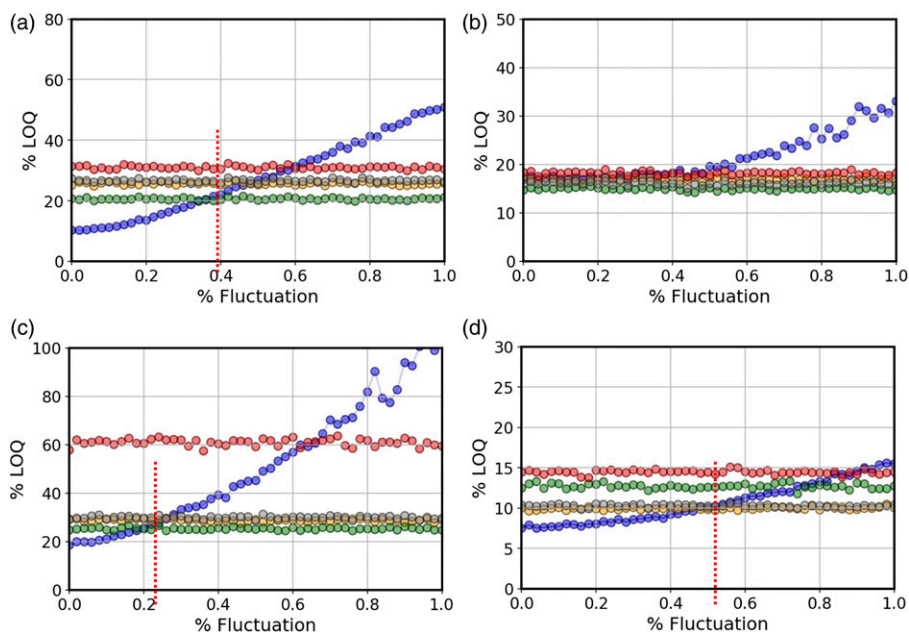


Figure 2. Limit-of-quantification versus spectral variation for various distance calculation methods. LOQ values were calculated using the following comparison sets by varying the entire spectrum by 0–1% under the condition $k = 9$: (a) Herceptin/IgG near-UV, (b) Herceptin/IgG far-UV, (c) VHH1/VHH2 near-UV, and (d) VHH1/VHH2 far-UV. Each point shows the average LOQ calculated by 100 independent iterations of adding noise. Blue: Euclidean distance, yellow: normalized Euclidean distance, green: correlation coefficient, gray: AOO, and red: MAO; the red dotted line indicates the intersection of the Euclidean distance and the LOQ of the method with the lowest LOQ among the normalized methods.

2.0 (JASCO Corporation). The molar absorbance spectra normalized by concentration and optical path length were then spiked numerically and reconverted to HT spectra. The spiked MRE spectral sets generated above do not contain random noise. Therefore, we added noise using random numbers to the data at each wavelength of the spiked MRE spectra. It is known that the magnitude of the noise in CD spectra is not constant for each wavelength. To reproduce this wavelength-dependent noise, we used the HT spectrum sets obtained by the above method. HT spectrum sets were converted to the standard deviation of the noise of CD value at each wavelength using Eq. 12. Based on the obtained standard deviation of noise at each wavelength, random numbers that follow a normal distribution were obtained and convolved with the bandwidth to reproduce the wavelength-dependent random noise set. This wavelength-dependent noise was added to each corresponding spiked MRE spectral set.

Calculation of the Limit-of-Quantification

The LOQs for each distance calculation method were obtained by plotting the distance between a spectrum of 100% Herceptin or VHH1, numerically spiked with a spectrum of IgG or VHH2, respectively, which are considered to be impurities, using Eq. 14 described in ICH-Q2 (R1). S is distance for slope concerning the mixing ratio, obtained by fitting with a linear equation. σ_{resi} is the residual standard deviation fitted

with an appropriate function for the data (Table S4 and Figure S3, Supplemental Material) using Eq. 15. Except for DCA, each function was selected from linear, quadratic, and cubic functions to minimize the residuals for the calculated distances. For DCA, the linear, quadratic, and cubic functions did not fit well because the Eq. 7 contains terms to the twenty-first power, so 21st-order functions were applied. As a representative example, Figure S4 (Supplemental Material) shows the change in the Euclidean distance for the 100% Herceptin spectrum when spiked with a fixed ratio of IgG and the calculated Euclidean distance changes linearly with the mixing ratio of IgG, which is consistent with previous reports.¹⁷

$$LOQ = \frac{10\sigma_{\text{resi}}}{S} \quad (14)$$

$$\sigma_{\text{resi}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}} \quad (15)$$

Smoothing

As a smoothing method for preprocessing of LOQ calculations, Savitzky–Golay filtering was performed under conditions that did not affect the spectral shape (order 2, window size 2.5 nm; near-UV, 5.1 nm; far-UV).

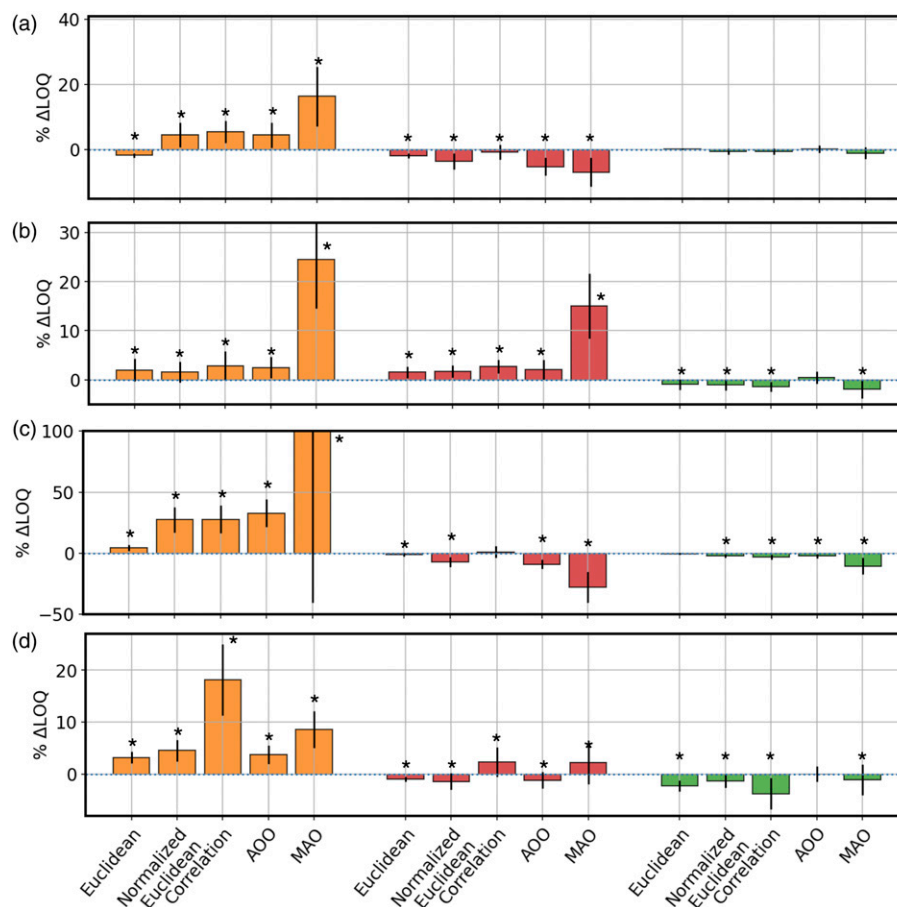


Figure 3. Effectiveness of weighting functions. Differences in LOQs with and without the use of a weighting function were calculated using the comparison sets of (a) Herceptin/IgG near-UV, (b) Herceptin/IgG far-UV, (c) VHH1/VHH2 near-UV, and (d) VHH1/VHH2 far-UV, with noise under the condition $k = 9$. Each bar graph and error bar show the mean and standard deviation of the difference of LOQs calculated 100 times independently with noise. The Manhattan distance and normalized Manhattan distance are omitted because their behavior is the same as the corresponding Euclidean distances; the DCA and derivative correlation coefficients are omitted because the LOQ exceeded 100%. Yellow: spectral intensity weighting function, green: noise weighting function, red: external stimulus weighting function. Two-sample t -tests were performed for LOQs with and without weighting for each method, and results showing $p < 0.05$ are marked with *.

Numerical Computations

All numerical computations and Welch's t -test were performed using Python 3.7.4. with the scientific computation library NumPy 1.17.2 and SciPy 1.3.1.

Results and Discussion

Spiked MRE Spectra with Random Noise

Supplemental Material Figure S5a shows the spiked MRE spectral set for Herceptin/IgG. Supplemental Figure S5b shows the random noise set for Herceptin/IgG in the near-UV region. The spiked MRE spectral set was converted to a CD spectral set, given a random noise set, and then re-converted to MRE spectra to obtain a spiked MRE spectral set with random noise (Figure S5c, Supplemental Material). In the same way, we also simulated the combination of Herceptin/IgG in the far-UV region, VHH1/VHH2 in the near-UV region,

and VHH1/VHH2 in the far-UV region to construct the spectral set. The simulated spectra with noise equivalent to one scan and the measured raw spectra for one scan are in good agreement (Figures S5d, e, f, and g, Supplemental Material).

Comparison of Limit-of-Quantification in the Presence of Spectral Noise

Figure 1 shows the LOQs calculated when wavelength-dependent random noise was added, which corresponds to the measurement condition $k = 9$. The LOQs for all the methods that use differentiation, such as the DCA and derivative correlation coefficient, exceed 100% for all the comparison sets. Also, in many comparison sets (Herceptin/IgG near-UV, VHH1/VHH2 near-UV, far-UV), the distance calculation methods that use spectral normalization (normalized Euclidean distance, normalized Manhattan distance,

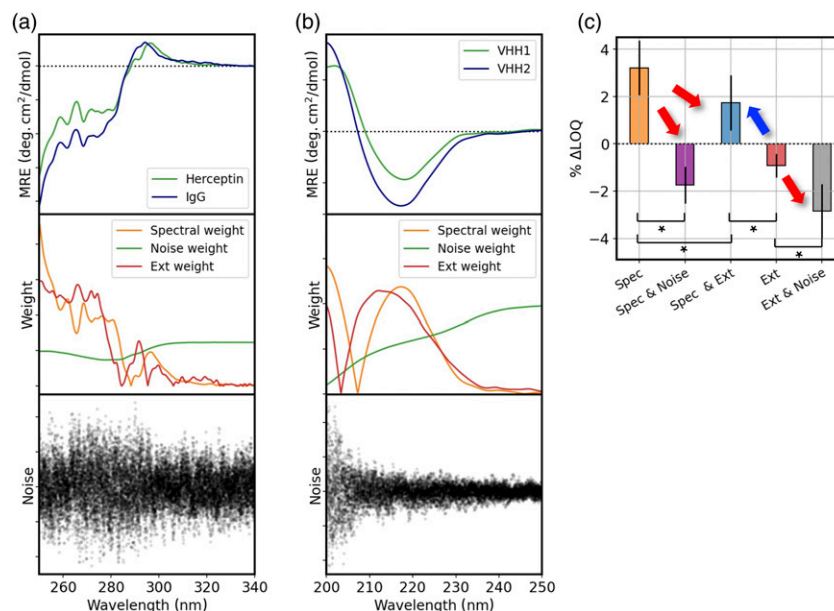


Figure 4. Comparison of the shapes of the various weighting functions and the distribution of the noise in the spectra. The original spectra (top), the weighting functions (middle), and the simulated noise under the condition $k = 9$ (bottom) are shown for the (a) Herceptin/IgG near-UV and (b) VHH1/VHH2 far-UV comparison sets. (c) The difference in LOQ with and without the weighting function. Each bar and error bar shows the mean and standard deviation of the difference in LOQ calculated by 100 independent iterations of noise assignment. Yellow: spectral intensity weighting, purple: combined spectral intensity and noise weighting, light blue: combined spectral intensity and external stimulus weighting, red: external stimulus weighting, gray: combined external stimulus and noise weighting. Two-sample *t*-tests were performed for ΔLOQ with weighting alone and with double weighting, and results showing $p < 0.05$ are marked with *.

correlation coefficient, AOO, and MAO) show higher LOQs than the methods that do not use normalization (Euclidean distance and Manhattan distance).

The advantage of spectral normalization is that it allows the evaluation of the distance between spectra without being affected by the variation of the entire spectrum due to sample preparation errors. On the other hand, as shown in Figures S6a, c, and d (Supplemental Material), the normalization process causes a reduction of the differences between reference and sample spectra. The high LOQ for the distance calculation method with spectral normalization is due to this property. For Herceptin/IgG far-UV, a difference in LOQ was not observed between the spectra with and without normalization. This is a rare case in which the difference between the two spectra is maintained even after normalization (Figure S6b, Supplemental Material).

The Euclidean distance and the Manhattan distance, known for the most famous distance calculation methods, did not show significant differences in many cases.

Effect of Variation of Absorbance on Limit-of-Quantification

As shown above, the Euclidean and Manhattan distances without normalization showed the best performance when only the noise in the spectrum was considered. However, in

actual measurements, it is necessary to consider not only the noise in the spectrum but also the variation of the entire spectrum caused by sample preparation errors. In general, the sample preparation error is assumed to be a few percent, which can be compensated for by converting the raw spectrum to the MRE spectrum using the concentration converted from the absorbance measured simultaneously with the CD spectrum. However, the photometric value for absorbance itself always contains some variation, resulting in an error in the conversion to MRE, and as a result, the variation of absorbance becomes a variation of the entire MRE spectrum. Therefore, we investigated the effect of the variation of the entire spectrum caused by the variation of the measured absorbance value on the LOQ for the various distance calculation methods.

Figure 2 shows the LOQ for each distance calculation method when the entire spectrum is subjected to variation for the MRE spectral set with spectral noise under the condition $k = 9$. The variation is assumed to follow a normal distribution with a mean of 0 and a standard deviation of 0–1%. In all comparison sets, the LOQ of the Euclidean distance increases with the variability added to the whole spectrum, while the LOQs of the normalized Euclidean distance, correlation coefficient, AOO, and MAO are unaffected. For each comparison set, the level of variation was estimated to be around 0.2–0.5% in regions where the LOQ of the Euclidean distance and the LOQ of the method with the smallest LOQ among the normalization methods are reversed.

Since the variation of the whole spectrum can be regarded as the variation of the measured absorbance value, it is suitable to choose the Euclidean distance or Manhattan distance without normalization when the variation of the measured absorbance value can be reduced to less than 0.2–0.5%. If this is difficult, it is appropriate to choose a distance calculation method that performs spectral normalization.

Effects of the Weighting Function

To confirm the effectiveness of the weighting functions, we performed a comprehensive analysis of the LOQ for each distance calculation method and weighting function combination (Fig. 3). The spectral intensity weighting function is based on the 100% MRE spectra of Herceptin and VHH1, and the noise weighting function is based on the HT spectra of each dataset. The external stimulus weighting functions are obtained from the subtraction spectra between 100% Herceptin and 100% IgG and the subtraction spectra between 100% VHH1 and 100% VHH2.

Each bar graph shows the change in LOQ before and after applying the weighting function to each distance calculation method, where a negative value of ΔLOQ indicates an improvement in LOQ. In the Herceptin/IgG near-UV comparison set, which is considered to be the closest to the conditions for comparability evaluation of antibody drugs currently in practice, the spectral intensity weighting function, in combination with the Euclidean distance, resulted in a negative ΔLOQ , indicating that it was effective in improving the LOQ. On the other hand, combining other comparison sets and distance calculation methods resulted in a positive ΔLOQ , which worsened the LOQ. The external stimulus weighting function worsened the LOQ in two cases but improved the LOQ in many combinations. Although the noise weighting function had a small effect on improving the LOQ, it consistently improved the LOQ for many comparison sets and distance calculation methods.

To investigate the cause of the LOQ increase caused by the spectral intensity weighting function, we examined the relationship between the weighted wavelength range, the range where the spectrum changes, and the distribution of noise in the comparison sets where the spectral intensity weighting function worked effectively or unfavorably (Fig. 4). In the comparison set for Herceptin/IgG near-UV and Euclidean distance where the spectral intensity weighting function worked effectively, the region around 250–260 nm was weighted (Fig. 4a). While this wavelength range is relatively consistent with the region of large spectral difference, it is not compatible with the wavelength near 280 nm, where the noise is significant.

In the VHH1/VHH2 far-UV comparison set, where the spectral intensity weighting function has an unfavorable effect, the spectral intensity weighting function weights the region around 200 nm (Fig. 4b). The region around 200 nm overlaps

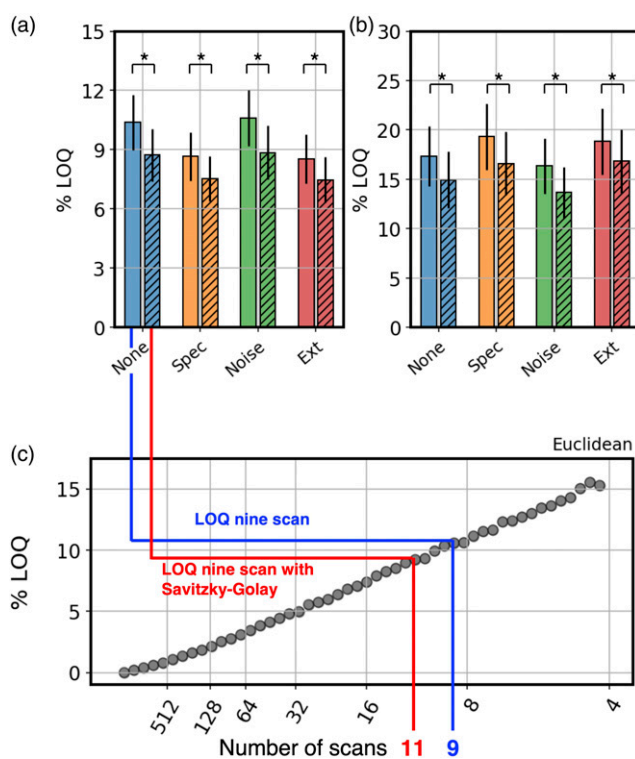


Figure 5. Effect of smoothing. The LOQs were calculated by combining the Euclidean distance and the Savitzky–Golay filter in the (a) Herceptin/IgG near-UV comparison set and (b) in the far-UV comparison set under the condition $k = 9$. The bars and error bars show the mean and standard deviation of the LOQs calculated after 100 independent trials of adding noise. Blue: no weighting function, yellow: spectral intensity weighting function, red: external stimulus weighting function. Open bar: no smoothing, hatched bar: with smoothing. A two-sample t -test was performed on the LOQs before and after noise removal, and the results showing $p < 0.05$ are marked with *. The LOQ values for Euclidean distance are calculated using spike spectral sets generated under the conditions of the various number of (c) scan k . The scatter plot shows the mean of the LOQs calculated after 100 independent trials of adding noise equivalent to the number of scan.

with the wavelength region where the spectrum changes, but it also coincides with the wavelength region where the noise is the largest. This may cause a deterioration of the LOQ, offsetting the improvement of the LOQ by the spectral intensity weighting. The same is true for the external stimulus weighting function.

Effects of Weighting Function Combination

As described above, the coincidence of the weighting wavelength range and the wavelength with high noise range leads to the deterioration of the LOQ. Therefore, we expected that the use of noise weighting functions in combination with spectral intensity weighting function and external stimulus weighting functions would be effective in

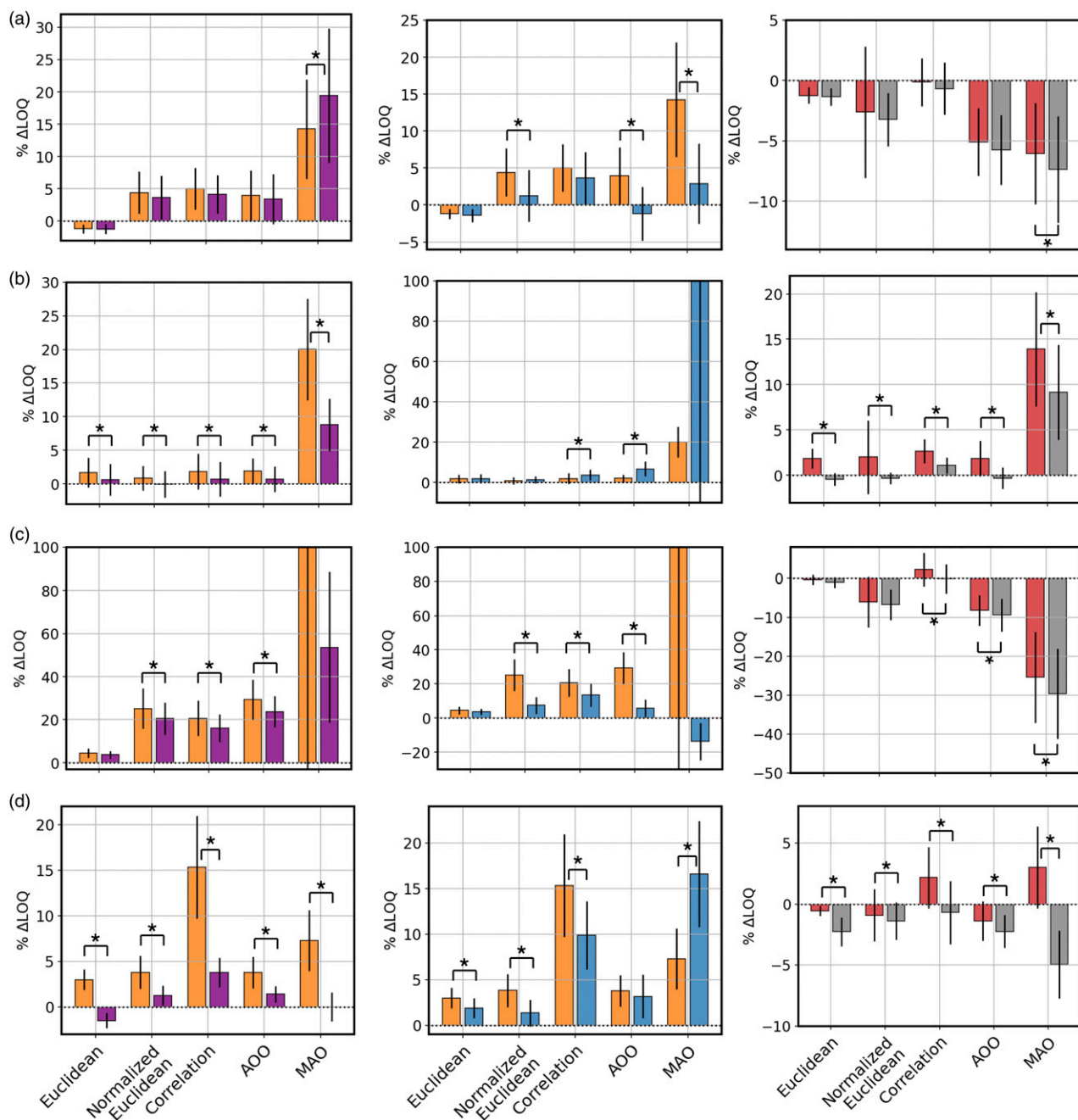


Figure 6. Comprehensive analysis of the effect of weighting function combination. Differences in LOQs with and without the use of weighting functions were calculated for (a) Herceptin/IgG near-UV, (b) Herceptin/IgG far-UV, (c) VHH1/VHH2 near-UV, and (d) VHH1/VHH2 far-UV, with noise under the condition $k = 9$. Yellow: spectral intensity weighting function alone, purple: dual-use of spectral intensity and noise weighting functions, light blue: dual-use of spectral intensity and external stimulus weighting functions, red: external stimulus weighting function alone, gray: dual-use of external stimulus and noise weighting functions. Each bar graph and error bar show the mean and standard deviation of the difference in LOQs calculated 100 times trials of adding noise. Each spectrum is preprocessed by a Savitzky–Golay filter. A two-sample t-test was performed on the Δ LOQ for the single weighted and double weighted double combination, and results showing $p < 0.05$ are marked with *.

preventing the deterioration of the LOQ and investigated the effect of the combined use of weighting functions (Fig. 4c). Interestingly, the positive Δ LOQ resulting from the application of spectral intensity weighting function

became negative when the noise weighting function was also used, confirming the improvement. The external stimulus weighting function was also found to improve the LOQ in the same way. The combination of spectral intensity

and external stimulus weighting function resulted in a smaller Δ LOQ than spectral intensity weighting alone and a larger Δ LOQ than external stimulus weighting alone. This result is consistent with the spectral intensity weighting function weights region with higher noise than the external stimulus weighting function.

Effects of Smoothing

Since the above examination results suggest that spectral noise has a significant effect on various distance methods and weighting functions, we investigated the effect of noise reduction using a Savitzky–Golay filter as a preprocessing step for spectral distance calculations. Figs. 5a and 5b show the LOQ before and after smoothing at the Euclidean distance for Herceptin/IgG as a representative example. It can be seen that the smoothing is effective regardless of the presence or absence of the weighting function and the type of the weighting function in the near-UV and far-UV comparison sets.

Figure 5c shows the relationship between the number of scan and the LOQ. This graph shows that although the Savitzky–Golay filter is effective, it is not able to remove the effect of noise completely, and its LOQ improvement effect is limited to an increase of about two scans.

Conclusion

The performance of various distance calculation methods and weighting methods was comprehensively evaluated by simulating spectral fluctuations caused by spectral noise and sample preparation errors. The results indicated that the method for calculating the distance using differentiation was not suitable for comparing the differences between CD spectra. It was also found that the distance calculation methods that did not use spectral normalization, such as the Euclidean distance, are appropriate when the sample concentration can be corrected by a device that can reduce the variation of absorbance to less than 0.2–0.5%, and spectral normalization methods are appropriate when that is difficult. No systematic performance differences were found between the Euclidean and Manhattan distances. Similarly, no systematic performance differences were observed among the normalized Euclidean distance, normalized Manhattan distance, correlation coefficient, and AOO.

The weighting functions effectively improved the LOQ for the spectral distance calculation method in many comparison sets. While the spectral intensity and external stimulus weighting functions were adequate for the Herceptin/IgG near-UV comparison sets, there were cases where the weighting wavelengths coincided with significant noise wavelengths, resulting in a deterioration of the LOQ. The noise weighting function consistently improved the LOQ, although the effect was moderate. The noise

weighting function was combined with the spectral intensity and external stimulus weighting functions to improve the LOQ, which was deteriorated by the application of these functions. To confirm the effectiveness and versatility (stability) of the triple combination method, which combines the double combination of weighting functions with an added Savitzky–Golay filter, we comprehensively analyzed the LOQ change for each comparison set and distance calculation method (Fig. 6). Each LOQ change was pre-processed using the Savitzky–Golay filter. The LOQ was improved by adding the noise weighting function to the spectral intensity weighting function, even under the condition that the noise reduction method was combined. There was only one case in which the LOQ deteriorated with the use of the noise weighting function. The combination of the external stimulus weighting function also improved the LOQ of the spectral intensity weighting function, but the LOQ deteriorated in three cases. When the external stimulus weighting function was used alone, the LOQ deteriorated in seven cases, but in all of these cases, the LOQ was improved by the combined use of the noise weighting function, confirming the effectiveness of the double use of noise weighting functions and the triple use of noise reduction methods.

Considering the above results, we can choose a non-normalizing method such as the Euclidean distance when using a modern CD spectrometer or measurement environment with a high measurement accuracy of CD and absorbance. As a preprocessing method, noise reduction using the Savitzky–Golay filter is effective. For the weighting function, it is preferable to combine the spectral intensity weighting function, which improves the LOQ under conditions closest to those used for comparability evaluation of antibody drugs, and the noise weighting function, which is effective in reducing unexpected deterioration of the LOQ. In addition, the introduction of the external stimulus weighting function should be considered to improve the sensitivity. Thus, combining the distance calculation method and the weighting function with consideration of the spectral changes and noise distributions will be possible to evaluate the spectral similarity of antibody drugs in a highly sensitive and robust manner.

The guidelines for the various distance calculation methods and the selection of weighting functions presented in this study were obtained by simulating spectral noise and sample preparation errors. More useful insights may be obtained by combining them with experimental evaluations in the future. On the other hand, this simulation-based methodology allowed us to statistically discuss the differences in the performance based on the results obtained from the number of trials, which is difficult to achieve in actual measurements. In addition to comparing two different molecules, a comparison of the performance under different conditions, such as denaturation and aggregation of antibodies with native

structures, may provide guidelines for the use of methods that did not show systematic differences in this study. Further comparison of the performance of spectral distance calculation methods will contribute to a more sensitive and robust evaluation of the comparability of the HOS of antibody drugs and will also lead to the development of objective evaluation methods for differences in HOS not only between antibodies, but also between various biomolecules.

Acknowledgments

We are very grateful to Dr. Akikazu Murakami for generously providing us with VHH antibodies.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

CRedit Author Statement

Taiji Oyama: Conceptualization, software, methodology, visualization, writing original draft. Satoko Suzuki: Conceptualization, investigation, methodology, writing, review, and editing. Yasuo Horiguchi: Conceptualization, investigation, writing, review, and editing. Ai Yamane: Data curation, investigation, writing, review, and editing. Kenichi Akao: Project administration, resources, writing, review, and editing. Koushi Nagamori: Conceptualization, methodology, supervision, writing, review, and editing. Kouhei Tsumoto: Supervision, writing, review, and editing.

ORCID iD

Taiji Oyama  <https://orcid.org/0000-0003-1403-073X>

Supplemental Material

All supplemental material mentioned in the text is available in the online version of the journal.

References

1. R.W. Woody. "Circular Dichroism". *Methods Enzymol.* 1995. 2(46): 34–71. doi: [10.1016/0076-6879\(95\)46006-3](https://doi.org/10.1016/0076-6879(95)46006-3).
2. N.J. Greenfield. "Using Circular Dichroism Spectra to Estimate Protein Secondary Structure". *Nat. Protoc.* 2006. 1(6): 2876–2980. doi: [10.1038/nprot.2006.202](https://doi.org/10.1038/nprot.2006.202).
3. R.W. Woody. "The Development and Current State of Protein Circular Dichroism". *Biomed. Spectrosc. Imaging.* 2015. 4(1): 5–34. doi: [10.3233/BSI-140098](https://doi.org/10.3233/BSI-140098).
4. B.A. Wallace. "The Role of Circular Dichroism Spectroscopy in the Era of Integrative Structural Biology". *Curr. Opin. Struct. Biol.* 2019. 58: 191–196. doi: [10.1016/j.sbi.2019.04.001](https://doi.org/10.1016/j.sbi.2019.04.001).
5. N.J. Greenfield. "Using Circular Dichroism Collected as a Function of Temperature to Determine the Thermodynamics of Protein Unfolding and Binding Interaction". *Nat. Protoc.* 2006. 1(6): 2527–2535. doi: [10.1038/nprot.2006.204](https://doi.org/10.1038/nprot.2006.204).
6. J.T. Yang, C.C. Wu, H.M. Martinez. "Calculation of Protein Conformation from Circular Dichroism". *Methods Enzymol.* 1986. 130: 208–269. doi: [10.1016/0076-6879\(86\)30013-2](https://doi.org/10.1016/0076-6879(86)30013-2).
7. N. Sreerama, R.W. Woody. "Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set". *Anal. Biochem.* 2000. 287(2): 252–260. doi: [10.1006/abio.2000.4880](https://doi.org/10.1006/abio.2000.4880).
8. A. Micsonai, F. Wien, L. Kernya, H.Y. Lee, et al. "Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy". *Proc. Natl. Acad. Sci. U.S.A.* 2015. 112(24): 3095–3103. doi: [10.1073/pnas.1500851112](https://doi.org/10.1073/pnas.1500851112).
9. A. Micsonai, F. Wien, E. Bulyáki, J. Kun, et al. "BeStSel: A Web Server for Accurate Protein Secondary Structure Prediction and Fold Recognition from the Circular Dichroism Spectra". *Nucleic Acids Res.* 2018. 46(1): 315–322. doi: [10.1093/nar/gky497](https://doi.org/10.1093/nar/gky497).
10. A. Micsonai, F. Wien, E. Bulyáki, J. Kardos. "BeStSel: From Secondary Structure Analysis to Protein Fold Prediction by Circular Dichroism Spectroscopy". *Methods Mol. Biol.* 2021. 2199: 175–189. doi: [10.1007/978-1-0716-0892-0_11](https://doi.org/10.1007/978-1-0716-0892-0_11).
11. C.H. Li, X. Nguyen, L. Narhi, L. Chemmalil, et al. "Applications of Circular Dichroism (CD) for Structural Analysis of Proteins: Qualification of Near- and Far-UV CD for Protein Higher Order Structural Analysis". *J. Pharm. Sci.* 2011. 100(11): 4642–4654. doi: [10.1002/jps.22695](https://doi.org/10.1002/jps.22695).
12. C.F. Kirchhoff, X.M. Wang, H.D. Conlon, S. Anderson, et al. "Biosimilars: Key Regulatory Considerations and Similarity Assessment Tools". *Biotechnol. Bioeng.* 2017. 114(12): 2696–2705. doi: [10.1002/bit.26438](https://doi.org/10.1002/bit.26438).
13. N. Lee, J.J. Lee, H. Yang, S. Baek, et al. "Evaluation of Similar Quality Attribute Characteristics in SB5 and Reference Product of Adalimumab". *MAbs.* 2019. 11(1): 129–144. doi: [10.1080/19420862.2018.1530920](https://doi.org/10.1080/19420862.2018.1530920).
14. K. Pisupati, A. Benet, Y. Tian, S. Okbazghi, et al. "Biosimilarity Under Stress: A Forced Degradation Study of Rmicade and Rmsima". *MAbs.* 2017. 9(7): 1197–1209. doi: [10.1080/19420862.2017.1347741](https://doi.org/10.1080/19420862.2017.1347741).
15. N.N. Dinh, B.C. Winn, K.K. Arthur, J.P. Gabrielson. "Quantitative Spectral Comparison by Weighted Spectral Difference for Protein Higher Order Structure Confirmation". *Anal. Biochem.* 2014. 464: 60–62. doi: [10.1016/j.ab.2014.07.011](https://doi.org/10.1016/j.ab.2014.07.011).
16. JASCO Corporation. "Spectrum QC Test Program Software Manual". 2017. 47–53.
17. B.M. Teska, C. Li, B.C. Winn, K.K. Arthur, et al. "Comparison of Quantitative Spectral Similarity Analysis Methods for Protein Higher-Order Structure Confirmation". *Anal. Biochem.* 2013. 434(1): 153–165. doi: [10.1016/j.ab.2012.11.018](https://doi.org/10.1016/j.ab.2012.11.018).
18. B.S. Kendrick, J.P. Gabrielson, C.W. Solsberg, E. Ma, et al. "Determining Spectroscopic Quantitation Limits for Misfolded Structures". *J. Pharm. Sci.* 2020. 109(1): 933–936. doi: [10.1016/j.xphs.2019.09.004](https://doi.org/10.1016/j.xphs.2019.09.004).
19. C. Jones. "Impact of Imperfect Data on the Performance of Algorithms to Compare Near-Ultraviolet Circular Dichroism

- Spectra". *Appl. Spectrosc.* 2021. 75(7): 857–866. doi: [10.1177/0003702821992370](https://doi.org/10.1177/0003702821992370).
20. A.J. Miles, R.W. Janes, B.A. Wallace. "Tools and Methods for Circular Dichroism Spectroscopy of Proteins: A Tutorial Review". *Chem. Soc. Rev.* 2021. 50(15): 8400–8413. doi: [10.1039/D0CS00558D](https://doi.org/10.1039/D0CS00558D).
21. J. D'antonio, B.M. Murphy, M.C. Manning, W.A. Al-Azzam. "Comparability of Protein Therapeutics: Quantitative Comparison of Second-Derivative Amide I Infrared Spectra". *J. Pharm. Sci.* 2012. 101(6): 2025–2033. doi: [10.1002/jps.23133](https://doi.org/10.1002/jps.23133).
22. A. Rinnan, L. Nørgaard, F. Van Den Berg, J. Thygesen, et al. "Data Pre-Processing". In: D.-W. Sun, editor. *Infrared Spectroscopy for Food Quality Analysis and Control*. Amsterdam: Elsevier/Academic Press, 2009. Chap. 2, Pp. 29–47. doi: [10.1016/B978-0-12-374136-3.X0001-6](https://doi.org/10.1016/B978-0-12-374136-3.X0001-6).
23. B.S. Kendrick, A. Dong, S.D. Allison, M.C. Manning, J.F. Carpenter. "Quantitation of the Area of Overlap Between Second-Derivative Amide I Infrared Spectra to Determine the Structural Similarity of a Protein in Different States". *J. Pharm. Sci.* 1996. 85(2): 155–158. doi: [10.1021/js950332f](https://doi.org/10.1021/js950332f).