Comparability of Higher Order Structure in Proteins: Chemometric Analysis of Second-Derivative Amide I Fourier Transform Infrared Spectra

GREGORY STOCKDALE,[1] BRIAN M. MURPHY,[2,3] JENNIFER D'ANTONIO,[4] MARK CORNELL MANNING,[2,3] WASFI AL-AZZAM[5]

[1]Department of Statistical Sciences, Projects Clinical Platform, R&D, GlaxoSmithKline, King of Prussia, Pennsylvania 19406
[2]Legacy BioDesign LLC, Johnstown, Colorado 80534
[3]Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523
[4]Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695
[5]Biopharm Analytical Sciences, BioPharmaceutical Development, R&D, GlaxoSmithKline, King of Prussia, Pennsylvania 19406

Received 14 December 2013; revised 2 September 2014; accepted 8 September 2014

Published online 7 November 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.24218

ABSTRACT: Comparing higher order structure (HOS) in therapeutic proteins is a significant challenge. Previously, we showed that changes in solution conditions produced detectable changes in the second-derivative amide I Fourier transform infrared (FTIR) spectra for a variety of model proteins. Those comparisons utilized vector-based approaches, such as spectral overlap and spectral correlation coefficients to quantify differences between spectra. In this study, chemometric analyses of the same data were performed, to classify samples into different groups based on the solution conditions received. The solution conditions were composed of various combinations of temperature, pH, and salt types. At first, principal component analysis (PCA) was used to visually demonstrate that FTIR spectra respond to changes in solution conditions, which, presumably indicates variations in HOS. This observed when samples from the same solution condition form clusters within a PCA score plot. The second approach, called soft independent modeling of class analogy (SIMCA), was conducted to account for the within-class experimental error for the lysozyme spectra. The DModX values, indicative of the distance of each spectra to their respective class models, was found to be a more sensitive quantitative indicator of changes in HOS, when compared with the modified area of overlap algorithm. The SIMCA approach provides a metric to determine whether new observations do, or do not belong to a particular class or group. Thus, SIMCA is the recommended approach when multiple samples from each condition are available.
© 2014 Wiley Periodicals, Inc. and the American Pharmacists Association J Pharm Sci 104:25–33, 2015
Keywords: protein structure; FTIR; spectroscopy; proteins; chemometrics; biopharmaceuticals characterization; stability

INTRODUCTION

Protein therapeutics requires a native-like structure in order to maintain biological activity and minimize immunogenicity. Therefore, the assessment of higher order structure (HOS) in proteins, that is, the analysis of the tertiary and secondary structure of proteins, is essential to ensure the quality of therapeutic proteins. The US Food and Drug Administration (FDA) has recently suggested that the biopharmaceutical industry should invest more effort into the HOS characterization of biological products as early as possible in their development stage.[1] Comparability studies are necessary to assess product quality and batch-to-batch variability to ensure quality and safety of protein therapeutics. One difficulty is how comparability should be determined when utilizing spectral data. In our previous study, different linear algorithms were evaluated.[2] Although these methods clearly provide methodology for quantitative assessment of spectral similarity, they do not consider more than one wavelength or portion of the spectrum at a time. They also fail to account for analytical variability when multiple samples are available. To conduct similarity testing with multiple replicates, one must be able to discern observed differences between samples from the noise, or variation, inherent in sampling and the precision of the measurement system. This may be achieved by employing chemometric (i.e., multivariate statistical) methods.[3] This study examines principal component analysis (PCA) and the related soft independent modeling of class analogy (SIMCA) approach as potential chemometric tools that could be used to assess spectral similarity, which would indicate comparability of HOS in protein samples. These chemometric methods are well known and have been widely used to deconvolute spectral data of complex mixtures, including detection of impurities,[4] analysis of tissue composition,[5] analysis of complex foodstuffs,[6] and changes in structure upon aggregation.[7]

It should be noted that this study focuses on demonstrating the utility of chemometric methods for quantitative comparison and visualization of spectral differences as a function of changing solution conditions. The work is not meant to identify a definitive metric that could be used in a regulatory setting as a determination of whether two samples are similar enough to be declared comparable. The purpose is to move beyond linear algorithms and equip researchers with viable mathematical tools for quantifying differences in spectral data sets that reflect distinct changes in HOS. Comparability testing has many facets, with the ultimate goal of relating these process-related changes to clinical outcomes. However, that connection has yet

Correspondence to: Wasfi Al-Azzam (Telephone: +610-270-4188; Fax: +610-270-6996; E-mail: wasfi.a.al-azzam@gsk.com)

Journal of Pharmaceutical Sciences, Vol. 104, 25–33 (2015)
© 2014 Wiley Periodicals, Inc. and the American Pharmacists Association

to be demonstrated in complete detail. Although the link between CMC-related measurements of the product and clinical qualities are generally assumed, one does not necessarily know how much of a change can be tolerated. What is clear from this work, along with previous studies,[2,8] is that changes in HOS can and do occur as a function of differences in batches, processing, handling, storage, and so on. Certainly, further work is needed to incorporate sources of variability beyond that seen in a single lab/day/analyst situation. However, the point of this study is to demonstrate that readily available mathematical tools can help differentiate samples from each other based on spectral properties and these techniques allow one to quantify how different samples are in terms of HOS, thereby providing an assessment of spectral differences based, not just on a single variable, but on differences seen across the entire spectrum.

MATERIALS AND METHODS

Materials

Human serum immunoglobin (IgG) (purity $\geq$ 95%), equine heart myoglobin (purity $\geq$ 95%), and chicken egg-white lysozyme (purity $\geq$ 95%) were purchased from Sigma–Aldrich(St. Louis, MO) and used without further purification. Sodium phosphate monobasic, sodium acetate trihydrate, sodium hydroxide, hydrochloric acid, acetic acid, and glycine were used as received from Sigma–Aldrich.

Preparation of Protein Samples

A 50-mg/mL protein solution was prepared by dissolving the appropriate amount of protein in the corresponding buffer. Sodium acetate, sodium phosphate, and sodium hydroxide/glycine were used to prepare pH 4, 7, and 10 buffer solutions (20 mM), respectively. All samples were freshly prepared prior to each measurement.

In addition, some protein samples, such as lysozyme, included either sodium sulfate (50 mM) or sodium thiocyanate (50 mM), representing anions at opposite ends of the Hofmeister series, with the expectation that these salts would either promote secondary structure or disrupt it, respectively.

Fourier Transform Infrared Spectroscopy

Fourier transform infrared (FTIR) spectra were acquired with a Nicolet 4700 instrument (Thermoelectron Corporation, Waltham, MA) equipped with KBr beam splitter and deuterium triglycine sulfate detector. Prior to each measurement, a 10-min wait period was allowed to purge the sample chamber with dry N_2 gas in order to reduce water vapor and CO_2 IR interference and equilibrate the sample to the corresponding temperature. Spectra were recorded in the mid infrared region of 4000–400 cm^{-1} in the transmission mode with a resolution of 4 cm^{-1} resolution and a 256 scans accumulation using the OMNIC software. A 20-μL aliquot of the aqueous protein solution was placed on a CaF_2 temperature controlled Biocell (Biotools Inc., Jupiter, FL) with a 6-μm path length. Spectra were collected at various temperatures: 5°C, 25°C, and 60°C. The temperature of the experiment was controlled through a circulating water bath (Julabo, Allentown, PA). All spectra were appropriately background subtracted. Spectra were obtained from three preparations from each sample. Reference spectra recorded for the buffer under identical conditions were subtracted using 6.01 GRAMS software (ThermoGalactic, Waltham, MA) from the sample spectra and the region near 1800 cm^{-1} was verified to display a flat line flat after buffer subtraction. The overlapping amide I bands were subjected to resolution enhancement using a second-derivative Savitzky–Golay function with a degree of 3 and smoothed using the Savitzky–Golay seven-point function. Baseline correction was performed using both end points of the amide I region (1600 and 1720 cm^{-1}) as well as the valleys between these points depending on the protein under analysis. Each spectrum was area normalized to an area of unity and then subjected to a 32x space-filling interpolation.

Spectral Comparison Algorithms

Spectral correlation coefficients (SCC) and area of overlap (AO) were calculated using literature algorithms[9] performed using Excel. The maximum signal between two spectra at each wavenumber was recorded to create an overlap spectrum. The area of the overlap spectrum was taken as a fraction of the total-normalized spectrum area, and recorded as a percentage.

$$\frac{(S_{\mathrm{2ndDer.}})^2}{A_{\mathrm{sq}}} = S_{\mathrm{ModAO}}$$

Principal Component Analysis

The PCA technique is a multivariate statistical method that has been in existence for many years under a variety of names.[3,10,11] It allows one to analyze and evaluate a multidimensional data matrix. It utilizes the correlation structure and accounts for random variation (noise) to reduce the number dimensions to a small set of new variables called principal components. By removing noise from the underlying "true" information, one can classify observations into appropriate groups and identify relationships between objects. Analysis of the IR spectra by PCA result in score plots shows the relationship between different spectra in terms of the primary principal components, as well as loading plots, which show how different variables (in this case, the intensities at each wavenumber) impact the final PCA model. On the score pots, one often plots the Hotelling T2 ellipse, which encircles all samples that are found to be within a 95% confidence region of the model (at a p value of 0.05).

Soft Independent Modeling of Class Analogy

The PCA model constructed allows one to visually see the clustering of samples exposed to the same solution conditions. However, it does not provide a quantitative way to determine whether the spectra obtained for one solution condition (class/group) is the same, or different, from another. This is because it does not separate the within-class variation from the between-class variation. The SIMCA approach constructs a model for each class, which, in this case, is each solution condition. Once a PCA model for each class is constructed, it is possible to compare any new sample to that model in order to determine whether the new sample belongs to the class described by the model.[11] In this study, an independent PCA model was constructed for each set of solution conditions for

Download English Version:

https://daneshyari.com/en/article/2484618

Download Persian Version:

https://daneshyari.com/article/2484618

Daneshyari.com