RAPID COMMUNICATION

Comparative Signature Diagrams to Evaluate Biophysical Data for Differences in Protein Structure Across Various Formulations

VIDYASHANKARA IYER,[1,2] NATHAN MADDUX,[3] LEI HU,[2] WEIQIANG CHENG,[2] AHMED K. YOUSSEF,[4] GERHARD WINTER,[4] SANGEETA B. JOSHI,[2] DAVID B. VOLKIN,[1,2] C. RUSSELL MIDDAUGH[1,2]

[1]Bioengineering Graduate Program, University of Kansas, Lawrence, Kansas 66045

[2]Macromolecule and Vaccine Stabilization Center, Department of Pharmaceutical Chemistry, University of Kansas, Lawrence, Kansas 66047

[3]Department of Physics, University of Kansas, Lawrence, Kansas 66047

[4]Pharmaceutical Technology and Biopharmaceutics, Department of Pharmacy, University of Munich, Munich 81377, Germany

Received 24 August 2012; revised 15 October 2012; accepted 19 October 2012

Published online 15 November 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.23367

ABSTRACT: A solution to the problem of being able to show statistically significant differences in the measurements of various levels of higher-order protein structure has been an elusive one. We propose the use of comparative signature diagrams (CSDs) to this end. CSDs compare datasets from different biophysical techniques that fingerprint the secondary, tertiary, and quaternary structures of a protein molecule and display statistically significant differences in these datasets. In this paper, we explore the differences in the structures of two proteins (Granulocyte Colony Stimulating Factor [GCSF] and a monoclonal antibody [mAb]) in various formulations. These proteins were chosen based on the extent of differences in structure observed in the formulations. As an initial test, we utilize data from circular dichroism, 8-anilino-1-naphthalene-sulfonate and intrinsic fluorescence spectroscopy, and static light scattering measurements to fingerprint protein structure in the different formulations. Several layers of statistics were explored to visualize the regions of significant differences in the protein spectra. This approach provides a rapid, high-resolution methodology to compare various structural levels of proteins using standard biophysical instrumentation. © 2012 Wiley Periodicals, Inc. and the American Pharmacists Association J Pharm Sci 102:43–51, 2013

Keywords: comparative signature diagrams; comparability; biosimilars; empirical phase diagrams; CD; light scattering; fluorescence

INTRODUCTION

An important biopharmaceutical issue is being able to detect statistically significant differences and similarities between two proteins or an individual protein under two different conditions (e.g., manufacturing process, formulations, etc.). This analytical challenge is especially important when dealing with comparability during the development of protein therapeutics[1,2] as well as with the preparation of biosimilars.[3–5] Proteins have several layers of higher-order structure and subtle differences in these structures may be crucial for the activity and safety of the protein. Although NMR and X-ray diffraction studies are able to provide accurate three-dimensional structures of proteins, these techniques are time consuming, labor intensive, and typically not applicable under relevant pharmaceutical conditions (excipients, protein concentration, etc.). In such cases, lower-resolution techniques such as circular dichroism (CD), Fourier transform infrared (FTIR) spectroscopy, intrinsic and extrinsic fluorescence spectroscopy, and light scattering may be used to study the secondary, tertiary, and quaternary structure of proteins. Large datasets from various biophysical techniques have been evaluated, combined, and visualized by constructing empirical phase diagrams (EPDs). EPDs have been successfully employed to study the overall structural integrity and conformational stability of various proteins and macromolecules.[6–18]

Correspondence to: C. Russell Middaugh (Telephone: +785-864-5813; Fax: +785-864-5814; E-mail: middaugh@ku.edu)

Weiqiang Cheng's present address is Genzyme, Framingham, Massachusetts 01701.

Journal of Pharmaceutical Sciences, Vol. 102, 43–51 (2013)

When comparing potential structural differences of a protein in different formulations or when evaluating a protein made after a process change (e.g., comparability) or made by different manufacturers (e.g., biosimilars), the currently employed EPD data analysis approach may not be an ideal analytical tool for this purpose. For example, much of the individual spectral information is lost during EPD analysis and only transitions in structural states of a protein as detected at certain, specific wavelengths are retained. By not employing the complete spectra and monitoring changes at multiple wavelengths, critical structural information that may be useful in assessing the structural similarity of two protein samples is lost. Although the analytical accuracy and precision of certain individual biophysical techniques [which examine one aspect of protein structure such as CD, FTIR, ultraviolet (UV) spectroscopy, and differential scanning calorimetry (DSC)] have recently been evaluated from a pharmaceutical perspective,[19–22] the use of data from individual techniques does not evaluate the overall structural integrity and stability of a protein in the same way as EPD analysis (which combines datasets across different techniques and various environmental stresses). The complete spectrum from CD, FTIR spectroscopy, intrinsic and extrinsic fluorescence, and light scattering signals combined with its change upon environmental stresses such as temperature and pH can provide a characteristic signature of the higher-order structure and stability of a protein. This signature could potentially be very useful in studies of comparability and in the development of biosimilars. In this communication, we propose the use of comparative signature diagrams (CSDs) for the visualization of statistically significant differences between large datasets from multiple spectroscopic techniques when analyzing the conformational stability of two different proteins (Granulocyte Colony Stimulating Factor [GCSF] and a monoclonal antibody [mAb]) in different formulations.

MATERIALS AND METHODS

Description of the CSD

Several CSDs were constructed based on the approach described below to determine the difference between structures a protein in several different formulations. The two proteins tested were GCSF and a mAb. These proteins were chosen because GCSF displays a variable stability profile in different buffers. In contrast, the mAb being highly stable is more resilient to small changes in formulation composition. The various protein formulations were analyzed using several different analytical techniques including CD, 8-anilino-1-naphthalein-sulfonate (ANS) and intrinsic fluorescence spectroscopy, and static light scattering.

The CSD between two protein samples is a pictorial representation of regions of significant differences between two datasets. In general, the x-axis and y-axis of a CSD are arbitrary to provide maximum flexibility in the use of the CSDs with different analytical techniques and formulation conditions. In the examples presented in the paper, however, the x-axis represents spectral wavelength (to reflect changes in the spectra of target proteins) and the y-axis defines a pharmaceutically relevant stress, in this case, temperature. Light scattering is shown in a separate bar on the right for ease of visualization. These axis can, however, be modified to suit the needs of the data. For instance, the x-axis could also represent time to accommodate chromatographic data. Figure 1 provides a simulated dataset of biophysical data for use in explaining the CSD methodology as outlined below.

Construction of the CSD

The first step in the construction of the CSD is to discard statistically insignificant differences between datasets. Start with a dataset X_{srt}, where $s = 1{,}2$ is an index for the two datasets being compared, $r = 1 \ldots n_{\text{runs}}$ is an index for n runs, and $t = 1 \ldots t_{\max}$ is an index for temperature. Plots of simulated datasets are shown in Figure 1a. The circles and crosses in Figure 1a correspond to two different simulated datasets and show a difference between the two datasets that rises with increasing temperature. Each dataset has three duplicate measurements at each temperature and incorporates simulated Gaussian error.

Let $\bar{X}_{st} = \frac{1}{n_{\text{runs}}} \sum_{r=1}^{n_{\text{runs}}} X_{srt}$, the mean over the runs. $\bar{X}_{1t} - \bar{X}_{2t}$ is the difference in means between the datasets at temperature t. Whether or not a difference in means $\bar{X}_{1t} - \bar{X}_{2t}$ is statistically significant can be determined by utilizing a decision rule based on the distribution of $\bar{X}_{1t} - \bar{X}_{2t}$, as follows. Assume that the error in dataset s at temperature t fits a normal distribution and let μ_{st}, σ_{st} be the mean and standard deviation of the underlying distribution at s, t. As n_{runs} approaches infinity, the difference in means $\bar{X}_{1t} - \bar{X}_{2t}$ has the limiting normal distribution[23]

$$N\left(\mu_{1t} - \mu_{2t}, \sqrt{\frac{\sigma_{1t}^2}{n_{\text{runs}}} + \frac{\sigma_{2t}^2}{n_{\text{runs}}}}\right)$$

Let the null hypothesis H_0 be that the two datasets are the same at temperature t, or H_0: $\mu_{1t} = \mu_{2t}$. The alternative hypothesis can be expressed as

 DOI 10.1002/jps