Journal of Pharmaceutical Sciences 105 (2016) 602–612

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Pharmaceutical Biotechnology

Biosimilarity Assessments of Model IgG1-Fc Glycoforms Using a Machine Learning Approach

Jae Hyun Kim [1, 2], Sangeeta B. Joshi [1, 2], Thomas J. Tolbert [1], C. Russell Middaugh [1, 2], David B. Volkin [1, 2], Aaron Smalter Hall [3, *]

[1] Department of Pharmaceutical Chemistry, University of Kansas, Lawrence, Kansas 66047
[2] Macromolecule and Vaccine Stabilization Center, University of Kansas, Lawrence, Kansas 66047
[3] Molecular Graphics and Modeling Lab, Molecular Structures Group, University of Kansas, Lawrence, Kansas 66047

ARTICLE INFO

Article history:
Received 28 August 2015
Revised 7 October 2015
Accepted 9 October 2015
Available online 30 December 2015

Keywords:
biosimilarity
comparability
machine learning
stability
formulation
protein
mAbs
glycoforms
Fc

ABSTRACT

Biosimilarity assessments are performed to decide whether 2 preparations of complex biomolecules can be considered "highly similar." In this work, a machine learning approach is demonstrated as a mathematical tool for such assessments using a variety of analytical data sets. As proof-of-principle, physical stability data sets from 8 samples, 4 well-defined immunoglobulin G1-Fragment crystallizable glycoforms in 2 different formulations, were examined (see More et al., companion article in this issue). The data sets included triplicate measurements from 3 analytical methods across different pH and temperature conditions (2066 data features). Established machine learning techniques were used to determine whether the data sets contain sufficient discriminative power in this application. The support vector machine classifier identified the 8 distinct samples with high accuracy. For these data sets, there exists a minimum threshold in terms of information quality and volume to grant enough discriminative power. Generally, data from multiple analytical techniques, multiple pH conditions, and at least 200 representative features were required to achieve the highest discriminative accuracy. In addition to classification accuracy tests, various methods such as sample space visualization, similarity analysis based on Euclidean distance, and feature ranking by mutual information scores are demonstrated to display their effectiveness as modeling tools for biosimilarity assessments.

Introduction

Owing to the structural complexity and inherent heterogeneity of biopharmaceutical drugs, they are sensitive to changes in bulk or drug product manufacturing processes.[1,2] Such changes in the manufacturing processes often occur during scale-up as part of clinical development and during postapproval life cycle management to meet market demand.[3,4] Moreover, biopharmaceutical drug products also must show reproducible batch-to-batch variation even when they are produced through the same steps. For follow-on biologics, new manufacturers typically develop their own production processes, which inevitably may lead to detectable analytical differences compared with the original bioproducts.[5] A major question is what degree of variation and what kinds of differences are permitted, for the data sets characterizing the 2 groups of biomolecules, for them to be considered as sufficiently similar (i.e., from a regulatory perspective, they are not necessarily identical but "highly similar") in terms of their clinical safety and efficacy.[6]

To address such questions, comparability and biosimilarity studies are performed with biopharmaceutical drugs to assess the effects of process changes (by the same manufacturer) and production of follow-on biologics (by a different manufacturer), respectively.[7-11] These studies first identify the critical quality attributes (CQAs) that are key physicochemical and biological properties of complex biomolecules to establish parameters such as their identify, purity, potency, and stability as ultimately linked to clinical safety and efficacy. To identify CQAs, proper assays must be developed by evaluating a variety of different potential methodologies and instruments. Once identified, CQAs can be compared, through a series of selected analytical tests (e.g., physicochemical, biological, and animal), across batches, between pre- and post-manufacturing changes, or between original and biosimilar

Present address for Aaron Smalter Hall: Center for the Advancement of Data and Research in Economics, Federal Reserve Bank of Kansas City, 1 Memorial Drive, Kansas City, MO 64198.

* *Correspondence to*: Aaron Smalter Hall (Telephone: 816-283-7883; Fax: 816-881-2569).

E-mail address: asmalterhall@gmail.com (A. Smalter Hall).

http://dx.doi.org/10.1016/j.xphs.2015.10.013
0022-3549/

products. The extent of uncertainty in such results ultimately must be decided to determine the need for, and extent of, human clinical trials that will be required to establish clinical safety and efficacy.[3,8-11]

A common method for comparability and biosimilarity assessments is to compare each CQA value directly and statistically, often in a 1-to-1 fashion, or using more sophisticated approaches.[7,12-16] For example, CQA numerical values could be calculated from experimental data (e.g., melting temperature: T_m) or determined directly from raw data (e.g., comparisons of spectra vs. temperature). In 1-to-1 comparisons, statistical tests are performed to see if there exists any statistically significant difference between 2 groups of samples. If the results are different, or not conclusive, for a certain set of CQAs derived from physicochemical and/or biological potency assays, additional studies involving animal models and/or human clinical studies may need to be performed. Analytical evaluations are therefore considered the anchor to any comparability assessment.[3,17] Previously, we reported in our laboratories the use of comparative signature diagrams (CSDs)[18,19] as data visualization tool to more easily assess if a wide variety of data components manifest statistically significant differences. CSDs are effective in identifying regions of raw data readouts (e.g., range of wavelengths and temperature conditions for spectra) by using color as a monitor if there exist statistically significant differences. Such features allow one to learn the characteristics of any differences in 2 sets of data and help decide whether such differences are important for similarity determination. These 1-to-1 comparison schemes, however, lack a way to combine differences found in individual data sets, which may be required to make overall decisions.

As an alternative method to 1-to-1 comparison of each CQA, it may be useful to estimate an overall picture from multiple aspects of all data sets as part of comparison, biosimilarity comparison. For example, recent draft guidance from the U.S. Food and Drug Administration refers to the possible use of fingerprint analysis (i.e., "…it may be useful to compare the quality attributes of the proposed biosimilar product with those of the reference product using a meaningful fingerprint-like analysis algorithm that covers a large number of product attributes and their combinations with high sensitivity using orthogonal methods…").[8] Previous work in our labs have used empirical phase diagrams (EPDs) and radar charts[17,20-22] as data analysis and visualization tools to locate major patterns extracted from multiple experimental data sets using principal component analysis (PCA) or normalized scales of relative changes. Because an EPD can summarize multiple CQAs, simply comparing EPDs allows macroscopic and multilateral comparison between 2 groups of samples. EPDs have been used to qualitatively evaluate the structural integrity and conformational stability of closely related macromolecules, for example, wild type and mutants,[17,23,24] differentially deglycosylated mAbs,[25] and glycosylated IgG-Fc proteins with different site occupancy or amino acid residues at the N297 glycosylation site.[26]

Data visualization tools such as CSDs, EPDs, and radar charts are helpful tools for qualitative and visual comparison of protein stability data sets. As a complementary and more quantitative approach, machine learning methods will be explored here for comparative assessments of protein samples using a variety of different analytical data sets. Machine learning methods are well suited for solving classification and regression problems. Comparative assessments can be framed as a classification problem where the goal is to determine whether 2 groups of samples, given their CQAs, are drawn from the same category. Similarly, comparative assessments could be framed as a regression problem to find the fractional similarity of a sample to a target category. For machine learning techniques, previously determined CQAs are used as features, or CQAs can be computationally determined from all available data by techniques such as feature selection or ranking. Reductions in the number of features (or CQAs) are important not only in improving machine learning accuracy but also in reducing the time and resource costs associated with obtaining them. Once minimal CQAs are determined by extensive testing of the original biopharmaceutical product, data from new batches from process changes or follow-on biosimilar products can be requested to provide appropriate information for comparability and biosimilarity assessments, respectively.

In this article, a number of analytical data sets generated with 4 well-defined IgG1-Fc glycoforms were used as a model system for biosimilarity assessments using machine learning techniques. As described in the companion articles in this special issue of *Journal of Pharmaceutical Sciences*, these IgG1-Fc glycoforms provide a model system for biosimilarity assessments, that can potentially be applied to monoclonal antibodies, in terms of varying levels of purity and potency (see Okbazghi et al.[27]), susceptibility to chemical degradation (see Mozziconacci et al.[28]), and physical stability profile (see More et al.[29]) based on the type and level of glycosylation (at the N297 N-linked glycosylation site in the C_H2 domain of IgGs). Furthermore, 2 different formulations (with NaCl or sucrose) were tested for each Fc glycoform, and the nature of the formulation has been found to alter the physical stability of Fc glycoforms (More et al[29]). The different formulations are tested to simulate follow-on biologics that may be formulated differently. Biophysical stability data sets (across different solution pH values and temperatures) using triplicate measurements from 3 different analytical methods were used in this study (a total of 2066 data features). Established machine learning techniques were used to evaluate whether they have sufficient discriminative power in this application. On successful identification by tailored mathematical models and classifiers, they could be further applied to assess similarity between these model reference and biosimilar samples.

Materials and Methods

Sample Preparation

Four different well-defined IgG1-Fc glycoforms were produced and characterized, as described in the companion articles in this special issue, in terms of structural integrity and potency (see Okbazghi et al.[27] in this issue), susceptibility to chemical degradation (see Mozziconacci et al.[28] in this issue), and physical stability profile (see More et al.[29] in this issue). Briefly, the high mannose IgG1-Fc (HM-Fc) was first expressed and purified from yeast. The Man5 and GlcNAc glycoforms (Man5-Fc and GlcNAc-Fc) were produced by enzymatic digestion from HM-Fc using a bacterial α-1, 2-mannosidase (GH92),[30] and endoglycosidase H, respectively. The N297Q nonglycosylated form (N297Q-Fc) was made by using site-directed mutagenesis to remove the N-linked glycosylation site (see Okbazghi et al.[27] in this issue). The 4 glycoforms were prepared at a concentration of 0.2 mg/mL and were dialyzed into 2 different formulations: 20 mM of citrate–phosphate buffer at pH 4.0-7.5 (0.5 pH unit increments) containing either (1) NaCl with a total ionic strength of 0.15 or (2) 10% sucrose (w/v; see More et al.[29] in this issue).

Analytical Methods

For the 8 different IgG-Fc glycoform samples described previously, 3 different analytical methods were used to monitor protein structural integrity and physical stability as a function of solution pH and temperature. Intrinsic (tryptophan [Trp]) fluorescence, extrinsic fluorescence (with SYPRO orange dye), and turbidity (by monitoring optical density at 350 nm) were measured in triplicate