Data in Brief 4 (2015) 583–586

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data set for mass spectrometric analysis of recombinant human serum albumin from various expression systems

Daryl G.S. Smith, Grant E. Frahm, Anita Kane, Barry Lorbetskie, Michel Girard, Michael J.W. Johnston, Terry D. Cyr*

Centre for Biologics Evaluation, Biologics and Genetic Therapies Directoriate, Health Canada, Ottawa, ON, Canada K1A0K9

ARTICLE INFO

Article history:
Received 11 June 2015
Received in revised form
20 July 2015
Accepted 20 July 2015
Available online 1 August 2015

Keywords:
Albumin
Recombinant proteins
Glycation
Mass spectrometry
Oryza sativa

ABSTRACT

Human serum albumin (HSA) is a versatile and important protein for the pharmaceutical industry (Fanali et al., Mol. Aspects Med. 33 (3) (2012) 209–290). Due to the potential transmission of pathogens from plasma sourced albumin, numerous expression systems have been developed to produce recombinant HSA (rHSA) (Chen et al., Biochim. Biophys. Acta (BBA)—Gen. Subj. 1830(12) (2013) 5515–5525; Kobayashi, Biologicals 34(1) (2006) 55–59). Based on our previous study showing increased glycation of rHSA expressed in Asian rice (Frahm et al., J. Phys. Chem. B 116(15) (2012) 4661–4670), both supplier-to-supplier and lot-to-lot variability of rHSAs from a number of expression systems were evaluated using reversed phase liquid chromatography linked with MS and MS/MS analyses. The data are associated with the research article 'Determination of Supplier-to-Supplier and Lot-to-Lot Variability in Glycation of Recombinant Human Serum Albumin Expressed in *Oryza sativa*' where further analysis of rHSA samples with additional biophysical methods can be found (Frahm et al., PLoS ONE 10(9) (2014) e109893). We determined that all rHSA samples expressed in rice showed elevated levels of arginine and lysine hexose glycation compared to rHSA expressed in yeast, suggesting that the extensive glycation of the recombinant

* Correspondence to: Correspondence to: Biologics and Genetic Therapies Directorate Health Canada, AL 2201E, 251 Sir Frederick Banting Driveway, Ottawa, Ontario, K1A 0K9 Canada. Tel.: +613 957 1068.
E-mail address: terry.cyr@hc-sc.gc.ca (T.D. Cyr).

http://dx.doi.org/10.1016/j.dib.2015.07.024

proteins is a by-product of either the expression system or purification process and not a random occurrence.

Specifications table

Subject area	Biology
More specific subject area	Recombinant proteins
Type of data	Raw Data, Peak Lists, Search Results and Exported Search Results
How data was acquired	Liquid chromatography–mass spectrometry (LC–MS) analysis—waters nanoAcquity UPLC and waters synapt HDMS system operating in data dependant acquisition (DDA) mode.
Data format	Waters raw data (.raw), Mascot generic file peak lists (.mgf), Mascot search results (.dat) and exported search results (.mzid)
Experimental factors	Samples were reduced and then alkylated with iodoacetamide followed by digestion with trypsin and chymotrypsin
Experimental features	Commercially available recombinant human serum albumin samples were digested and analyzed by LC–MS/MS
Data source location	Ottawa, Ontario, Canada
Data accessibility	Data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PXD001248 and DOI 10.6019/PXD001248).

1. Value of the data [describe in 3–5 bulleted points why this data is of value to the scientific community]

- Represents a robust method for profiling variation in glycation of commercial products
- Demonstrates the utility of multi-enzyme approach for extensive protein sequence coverage allowing detailed analysis/comparison of non-complex proteomic samples
- Provides a model data set for qualitative and quantitative proteomics studies
- Covers the entire data generation/analysis process, from raw data, to processed peak lists, to search results, to exported result files.

2. Experimental design, materials and methods

As described in 'Determination of Supplier-to-Supplier and Lot-to-Lot Variability in Glycation of Recombinant Human Serum Albumin Expressed in *Oryza sativa*' [1,2,3,5].

3. Materials

Chemicals, essentially FA-free pHSA (A3872 Lot090M7001V, $\geq$99% purity), recombinant human serum albumin expressed in Saccharomyces cerevisiae (ScrHSA, Lot SLBD2407, $\geq$99% purity, Albucult), *O. sativa* [OsrHSA, Lot SLBC7527V (OsrHSA-sig-C), Lot SLBG7405V (OsrHSA-sig-G), Lot SLBH9636V (OsrHSA-sig-H) and Lot SLBJ1196V (OsrHSA-sig-J) 100% purity, Cellastim] and Pichia pastoris (PprHSA, Lot 080M1580V, $\geq$99% purity, Albagen) were sourced from Sigma-Aldrich (St. Louis, MO, USA). OsrHSA was also obtained from eEnzyme LLC (Gaithersburg, MD, USA) (OsrHSA-phy) (Lot 20130110, $>$99% purity, Phyto-HSA), ScienCell Research Laboratories (Carlsbad, CA, USA) (OsrHSA-sci) (Lot BJABAA42, $\geq$99% purity, Oryzogen) and amsbio LLC (Cambridge, MA, USA) (OsrHSA-ams) (Lot 20101008, $>$95% purity, ecoHSA). Recombumin (Lot PDP100106) was donated by Novozymes Biopharma (Cambridge, MA, USA). Amicon Ultra 0.5 ml 3000 Da molecular weight cut-off (MWCO) centrifugal filter devices were purchased from Millipore (Billerica, MA, USA). Trypsin and chymotrypsin were Promega sequencing grade purchased from Fisher Scientific Canada (Ottawa, ON, Canada). Vivacon 500 10 kDa molecular weight cut-off filters were from Sartorius Stadium Biotech North America (Bohemia, NY, USA).