

ULTRAVIOLET MONITORING OF CHROMATOGRAPHY PERFORMANCE BY ORTHOGONAL PARTIAL LEAST SQUARES

FIELD OF THE INVENTION

[001] The present invention pertains to chromatography, and in particular, to a method for ultraviolet (UV) monitoring of chromatography performance, such as by use of an orthogonal partial least squares (OPLS) model.

BACKGROUND

[002] Chromatography is a separations technique used to purify proteins, including antibodies, in the biopharmaceutical industry. Chromatography uses a column packed with small particles (resin) which interacts with the product (desired protein) to separate it from impurities. Chromatograms are time based graphical records of a chromatographic separation to see how parameters such as UV absorbance, conductivity, and pH change over time during the chromatography run. The packed bed of resin in a chromatography column can deteriorate over time which can impact the efficiency of the chromatography separation and affect product quality.

[003] It is beneficial to know when a chromatography column has begun to deteriorate so it can be repacked before the column deteriorates to the point where product quality can be impacted. Signs of column deterioration often appear in the UV absorbance trace in the chromatogram; however, features of deterioration can be difficult to visually identify. Chromatography-based protein analysis could benefit tremendously from improved methods of detecting and monitoring column deterioration, which can subsequently lead to improved protein purification, including improved antibody purification.

BRIEF SUMMARY OF THE INVENTION

[004] In one aspect, the present invention provides a method of monitoring column chromatography performance, comprising: acquiring one or more chromatogram ultraviolet (UV) traces generated by a chromatography system during sample purification and/or separation; and analyzing the one or more acquired chromatogram UV traces with an orthogonal partial least squares (OPLS) model, thereby allowing detection of column deterioration prior to column failure and quantitative analysis of UV signal in the one or more chromatogram UV traces.

[005] In some embodiments, the method further comprises creating an OPLS model.

[006] In some embodiments creating the OPLS model comprises: selecting a process or unit operation of the OPLS model; collecting raw data for a UV trace of one or more column cycles/lots/runs for the selected process or unit operation; normalizing and aligning the collected

raw data; optionally generating artificially created curves from normalized raw data; classifying and formatting data for importation into a multivariate tool; importing classified and formatted data into the multivariate tool to generate a training set; and generating the OPLS model.

[007] In some embodiments, the method further comprises optimizing the generated OPLS model, validating and testing the optimized OPLS model.

[008] In some embodiments, the process or unit operation is a chromatography unit operation for a single molecule.

[009] In some embodiments, the process or unit operation is a protein affinity chromatography step for the single molecule.

[0010] In some embodiments, collecting raw data for a UV trace of one or more column cycles/lots/runs for the selected process or unit operation comprises collecting UV absorbance values at corresponding column volumes for the one or more cycles/lots/runs.

[0011] In some embodiments, normalizing and aligning the collected raw data comprises normalizing and aligning UV absorbance values and aligning column volumes.

[0012] In some embodiments, normalizing UV absorbance values comprises removing variation in magnitude differences in UV raw signal.

[0013] In some embodiments, optionally generating artificially created curves from normalized raw data comprises generating artificially created curves from normalized raw data when one or more unacceptable UV chromatogram traces are not available.

[0014] In some embodiments, the method further comprises providing a sample to the chromatography system prior to acquiring the one or more chromatogram ultraviolet (UV) traces generated by a chromatography system during sample purification and/or separation.

[0015] In some embodiments, the chromatography system is a liquid chromatography system.

[0016] In some embodiments, the sample comprises a protein.

[0017] In some embodiments, the protein is an antibody, a fusion protein, recombinant protein, or a combination thereof.

[0018] In some embodiments, the antibody is a monoclonal antibody.

[0019] In some embodiments, the monoclonal antibody is of isotype IgG1, IgG2, IgG3, IgG4, or mixed isotype.

[0020] Also disclosed is a method of creating an orthogonal partial least square (OPLS) model for UV monitoring of a chromatography column performance, comprising: selecting a process or unit operation of the OPLS model; collecting raw data for a ultraviolet (UV) chromatogram trace of one or more column cycles/lots/runs for the selected process or unit operation; normalizing and aligning the collected raw data; optionally generating artificially created curves from normalized raw data; classifying and formatting data for importation into a multivariate tool; importing classified and

formatted data into the multivariate tool to generate a training set; and generating the OPLS model.

[0021] In some embodiments, the method further comprises optimizing the generated OPLS model, validating and testing the optimized OPLS model.

[0022] In some embodiments, the process or unit operation is a chromatography unit operation for a single molecule.

[0023] In some embodiments, the process or unit operation is a protein affinity chromatography step for the single molecule.

[0024] In some embodiments, collecting raw data for a UV trace of one or more column cycles/lots/runs for the selected process or unit operation comprises collecting UV absorbance values at corresponding column volumes for the one or more cycles/lots/runs.

[0025] In some embodiments, normalizing and aligning the collected raw data comprises normalizing and aligning UV absorbance values and aligning column volumes.

[0026] In some embodiments, normalizing UV absorbance values comprises removing variation in magnitude differences in UV raw signal.

[0027] In some embodiments, optionally generating artificially created curves from normalized raw data comprises generating artificially created curves from normalized raw data when one or more unacceptable UV chromatogram traces are not available.

[0028] In some embodiments, the method further comprises providing a sample to the chromatography system prior to acquiring the one or more chromatogram ultraviolet (UV) traces generated by a chromatography system during sample purification and/or separation.

[0029] In some embodiments, the chromatography system is a liquid chromatography system.

[0030] In some embodiments, the sample comprises a protein.

[0031] In some embodiments, the protein is an antibody, a fusion protein, recombinant protein, or a combination thereof.

[0032] In some embodiments, the antibody is a monoclonal antibody.

[0033] In some embodiments, the monoclonal antibody is of isotype IgG1, IgG2, IgG3, IgG4, or mixed isotype.

[0034] In embodiments, a non-transitory computer-readable storage medium with an executable program stored thereon for monitoring column chromatography performance, wherein the program instructs a microprocessor to perform the steps of: acquiring one or more chromatogram ultraviolet (UV) traces generated by a chromatography system during sample purification and/or separation; and analyzing the one or more acquired chromatogram UV traces with an orthogonal partial least squares (OPLS) model, thereby allowing detection of column deterioration prior to column failure and quantitative analysis of UV signal in the one or more chromatogram UV traces.

[0035] In some embodiments, the non-transitory computer-readable storage medium further

comprises instructions for creating the OPLS model.

[0036] In some embodiments creating the OPLS model comprises: selecting a process or unit operation of the OPLS model; collecting raw data for a UV trace of one or more column cycles/lots/runs for the selected process or unit operation; normalizing and aligning the collected raw data; optionally generating artificially created curves from normalized raw data; classifying and formatting data for importation into a multivariate tool; importing classified and formatted data into the multivariate tool to generate a training set; and generating the OPLS model.

[0037] In some embodiments, the method further comprises optimizing the generated OPLS model, validating and testing the optimized OPLS model.

[0038] In some embodiments, the process or unit operation is a chromatography unit operation for a single molecule.

[0039] In some embodiments, the process or unit operation is a protein affinity chromatography step for the single molecule.

[0040] In some embodiments, collecting raw data for a UV trace of one or more column cycles/lots/runs for the selected process or unit operation comprises collecting UV absorbance values at corresponding column volumes for the one or more cycles/lots/runs.

[0041] In some embodiments, normalizing and aligning the collected raw data comprises normalizing and aligning UV absorbance values and aligning column volumes.

[0042] In some embodiments, normalizing UV absorbance values comprises removing variation in magnitude differences in UV raw signal.

[0043] In some embodiments, optionally generating artificially created curves from normalized raw data comprises generating artificially created curves from normalized raw data when one or more unacceptable UV chromatogram traces are not available.

[0044] In some embodiments, the method further comprises providing a sample to the chromatography system prior to acquiring the one or more chromatogram ultraviolet (UV) traces generated by a chromatography system during sample purification and/or separation.

[0045] In some embodiments, the chromatography system is a liquid chromatography system.

[0046] In some embodiments, the sample comprises a protein.

[0047] In some embodiments, the protein is an antibody, a fusion protein, recombinant protein, or a combination thereof.

[0048] In some embodiments, the antibody is a monoclonal antibody.

[0049] In some embodiments, the monoclonal antibody is of isotype IgG1, IgG2, IgG3, IgG4, or mixed isotype.

[0050] In various embodiments, any of the features or components of embodiments discussed above or herein may be combined, and such combinations are encompassed within the scope of

the present disclosure. Any specific value discussed above or herein may be combined with another related value discussed above or herein to recite a range with the values representing the upper and lower ends of the range, and such ranges and all values falling within such ranges are encompassed within the scope of the present disclosure. Each of the values discussed above or herein may be expressed with a variation of 1%, 5%, 10% or 20%. Other embodiments will become apparent from a review of the ensuing detailed description.

DESCRIPTION OF THE FIGURES

[0051] Figure 1A shows an exemplary process workflow in accordance with embodiments disclosed herein.

[0052] Figure 1B shows a geometric interpretation of OPLS.

[0053] Figure 1C shows a schematic illustrating OPLS predictions.

[0054] Figure 1D shows a flow-chart illustrating an exemplary method of creating a UV OPLS model.

[0055] Figure 2 shows an exemplary bind-elute chromatogram with the elution block highlighted.

[0056] Figure 3 shows an exemplary flow-through chromatogram with UV liftoff highlighted.

[0057] Figure 4 shows exemplary exported raw data generated in accordance with embodiments disclosed herein.

[0058] Figure 5-1 and Figure 5-2 show exemplary normalization of data in accordance with embodiments disclosed herein.

[0059] Figures 6-1, 6-2, 6-3, 7, 8, 9-1, 9-2, 10-1, and 10-2 show creating artificially generated curves from normalized data in accordance with embodiments disclosed herein.

[0060] Figure 11-1 and Figure 11-2 show exemplary classifying and formatting data for Simca import in accordance with embodiments disclosed herein.

[0061] Figures 12A, 12B, 13A, 13B-1, 13B-2, 14A, 14B, 15A, 15B-1, 15B-2, and 16 show exemplary importing data into a multivariate tool in accordance with embodiments disclosed herein.

[0062] Figures 17, 18, 19, and 20 show exemplary generating an OPLS model in accordance with embodiments disclosed herein.

[0063] Figures 21, 22A, 22B, 23A, 23B, 24A, 24B, 25-1, and 25-2 show exemplary optimization of an OPLS model in accordance with embodiments disclosed herein.

[0064] Figures 26A, 26B, 27A, 27B, 28, 29-1, 29-2, 30, 31-1, and 31-2 show exemplary application of the created OPLS model to classify new chromatograms as acceptable or unacceptable in accordance with embodiments disclosed herein.

[0065] Figures 32-1, 32-2, 33, 34, 35, 36, 37, 38A, 38B, 39A, 39B, and 40 show results generated by use of OPLS models in accordance with embodiments disclosed herein.

[0066] Figure 41 shows a schematic representation of an exemplary computing environment for performing aspects of the methods disclosed herein.

DETAILED DESCRIPTION OF THE INVENTION

[0067] Before the present invention is described, it is to be understood that this invention is not limited to particular methods and experimental conditions described, as such methods and conditions may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims. Any embodiments or features of embodiments can be combined with one another, and such combinations are expressly encompassed within the scope of the present invention.

[0068] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. As used herein, the term "about," when used in reference to a particular recited numerical value, means that the value may vary from the recited value by no more than 1%. For example, as used herein, the expression "about 100" includes 99 and 101 and all values in between (e.g., 99.1, 99.2, 99.3, 99.4, etc.).

[0069] Various operations may be described as multiple discrete operations in turn, in a manner that may be helpful in understanding embodiments; however, the order of description should not be construed to imply that these operations are order dependent.

[0070] The terms "coupled" and "connected," along with their derivatives, may be used. These terms are not intended as synonyms for each other. Rather, aspects, "connected" may be used to indicate that two or more elements are in direct physical or electrical contact with each other. "Coupled" may mean that two or more elements are in direct physical or electrical contact. However, "coupled" may also mean that two or more elements are not in direct contact with each other, but still cooperate or interact with each other.

[0071] As used herein, the terms "include," "includes," and "including," are meant to be non-limiting and are understood to mean "comprise," "comprises," and "comprising," respectively.

[0072] Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. All patents, applications and non-patent publications mentioned in this specification are incorporated herein by reference in their entireties.

Abbreviations Used Herein

[0073]	CHO: Chinese Hamster Ovary
[0074]	CV: Column Volumes
[0075]	DDA: Data-Dependent Acquisition
[0076]	EIC: Extracted Ion Chromatograph
[0077]	HC: Heavy Chain
[0078]	HIC: Hydrophobic Interaction Chromatography
[0079]	HILIC: Hydrophilic Interaction Liquid Chromatography
[0080]	HMW: High Molecular Weight
[0081]	IgG: Immunoglobulin G
[0082]	IPA: Isopropanol
[0083]	LC: Light Chain
[0084]	LMW: Low Molecular Weight
[0085]	mAb: Monoclonal Antibody
[0086]	MW: Molecular Weight
[0087]	OPLS: Orthogonal Partial Least Square
[0088]	PK: Pharmacokinetics
[0089]	RMSE: Root Mean Square Error
[0090]	RP-LC: Reversed Phase Liquid Chromatography
[0091]	SME: Subject Matter Expert
[0092]	SIM: Selected Ion Monitoring
[0093]	UV: Ultraviolet

Definitions

[0094] As used herein, the term "protein" includes any amino acid polymer having covalently linked amide bonds. Proteins comprise one or more amino acid polymer chains, generally known in the art as "polypeptides." "Polypeptide" refers to a polymer composed of amino acid residues, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof linked via peptide bonds, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof. "Synthetic peptides or polypeptides" refers to a non-naturally occurring peptide or polypeptide. Synthetic peptides or polypeptides can be synthesized, for example, using an automated polypeptide synthesizer. Various solid phase peptide synthesis methods are known to those of skill in the art. A protein may contain one or multiple polypeptides to form a single functioning biomolecule. A protein can include any of bio-therapeutic proteins, recombinant proteins used in research or therapy, trap proteins and other chimeric receptor Fc-fusion proteins, chimeric proteins, antibodies, monoclonal antibodies, polyclonal antibodies, human antibodies, and

bispecific antibodies. In another exemplary aspect, a protein can include antibody fragments, nanobodies, recombinant antibody chimeras, cytokines, chemokines, peptide hormones, and the like. Proteins may be produced using recombinant cell-based production systems, such as the insect baculovirus system, yeast systems (e.g., *Pichia* sp.), mammalian systems (e.g., CHO cells and CHO derivatives like CHO-K1 cells). For a recent review discussing biotherapeutic proteins and their production, see Ghaderi et al., "Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation," (*Biotechnol. Genet. Eng. Rev.* (2012) 147-75). In some embodiments, proteins comprise modifications, adducts, and other covalently linked moieties. Those modifications, adducts and moieties include for example avidin, streptavidin, biotin, glycans (e.g., N-acetylgalactosamine, galactose, neuraminic acid, N-acetylglucosamine, fucose, mannose, and other monosaccharides), PEG, polyhistidine, FLAGtag, maltose binding protein (MBP), chitin binding protein (CBP), glutathione-S-transferase (GST) myc-epitope, fluorescent labels and other dyes, and the like. Proteins can be classified on the basis of compositions and solubility and can thus include simple proteins, such as, globular proteins and fibrous proteins; conjugated proteins, such as, nucleoproteins, glycoproteins, mucoproteins, chromoproteins, phosphoproteins, metalloproteins, and lipoproteins; and derived proteins, such as, primary derived proteins and secondary derived proteins.

[0095] Variant protein" or "protein variant", or "variant" as used herein can include a protein that differs from a target protein by virtue of at least one amino acid modification. Protein variant may refer to the protein itself, a composition comprising the protein, or the amino sequence that encodes it. Preferably, the protein variant has at least one amino acid modification compared to the parent protein, e.g. from about one to about ten amino acid modifications, and preferably from about one to about five amino acid modifications compared to the parent. The protein variant sequence herein will preferably possess at least about 80% homology with a parent protein sequence, and most preferably at least about 90% homology, more preferably at least about 95% homology. In some exemplary embodiments, the protein can be an antibody, a bispecific antibody, a multispecific antibody, antibody fragment, monoclonal antibody, or combinations thereof.

[0096] The term "antibody", as used herein, is intended to refer to immunoglobulin molecules comprised of four polypeptide chains, two heavy (H) chains and two light (L) chains inter-connected by disulfide bonds (*i.e.*, "full antibody molecules"), as well as multimers thereof (e.g. IgM) or antigen-binding fragments thereof. Each heavy chain is comprised of a heavy chain variable region ("HCVR" or "V_H") and a heavy chain constant region (comprised of domains C_{H1}, C_{H2} and C_{H3}). In various embodiments, the heavy chain may be an IgG isotype. In some cases, the heavy chain is selected from IgG1, IgG2, IgG3 or IgG4. In some embodiments, the heavy chain is of isotype IgG1 or IgG4, optionally including a chimeric hinge region of isotype IgG1/IgG2 or IgG4/IgG2. Each light

chain is comprised of a light chain variable region ("LCVR or "V_L") and a light chain constant region (C_L). The V_H and V_L regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDR), interspersed with regions that are more conserved, termed framework regions (FR). Each V_H and V_L is composed of three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4. The term "antibody" includes reference to both glycosylated and non-glycosylated immunoglobulins of any isotype or subclass. The term "antibody" includes antibody molecules prepared, expressed, created or isolated by recombinant means, such as antibodies isolated from a host cell transfected to express the antibody. For a review on antibody structure, see Lefranc et al., *IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains*, 27(1) Dev. Comp. Immunol. 55-77 (2003); and M. Potter, *Structural correlates of immunoglobulin diversity*, 2(1) Surv. Immunol. Res. 27-42 (1983).

[0097] The term antibody also encompasses "bispecific antibody", which includes a heterotetrameric immunoglobulin that can bind to more than one different epitope. One half of the bispecific antibody, which includes a single heavy chain and a single light chain and six CDRs, binds to one antigen or epitope, and the other half of the antibody binds to a different antigen or epitope. In some cases, the bispecific antibody can bind the same antigen, but at different epitopes or non-overlapping epitopes. In some cases, both halves of the bispecific antibody have identical light chains while retaining dual specificity. Bispecific antibodies are described generally in U.S. Patent App. Pub. No. 2010/0331527 (Dec. 30, 2010).

[0098] The term "antigen-binding portion" of an antibody (or "antibody fragment"), refers to one or more fragments of an antibody that retain the ability to specifically bind to an antigen. Examples of binding fragments encompassed within the term "antigen-binding portion" of an antibody include (i) a Fab fragment, a monovalent fragment consisting of the VL, VH, CL and CH1 domains; (ii) a F(ab')₂ fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a Fd fragment consisting of the VH and CH1 domains; (iv) a Fv fragment consisting of the VL and VH domains of a single arm of an antibody, (v) a dAb fragment (Ward et al. (1989) Nature 241:544-546), which consists of a VH domain, (vi) an isolated CDR, and (vii) an scFv, which consists of the two domains of the Fv fragment, VL and VH, joined by a synthetic linker to form a single protein chain in which the VL and VH regions pair to form monovalent molecules. Other forms of single chain antibodies, such as diabodies are also encompassed under the term "antibody" (see e.g., Holliger et al. (1993) 90 PNAS U.S.A. 6444-6448; and Poljak et al. (1994) 2 Structure 1121-1123).

[0099] Moreover, antibodies and antigen-binding fragments thereof can be obtained using standard recombinant DNA techniques commonly known in the art (see Sambrook et al., 1989).

Methods for generating human antibodies in transgenic mice are also known in the art. For example, using VELOCIMMUNE® technology (see, for example, US 6,596,541, Regeneron Pharmaceuticals, VELOCIMMUNE®) or any other known method for generating monoclonal antibodies, high affinity chimeric antibodies to a desired antigen are initially isolated having a human variable region and a mouse constant region. The VELOCIMMUNE® technology involves generation of a transgenic mouse having a genome comprising human heavy and light chain variable regions operably linked to endogenous mouse constant region loci such that the mouse produces an antibody comprising a human variable region and a mouse constant region in response to antigenic stimulation. The DNA encoding the variable regions of the heavy and light chains of the antibody are isolated and operably linked to DNA encoding the human heavy and light chain constant regions. The DNA is then expressed in a cell capable of expressing the fully human antibody

[00100] The term "human antibody", is intended to include antibodies having variable and constant regions derived from human germline immunoglobulin sequences. The human mAbs of the invention may include amino acid residues not encoded by human germline immunoglobulin sequences (e.g., mutations introduced by random or site-specific mutagenesis *in vitro* or by somatic mutation *in vivo*), for example in the CDRs and in particular CDR3. However, the term "human antibody", as used herein, is not intended to include mAbs in which CDR sequences derived from the germline of another mammalian species (e.g., mouse), have been grafted onto human FR sequences. The term includes antibodies recombinantly produced in a non-human mammal, or in cells of a non-human mammal. The term is not intended to include antibodies isolated from or generated in a human subject.

[00101] As used herein, the term "impurity" can include any undesirable protein present in the biopharmaceutical product. Impurity can include process and product-related impurities. The impurity can further be of known structure, partially characterized, or unidentified. Process-related impurities can be derived from the manufacturing process and can include the three major categories: cell substrate-derived, cell culture-derived and downstream derived. Cell substrate-derived impurities include, but are not limited to, proteins derived from the host organism and nucleic acid (host cell genomic, vector, or total DNA). Cell culture-derived impurities include, but are not limited to, inducers, antibiotics, serum, and other media components. Downstream-derived impurities include, but are not limited to, enzymes, chemical and biochemical processing reagents (e.g., cyanogen bromide, guanidine, oxidizing and reducing agents), inorganic salts (e.g., heavy metals, arsenic, nonmetallic ion), solvents, carriers, ligands (e.g., monoclonal antibodies), and other leachables. Product-related impurities (e.g., precursors, certain degradation products) can be molecular variants arising during manufacture and/or storage that do not have properties

comparable to those of the desired product with respect to activity, efficacy, and safety. Such variants may need considerable effort in isolation and characterization in order to identify the type of modification(s). Product-related impurities can include truncated forms, modified forms, and aggregates. Truncated forms are formed by hydrolytic enzymes or chemicals which catalyze the cleavage of peptide bonds. Modified forms include, but are not limited to, deamidated, isomerized, mismatched S-S linked, oxidized, or altered conjugated forms (e.g., glycosylation, phosphorylation). Modified forms can also include any post-translational modification form. Aggregates include dimers and higher multiples of the desired product (Q6B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products, ICH August 1999, U.S. Dept. of Health and Humans Services).

[00102] The term "low molecular weight (LMW) protein drug impurity" includes but is not limited to precursors, degradation products, truncated species, proteolytic fragments including Fab fragments, Fc or heavy chain fragments, ligand or receptor fragments, H2L (2 heavy chains and 1 light chain), H2 (2 heavy chains), HL (1 heavy chain and 1 light chain), HC (1 heavy chain), and LC (1 light chain) species. A LMW protein drug impurity can be any variant which is an incomplete version of the protein product, such as one or more components of a multimeric protein. Protein drug impurity, drug impurity or product impurity are terms that may be used interchangeably throughout the specification. LMW drug or product impurities are generally considered molecular variants with properties such as activity, efficacy, and safety that may be different from those of the desired drug product.

[00103] Degradation of protein product is problematic during production of the protein drug product in cell culture systems. For example, proteolysis of a protein product may occur due to release of proteases in cell culture medium. Medium additives, such as soluble iron sources added to inhibit metalloproteases, or serine and cysteine proteases inhibitors, have been implemented in cell culture to prevent degradation (Clincke, M.-F., et al, BMC Proc. 2011, 5, P115). C-terminal fragments may be cleaved during production due to carboxyl peptidases in the cell culture (Dick, LW et al, Biotechnol Bioeng 2008; 100:1132-43).

[00104] The term "high molecular weight (HMW) protein drug impurity" includes but is not limited to mAb trimers and mAb dimers. HMW species can be divided into two groups: 1) monomer with extra light chains (H2L3 and H2L4 species) and 2) monomer plus Fab fragments complexes. In addition, after treatment with IdeS enzymatic digestion, different dimerized fragments (Fab2-Fab2, Fc-Fc and Fab2-Fc) are formed.

[00105] The term as used herein, "glycopeptide/glycoprotein" is a modified peptide/protein, during or after their synthesis, with covalently bonded carbohydrates or glycan. In certain embodiments, a

glycopeptide is obtained from a monoclonal antibody, for example, from a protease digest of a monoclonal antibody.

[00106] The term as used herein, “glycan” is a compound comprising one or more of sugar units which commonly include glucose (Glc), galactose (Gal), mannose (Man), fucose (Fuc), N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc) and N-acetylneuraminic acid (NeuNAc) (Frank Kjeldsen, et al. Anal. Chem. 2003, 75, 2355-2361). The glycan moiety in glycoprotein, such as a monoclonal antibody, is an important character to identify its function or cellular location. For example, a specific monoclonal antibody is modified with specific glycan moiety.

[00107] The term “sample,” as used herein, includes at least an analyte molecule, e.g., glycopeptide, such as obtained from a monoclonal antibody, that is subjected to manipulation in accordance with the methods of the invention, including, for example, separating, analyzing, extracting, concentrating or profiling.

[00108] The terms “analysis” or “analyzing,” as used herein, are used interchangeably and refer to any of the various methods of separating, detecting, isolating, purifying, solubilizing, detecting and/or characterizing molecules of interest. Examples include, but are not limited to, chromatography, solid phase extraction, solid phase micro extraction, electrophoresis, mass spectrometry, e.g., liquid chromatography, e.g., high performance, e.g., reverse phase, normal phase, or size exclusion, ion-pair liquid chromatography, liquid-liquid extraction, e.g., accelerated fluid extraction, supercritical fluid extraction, microwave-assisted extraction, membrane extraction, soxhlet extraction, precipitation, clarification, electrochemical detection, staining, elemental analysis, Edmund degradation, nuclear magnetic resonance, infrared analysis, flow injection analysis, capillary electrochromatography, ultraviolet detection, and combinations thereof.

[00109] The term “profiling,” as used herein, refers to any of various methods of analysis which are used in combination to provide the content, composition, or characteristic ratio of compounds, such as proteins.

[00110] As used herein, the term “digestion” refers to hydrolysis of one or more peptide bonds of a protein. There are several approaches to carrying out digestion of a protein in a sample using an appropriate hydrolyzing agent, for example, enzymatic digestion or non-enzymatic digestion. As used herein, the term “hydrolyzing agent” refers to any one or combination of a large number of different agents that can perform digestion of a protein. Non-limiting examples of hydrolyzing agents that can carry out enzymatic digestion include trypsin, endoproteinase Arg-C, endoproteinase Asp-N, endoproteinase Glu-C, outer membrane protease T (OmpT), immunoglobulin-degrading enzyme of *Streptococcus pyogenes* (IdeS), chymotrypsin, pepsin, thermolysin, papain, pronase, and protease from *Aspergillus Saitoi*. Non-limiting examples of

hydrolyzing agents that can carry out non-enzymatic digestion include the use of high temperature, microwave, ultrasound, high pressure, infrared, solvents (non-limiting examples are ethanol and acetonitrile), immobilized enzyme digestion (IMER), magnetic particle immobilized enzymes, and on-chip immobilized enzymes. For a recent review discussing the available techniques for protein digestion see Switazar et al., "Protein Digestion: An Overview of the Available Techniques and Recent Developments" (J. Proteome Research 2013, 12, 1067-1077). One or a combination of hydrolyzing agents can cleave peptide bonds in a protein or polypeptide, in a sequence-specific manner, generating a predictable collection of shorter peptides.

[00111] Several approaches are available that can be used to digest a protein. One of the widely accepted methods for digestion of proteins in a sample involves the use of proteases. Many proteases are available and each of them has their own characteristics in terms of specificity, efficiency, and optimum digestion conditions. Proteases refer to both endopeptidases and exopeptidases, as classified based on the ability of the protease to cleave at non-terminal or terminal amino acids within a peptide. Alternatively, proteases also refer to the six distinct classes - aspartic, glutamic, and metalloproteases, cysteine, serine, and threonine proteases, as classified on the mechanism of catalysis. The terms "protease" and "peptidase" are used interchangeably to refer to enzymes which hydrolyze peptide bonds. Proteases can also be classified into specific and non-specific proteases. As used herein, the term "specific protease" refers to a protease with an ability to cleave the peptide substrate at a specific amino acid side chain of a peptide. As used herein, the term "non-specific protease" refers to a protease with a reduced ability to cleave the peptide substrate at a specific amino acid side chain of a peptide. A cleavage preference may be determined based on the ratio of the number of a particular amino acid as the site of cleavage to the total number of cleaved amino acids in the protein sequences.

[00112] The protein can optionally be prepared before characterizing. In some exemplary embodiments, the protein preparation includes a step of protein digestion. In some specific exemplary embodiments, the protein preparation includes a step of protein digestion, wherein the protein digestion can be carried out using trypsin.

[00113] In some exemplary embodiments, the protein preparation can include a step for denaturing the protein, reducing the protein, buffering the protein, and/or desalting the sample, before the step of protein digestion. These steps can be accomplished in any suitable manner as desired.

[00114] As used herein, the term "chromatography" refers to a process technique for separating the components, or solutes, of a mixture on the basis of the relative amounts of each solute distributed between a moving fluid stream, called the mobile phase, and a contiguous stationary phase. The mobile phase may be either a liquid or a gas, while the stationary phase is either a

solid or a liquid.

[00115] As used herein, the term “liquid chromatography” refers to a process in which a chemical mixture carried by a liquid can be separated into components as a result of differential distribution of the chemical entities as they flow around or over a stationary liquid or solid phase. Non-limiting examples of liquid chromatography include reverse phase liquid chromatography, ion-exchange chromatography, size exclusion chromatography, affinity chromatography, and hydrophobic chromatography.

[00116] As used herein, the term “multivariate tool” refers to a statistical tool that uses multiple variables to forecast outcomes. A multivariate tool can allow data to be explored, analyzed and/or interpreted. The tool can facilitate data diving by revealing trends and clusters, analyze process variations, identify parameters and/or predict final product quality. In some examples, a multivariate tool is one that is commercially available, such as SIMCA (umetrics, Umea, Sweden).

[00117] As used herein, the term “protein sequence coverage” refers to the percentage of the protein sequence covered by identified peptides. The percent coverage can be calculated by dividing the number of amino acids in all found peptides by the total number of amino acids in the entire protein sequence.

[00118] As used herein, the term “database” refers to bioinformatic tools which provide the possibility of searching the uninterpreted MS-MS spectra against all possible sequences in the database(s). Non-limiting examples of such tools are Mascot (www.matrixscience.com), Spectrum Mill (www.chem.agilent.com), PLGS (www.waters.com), PEAKS (www.bioinformaticssolutions.com), Proteinpilot (download.appliedbiosystems.com/proteinpilot), Phenyx (<http://www.phenyx-ms.com>), Sorcerer (www.sagenresearch.com), OMSSA (www.pubchem.ncbi.nlm.nih.gov/omssa/), XITandem (www.thegpm.org/TANDEM/), Protein Prospector (<http://prospector.ucsf.edu/prospector/mshome.htm>), Byonic (www.proteinmetrics.com/products/byonic) or Sequest (fields.scripps.edu/sequest).

General Description

[00119] From the foregoing, it will be appreciated that a need exists for improved methods and systems to improve protein purification, including antibody purification. The disclosed invention meets that need. Disclosed herein are methods utilizing OPLS modelling including examples of acceptable (no deterioration) and unacceptable (deterioration) chromatogram UV traces, such as UV, infrared (IR) or Ramen traces, to provide early detection of column failures and quantitative analysis of the UV signal in the chromatogram. In some embodiments, the methods utilize UV OPLS modeling. The disclosed methods combine process knowledge gained by subject matter experts (SMEs) with equations and procedures to create chromatograms to create comprehensive data sets ranging from robust to acceptable to incremental failures to catastrophic (see Figure 1A).

The disclosed methods utilize OPLS to address many challenges such as automation and comprehensive data assessment which was not previously possible. The disclosed methods achieve process monitoring through use of a multivariate tool (e.g., SIMCA) for predicting column failures. The disclosed OPLS models can be built without historical column failures and can be automated (such as by utilizing Python scripting) eliminating the manual process to perform overlays and analysis. Procedurally generated operation failures by use of mathematical equations and protocols overcome the weakness of limited historical data sets not containing failures. The disclosed methods are able to provide quantitative measurement of the chromatogram and column performance and allow for dynamic identification of variation. Thus, the disclosed methods ensure process consistency and robustness as well as consistent performance of equipment while eliminating disadvantages associated with current methods.

[00120] In embodiments, the disclosed methods utilize OPLS. OPLS regression analyzes relationships between two blocks of data by regression extensions of Principle Component Analysis (PCA). For example, a dataspace for X and Y matrix is constructed. A first component fits a line in the X and Y space so that the correlation between the X and Y projection is maximized. A second component is orthogonal to the first component in the X space. The first component is predictive and maximizes covariance between X and Y while the second component and onwards are orthogonal and represent structured variation in X that is not related to Y (Figure 1B). OPLS-DA is a model where qualitative categorical variables (e.g., good/satisfactory or bad/unsatisfactory) are assigned quantitative values so that an OPLS model can be constructed. Figure 1C shows a schematic illustrating OPLS predictions.

[00121] In embodiments, a method of monitoring column chromatography performance, comprises acquiring one or more chromatogram ultraviolet (UV) traces generated by a chromatography system during sample purification and/or separation; and analyzing the one or more chromatogram UV traces with an OPLS model, thereby allowing detection of column deterioration prior to column failure and quantitative analysis of UV signal in the one or more chromatogram UV traces.

[00122] In embodiments, the method of monitoring column chromatography performance, includes creating a UV OPLS model. Figure 1D provides a flow chart illustrating an exemplary protocol for creating a UV OPLS model. As illustrated in Figure 1D, creating a UV OPLS model, includes step 102, identifying a process or unit operation the UV OPLS model is being built. In some embodiments, this is a particular chromatography unit operation for a certain product/molecule, for example, a protein affinity, ion exchange, hydrophobic interaction, or size exclusion chromatography step in either bind and elute or flow through design for a specific molecule. In embodiments, at least a portion of the UV signal in the chromatogram is fed into the model. Some portions of the chromatogram are not helpful for determining column integrity. For example, at least

one block or section of the chromatogram is used. In some embodiments, if the unit operation column modality is bind and elute, then the elution block is used for analysis (e.g., where the bound product elutes off the column). Figure 2 provides an example of a bind-elute chromatogram with elution block highlighted. In some embodiments, if the unit operation column modality is flow through, then the point where collection starts during the liftoff of the UV signal will be analyzed. Figure 3 provides an example of a flow-through chromatogram with UV liftoff highlighted.

[00123] The method illustrated in Figure 1D for creating a UV OPLS model further includes step 200, gathering raw data for a UV trace. In some embodiments, gathering raw data for the UV trace includes gathering raw data of multiple cycles, lots, and/or runs for the selected unit operation, for example UV absorbance values at corresponding volumes. For example, raw data for UV and logbook from the chromatography software (e.g., Unicorn) is exported, such as into an electronic storage file (e.g., Microsoft Excel file). The logbook is a tracker that identifies what events occurred at what volume during the chromatogram and identifies where to pull the block of data to input into the model (e.g., the elution block). The process can then be repeated for each run/cycle of the chromatography step. In embodiments, the run/cycle of raw data is stored, such as in a separate file (e.g., separate Microsoft Excel files) as shown in Figure 4.

[00124] In embodiments, the exemplary method includes step 106, normalizing and aligning raw data, such as normalizing UV values and aligning volumes. For example, the data is normalized to remove any variation in magnitude differences (from UV meter functionality) in the raw signal. In embodiments, the raw UV and logbook data imported, data is normalized and formatted for import into a multivariate tool (e.g., SIMCA), such as by use of a macro. In embodiments, the steps performed by a macro can include:

1. Macro first prompts to select what files the macro will run on (e.g., what runs will be fed in to build the model);
2. Next, the macro finds the section of the chromatogram that will be analyzed by the model (e.g., the Elution block);
3. The macro finds the maximum and minimum UV values and normalizes the UV values from -1 to 1;
4. The macro normalizes the volume in terms of Column Volumes (CVs);
5. The macro pulls the normalized UV data at a specified CV interval so that all the UV data is aligned (e.g., pull UV signal value every 0.02 CVs in the Elution block);
6. The macro pastes the normalized data into a table that is formatted for Simca import; and
7. Macro repeats steps 2-6 if more than one run is selected.

In embodiments, column volumes (CVs) are optimized as a volume interval to maximize data resolution and minimize required computational resources. In one embodiment, 0.02 CVs are used. Figures 5-1 and 5-2 provide an example of normalizing the data in accordance with embodiments disclosed herein.

[00125] In embodiments, the method includes evaluating the occurrence of column failures in the data, prior to generating curves. For example, if there are empirical examples of failed chromatograms, then artificially generated chromatograms may not be necessary. If there are a sufficient number of examples (e.g., 7 or greater, such as 8, 9, 10, 11, 12, 13, 14, 15 or greater) of failed chromatograms, then these chromatogram serve as the unacceptable data set and artificial curve generation is not needed.

[00126] In embodiments, the method optionally includes step 108, creating artificially generated curves from normalized data to provide the OPLS model examples of undesirable chromatograms. For example, step 108 is performed if there are no examples of unacceptable chromatograms. In embodiments, if the chromatogram is for a Bind-Elute Column, an elution peak generator tool, such as an Elution peak generator Microsoft Excel tool is used. In embodiments, if the chromatogram is for a flow-through column, a flow through curve generator excel tool is used.

[00127] In some embodiments, creating artificially generated curves includes determining which runs to use to generate the curves. For example, generally about 10-15 runs are a sufficient number of chromatograms to evaluate. If the data set is large (e.g., greater than 100 runs), runs are selected so that variability in the process is captured (e.g., only acceptable chromatograms are selected). In embodiments, creating artificially generated curves further includes copying data into a generation tab under F- Q and copying data for one of the runs into column E (e.g., chart displays the chromatogram in a first color, such as blue, in the figure next to the data whereas the mathematically generated chromatogram is displayed in a second color, such as red. In some embodiments, the method of creating artificially generated curves includes aligning a curve generated from the model to the real run data. For example, alignment can continue to be performed until root mean square error (RMSE) value in cell X1 stops decreasing which indicates that the model has been fitted to the chromatogram. Equation parameter data is copied for the curve and such values will be used to procedurally generate the new chromatograms. One or more of the prior actions may be repeated for selected run data. For example, runs with variability are selected as it is not required to fit multiple runs if they are relatively mirror images.

[00128] In embodiments, creating artificially generated curves for normalized data includes generating chromatograms that are representative of deterioration. For example, this may be performed by use of a computer program, such as Microsoft Excel which copies the mean and standard deviation values for each equation parameter into the table in Columns X and Y. In

embodiments, for the elution peak generator four terms are utilized in the mathematical model: (1) Tm1: determines where the peak liftoff occurs (e.g., decreasing tm1 can shift the peak start to the left and increasing it can shift it to the right); (2) S1: determines how steep the peak liftoff is (e.g., decreasing s1 causes the peak liftoff to be sharper while increasing it makes it broader); (3) Tm2: determines where the peak end occurs (e.g., decreasing tm2 shifts peak end to the left and increasing tm2 shifts peak end to the right); and S2: determines how steep the peak dropdown is (e.g., decreasing s2 makes the peak end sharper while increasing s2 makes the peak end broader). In embodiments, additional terms for fitting the peak maximum may be utilized, such as if the top of the elution peak is not flat or if the UV sensor utilized requires additional input. In embodiments, the equation parameters are initially set to the mean parameter values from the generation tab and standard deviation (SD) values initially set to 0. Changing these values will change the corresponding equation parameter by the number of standard deviations entered into the cell (e.g., entering 3 into the SD cell next to s1 will increase the value of s1 by 3 SDs). In embodiments, each parameter is accessed and the SD value is either increased or decreased to create elution peaks indicative of deterioration such as broadening (increasing tm2), tailing (increasing s2); fronting and/or bimodal peaks are other common deterioration indicators which can be modeled as well. The action is performed allowing minor, moderate, and/or major variation examples to be created. In some embodiments, between 10-15 undesirable chromatograms are created using the aforementioned process, for example one or more minor deviation, one or more moderate and one or more major variation examples are created. In embodiments, the undesirable chromatograms include a greater number of minor deviation examples as compared to the moderate and major deviation examples. In embodiments, the undesirable chromatograms include a greater number of moderate deviation examples as compared to the major deviation examples. For example, in one embodiment, the undesirable chromatograms created include 7 minor deviation, 5 moderate deviation, and 3 as major deviation examples. Figures 6-1, 6-2, 6-3, 7, 8, 9-1, 9-2, 10-1, and 10-2 illustrate creating artificially generated curves from normalized data in accordance with embodiments disclosed herein.

[00129] In embodiments, the method further includes step 110, classifying data and formatting data for import into a multivariate tool, such as SIMCA, to determine what data is acceptable and unacceptable to train model. Figure 11-1 and Figure 11-2 show exemplary classifying and formatting data for SIMCA import in accordance with embodiments disclosed herein. For example, in some embodiments classifying and formatting data for importation into a multivariate tool includes (1) accessing the file utilized to normalize the data; (2) copying data for the artificially generated good and bad curves after the data already imported; (3) classifying each run/cycle into two groups – an acceptable or “good” group which contains chromatograms that do not show signs

of column deterioration and an unacceptable or “bad” group which contains chromatograms that have indicators of column deterioration, such as tailing, peak broadening or other indicators known to those of ordinary skill in the art to indicate column deterioration; and (4) provide additional categories to classify runs, such as column packing number of resin cycle number, facilitating the ability to organize and read the data with a multivariate tool, such as SIMCA (for example, the generated groups can be colored and sorted in SIMCA).

[00130] In embodiments, the method includes importing data into a multivariate platform, such as SIMCA (step 112) and generating an OPLS model (step 114). Figures 12A, 12B, 13A, 13B-1, 13B-2, 14A, 14B, 15A, 15B-1, 15B-2, and 16 illustrate a protocol for importing data into SIMCA in accordance with embodiments disclosed herein. Figures 17, 18, 19 and 20 illustrate a protocol for generating an OPLS model in accordance with embodiments disclosed herein.

[00131] After generating an OPLS model, the method includes optimizing the OPLS model (step 116), validating and testing (step 118) and model application (step 120). In some embodiments, model application can include classifying new data as acceptable or unacceptable. In embodiments, the resulting optimized OPLS model is further augmented, such as by increasing, decreasing and/or modifying the number and/or content of the mathematically generated chromatogram inputs. For example, if the OPLS model is not satisfactory after optimizing OPLS settings, then it is further augmented by increasing, decreasing and/or modifying the number and/or content of mathematically generated chromatogram inputs. In embodiments, the OPLS model optimization are iterative with the mathematically generated chromatogram failure examples described herein. Figures 21, 22A, 22B, 23A, 23B, 24A, 24B, 25-1, and 25-2 show exemplary optimization of an OPLS model in accordance with embodiments disclosed herein. In embodiments, optimizing the OPLS includes determining if there is another set of model making parameters that fits a better model to the data. The Q^2 value represents the predictive power of the model (e.g., the higher Q^2 , the higher the predictive power of the model). In embodiments, a Q^2 value of 0.7 or greater is considered acceptable. In some embodiments, additional models are created by comparing the Q^2 values to determine which model is better able to predict new data. Further, additional data transformations can be performed with a multivariate tool, such as SIMCA, to analyze a derived or transformed version of the imported data. In embodiments, one or more of the following parameters are adjusted to create additional OPLS models: (1) Derivative level (e.g., 0 or non-derived data, 1 or first derivative, 2 or second derivative, etc.); (2) Smoothing level (e.g., when taking derivatives or for additional data smoothing); and/or (3) Excluding observations (e.g., UV meter malfunction during the chromatogram).

[00132] In embodiments, an optimized UV OPLS model is analyzed to determine the predictive ability of the model. For example, a multivariate tool, such as SIMCA, is used to test the predictive ability of the optimized UV OPLS model. In embodiments, permutations are determined which indicate the statistical significance of the R^2 and Q^2 values by finding reference distributions of the R^2 and Q^2 values from permutation testing of the Y variable. If the model is a satisfactory model, the reference R^2 and Q^2 values will have lower values than the values of the model. In embodiments, a CV scores plot is generated which shows the cross validated component to the regular scores plot. A CV Scores plot which closely matches the regular scores plot indicates the model is satisfactory.

[00133] Figures 26A, 26B, 27A, 27B, 28, 29-1, 29-2, 30, 31-1, and 31-2 show exemplary application of the created OPLS model to classify new chromatograms as acceptable or unacceptable in accordance with embodiments disclosed herein.

[00134] In embodiments, the method further includes providing a sample to the chromatography system prior to acquiring the one or more chromatogram UV traces generated by a chromatography system during sample purification and/or separation.

[00135] Although the description herein describes in detail the use of the disclosed OPLS model with UV spectrophotometry it is contemplated that the disclosed model and methods may be used with any form of spectrophotometry that monitors the outlet of the column for the eluate, including, but not limited to RAMAN or IR.

[00136] In some embodiments, the chromatography system capable of sample separation to purify and/or separate sample components comprises a liquid chromatography system. In some embodiments, the system is a chromatography system is a hydrophobic chromatography system, reverse phase liquid chromatography system, ion-exchange chromatography system, size exclusion chromatography system, affinity chromatography system, or hydrophilic-interaction chromatography system.

[00137] In some embodiments, the chromatography column temperature can be maintained at a constant temperature throughout the chromatography run, e.g., using a commercial column heater. In some embodiments, the column is maintained at a temperature between about 18°C to about 70°C, e.g., about 30°C to about 60°C, about 40°C to about 50°C, e.g., at about 20°C, about 25°C, about 30°C, about 35°C, about 40°C, about 45°C, about 50°C, about 55°C, about 60°C, about 65°C, or about 70°C. In some embodiments, the column temperature is about 40°C. In some embodiments, the run time can be between about 15 to about 240 minutes, e.g., about 20 to about 70 min, about 30 to about 60 min, about 40 to about 90 min, about 50 min to about 100 min, about 60 to about 120 min, about 50 to about 80 min.

[00138] In some embodiments, the mobile phase is an aqueous mobile phase. A representative aqueous mobile phase contains 208 mM sodium acetate and 10 mM ammonium bicarbonate. The UV traces are typically recorded at 215 and 280 nm.

[00139] In some exemplary embodiments, the mobile phase used to elute the protein can be a mobile phase that can be compatible with a mass spectrometer.

[00140] In some exemplary embodiments, the mobile phase used in the liquid chromatography device can include water, acetonitrile, trifluoroacetic acid, formic acid, or combination thereof.

[00141] In some exemplary embodiments, the mobile phase for manufacturing operations can have a flow rate varying within operation and operation to operation from 60 L/hr to 1800 L/hr.

[00142] In some embodiments, the sample is a protein or cell culture medium including a protein, exemplary proteins including, but not limited to, an antibody, a fusion protein, recombinant protein, or a combination thereof.

[00143] In some embodiments, the antibody is a bispecific antibody, antibody fragment or a multispecific antibody.

[00144] In some exemplary embodiments, the antibody is a monoclonal antibody, such as, but not limited to, a monoclonal antibody of isotype IgG1, IgG2, IgG3, IgG4, or mixed isotype.

[00145] In some exemplary embodiments, the protein is be a therapeutic protein.

[00146] In some exemplary embodiments, the protein can be an immunoglobulin protein.

[00147] In one exemplary embodiment, the protein can be a protein variant.

[00148] In one exemplary embodiment, the protein can be a post-translationally modified protein.

[00149] In one exemplary embodiment, the post-translationally modified protein can be a formed by cleavage, N-terminal extensions, protein degradation, acylation of the N-terminus, biotinylation, amidation of the C-terminal, oxidation, glycosylation, iodination, covalent attachment of prosthetic groups, acetylation, alkylation, methylation, adenylation, ADP-ribosylation, covalent cross links within, or between, polypeptide chains, sulfonation, prenylation, Vitamin C dependent modifications, Vitamin K dependent modification, glutamylation, glycylation, glycosylation, deglycosylation, isoprenylation, lipoylation, phosphopantetheinylation, phosphorylation, sulfation, citrullination, deamidation, formation of disulfide bridges, proteolytic cleavage, ISGylation, SUMOylation or ubiquitination (covalent linkage to the protein ubiquitin).

[00150] In one exemplary embodiment, the post-translationally modified protein can be formed on oxidation of a protein.

[00151] In embodiments, the disclosed methods are used to monitor column deterioration due to a change in the column packing status, an accumulation of contaminant components, channeling through the column, microparticle blockage, desorption from the solid phase, or a combination thereof. In embodiments, the disclosed methods detect column deterioration prior to column failure.

In embodiments the disclosed methods detect imminent column deterioration prior to a sign of column deterioration is manifested, such as increased column pressure, decreased theoretical plates, shortened retention time, poor peak shape, and/or decreased resolution.

[00152] It is contemplated that the methods described herein can be performed by software, hardware, or both, of a computing environment, such as one or more computing devices. For example, computing devices include server computers, desktop computers, laptop computers, notebook computers, handheld devices, netbooks, tablet devices, mobile devices, and other types of computing devices.

[00153] FIG. 41 illustrates an exemplary computing environment 200 for implementation of various aspects of the methods disclosed herein, including creating and/or utilizing an OPLS for monitoring methods of monitoring chromatography performance. The computing environment 200 is not intended to suggest any limitation as to scope of use or functionality, as the technologies may be implemented in diverse general-purpose or special-purpose computing environments. For example, the disclosed technology may be implemented using a computing device comprising a processing unit, memory, and storage, storing computer-executable instructions implementing methods disclosed herein. The disclosed technology may also be implemented with other computer system configurations, including hand held devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, a collection of client/server systems, and the like. The disclosed technology may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices

[00154] With reference to FIG. 41, the computing environment 200 includes at least one processing unit 210 coupled to memory 220. In FIG. 41, this basic configuration 230 is included within a dashed line. The processing unit 210 executes computer-executable instructions and may be a real or a virtual processor. In a multi-processing system, multiple processing units execute computer-executable instructions to increase processing power. The memory 220 may be volatile memory (e.g., registers, cache, RAM), non-volatile memory (e.g., ROM, EEPROM, flash memory, etc.), or some combination of the two. The memory 220 can store software 280 implementing any of the technologies described herein.

[00155] A computing environment may have additional features. For example, the computing environment 200 includes storage 240, one or more input devices 250, one or more output devices 260, and one or more communication connections 270. An interconnection mechanism (not shown) such as a bus, controller, or network interconnects the components of the computing environment 200. Typically, operating system software (not shown) provides an operating

environment for other software executing in the computing environment 200, and coordinates activities of the components of the computing environment 200.

[00156] The storage 240 may be removable or non-removable, and includes magnetic disks, magnetic tapes or cassettes, CD-ROMs, CD-RWs, DVDs, or any other computer-readable media which can be used to store information and which can be accessed within the computing environment 200. The storage 240 can store software 280 containing instructions for any of the technologies described herein.

[00157] The input device(s) 250 may be a touch input device such as a keyboard, mouse, pen, or trackball, a voice input device, a scanning device, or another device that provides input to the computing environment 200. For audio, the input device(s) 250 may be a sound card or similar device that accepts audio input in analog or digital form, or a CD-ROM reader that provides audio samples to the computing environment. The output device(s) 260 may be a display, printer, speaker, CD-writer, or another device that provides output from the computing environment 200.

[00158] The communication connection(s) 270 enable communication over a communication mechanism to another computing entity. The communication mechanism conveys information such as computer-executable instructions, audio/video or other information, or other data. By way of example, and not limitation, communication mechanisms include wired or wireless techniques implemented with an electrical, optical, RF, infrared, acoustic, or other carrier.

[00159] The techniques herein can be described in the general context of computer-executable instructions, such as those included in program modules, being executed in a computing environment on a target real or virtual processor. Generally, program modules include routines, programs, libraries, objects, classes, components, data structures, etc., that perform particular tasks or implement particular abstract data types. The functionality of the program modules may be combined or split between program modules as desired in various embodiments. Computer-executable instructions for program modules may be executed within a local or distributed computing environment.

[00160] Any of the disclosed methods can be implemented as computer-executable instructions or a computer program product stored on one or more computer-readable storage media (e.g., non-transitory computer-readable media, such as one or more optical media discs such as DVD or CD, volatile memory components (such as DRAM or SRAM, or non-volatile memory components such as hard drives) and executed on a computer (e.g., any commercially available computer, including smart phones or other mobile devices that include computing hardware). Computer-readable media does not include propagated signals. Any of the computer-executable instructions for implementing the disclosed methods as well as any data created and used during implementation

of the disclosed embodiments can be stored on one or more computer-readable media (e.g., non-transitory computer-readable media).

[00161] The computer-usable or computer-readable medium may be, for example but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific examples (a non-exhaustive list) of the computer-readable medium would include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a transmission media such as those supporting the Internet or an intranet, or a magnetic storage device. Note that the computer-usable or computer-readable medium can even be paper or another suitable medium upon which the program is printed, as the program can be electronically captured, via, for instance, optical scanning of the paper or other medium, then compiled, interpreted, or otherwise processed in a suitable manner, if necessary, and then stored in a computer memory. In the context of this document, a computer-usable or computer-readable medium may be any medium that can contain, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device. The computer-usable medium may include a propagated data signal with the computer-usable program code embodied therewith, either in baseband or as part of a carrier wave. The computer usable program code may be transmitted using any appropriate medium, including but not limited to wireless, wireline, optical fiber cable, RF, etc.

[00162] The computer-executable instructions can be part of, for example, a dedicated software application or a software application that is accessed or downloaded via a web browser or other software application (such as a remote computing application). Such software can be executed, for example, on a single local computer (e.g., any suitable commercially available computer) or in a network environment (e.g., via the internet, a wide-area network, a local-area network, a client-server network (such as a cloud computing network), or other such network using one or more network computers.

[00163] For clarity, only certain selected aspects of the software-based implementations are described. Other details that are well known in the art are omitted. For example, it should be understood that the disclosed technology is not limited to any specific computer language or program. For instance, the disclosed technology can be implemented by software written in C++, Java, Perl, JavaScript, Adobe Flash, Python or any other suitable programming language. Likewise, the disclosed technology is not limited to any particular computer or type of hardware.

[00164] Furthermore, example embodiments may be implemented by hardware, software, firmware, middleware, microcode, hardware description languages, or any combination thereof. When implemented in software, firmware, middleware or microcode, the program code or code segments to perform the necessary tasks may be stored in a machine or computer readable medium. A code segment may represent a procedure, a function, a subprogram, a program, a routine, a subroutine, a module, program code, a software package, a class, or any combination of instructions, data structures, program statements, and the like.

[00165] In embodiments, any of the software-based embodiments (including, for example, computer-executable instructions for causing a computer to perform any of the disclosed methods) can be uploaded, downloaded or remotely accessed through a suitable communication means. Such suitable communication means include, for example, the internet, the World Wide Web, an intranet, cable (including fiber optic cable), magnetic communications, electromagnetic communications (including RF, microwave, and infrared communications), electronic communications, or other such communication means.

[00166] In various embodiments, an article of manufacture may be employed to implement one or more methods as disclosed herein. The article of manufacture may include a computer-readable non-transitory storage medium and a storage medium. The storage medium may include programming instructions configured to cause an apparatus to practice some or all aspects of a disclosed method using a computing device, in accordance with embodiments of the present disclosure. The storage medium may represent a broad range of persistent storage medium known in the art, including but not limited to flash memory, optical disks or magnetic disks. The programming instructions, in particular, may enable an apparatus, in response to their execution by the apparatus, to perform various operations described herein. For example, the storage medium may include programming instructions configured to cause an apparatus to practice some or all aspects of a disclosed method, such as a method of monitoring column chromatography performance, including creating an OPLS model, in accordance with embodiments of the present disclosure.

[00167] Although various example methods, apparatus, systems, and articles of manufacture have been described herein, the scope of coverage of the present disclosure is not limited thereto. On the contrary, the present disclosure covers all methods, apparatus, and articles of manufacture fairly falling within the scope of the appended claims either literally or under the doctrine of equivalents. For example, although the above discloses example systems including, among other components, software or firmware executed on hardware, it should be noted that such systems are merely illustrative and should not be considered as limiting. In particular, it is contemplated that any or all of the disclosed hardware, software, and/or firmware components can be embodied

exclusively in hardware, exclusively in software, exclusively in firmware or in some combination of hardware, software, and/or firmware.

EXAMPLES

[00168] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the methods of the invention, and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperature, *etc.*) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees Centigrade, room temperature is about 25°C, and pressure is at or near atmospheric.

Example 1: Case study: UV OPLS Model

[00169] A UV OPLS model was generated using methods disclosed herein. As illustrated in Figure 32-1 and Figure 32-2, the generated OPLS model was able to differentiate between good/satisfactory lots (circles on left side) and bad/unsatisfactory lots (circles on right side). The generated model was able to identify column deterioration before it was readily apparent from historical overlay or through transition analysis. Figure 33 is a plot generated using an OPLS model on Column Pack 1122600028. The model was able to detect column deterioration before it was apparent through visual observation (column data was not in training set). Figure 34 is a plot generated from a Column Pack 1122600018. This column did not deteriorate as the model shows that column stayed within the good/satisfactory range (left side). Column remained robust entire lifetime demonstrates that model is working.

Example 2: Expanding Application of OPLS Models – Procedurally generated chromatograms (e.g. chromatograms artificially generated, not from actual experimental data)

[00170] This example shows UV OPLS models can be created that are satisfactory without using lots where column deterioration was observed. For flow-through columns, the initial UV liftoff at the start of collection is expected to broaden as the column bed degrades. The liftoff can be modeled by the equation for logistic growth. Half of the data was selected and augmented with procedurally generated curves allowing the column to be calibrated. Figure 35 illustrates the curves generated, blue was real data and green was information fed in based upon prediction.

[00171] Figure 36 shows the results of an OPLS model constructed using sample set of real acceptable historical data and procedurally generated curves representative of deterioration (e.g.,

chromatogram curves generated not from actual data) in which to make model more conservative, upper limit was set as the highest value for the acceptable lots used. Higher values were indicative of more column deterioration (not shown procedurally generated failures)

[00172] Figure 37 shows the model was used to predict the performance of column pack 1194400001 in which the model was able to predict deterioration approximately 10 lots before column failure. This example indicates that OPLS modeling can be applied to UV signaling of chromatography to detect subtle, but important changes in column performance. Further, procedurally generated data can be used instead of a comprehensive training set with the disclosed models and still allow column deterioration to be modeled.

Example 3: OPLS Model Prediction

[00173] This example provides exemplary OPLS Model Prediction Data generated by the disclosed methods. Figures 38A and 38B illustrate the model prediction for column pack 1122600011. TA did not clearly show development of deterioration. Bifurcation was present in the next lot after the first lot to exceed $x=4$. Operation preceded PS repack recommendations. An OPLS model for hydrophobic interaction chromatography (HIC) using SME generated chromatograms was constructed. A procedure was developed to create curves based on the standard deviation of the values for the sample set of acceptable curves at 2 points. The results are shown in Figures 39A and 39B (Set $x = -2.46$ as upper limit (highest value for acceptable lots)). Figure 40 illustrates the generated OPLS model was able to predict deterioration for column packs 20 & 21 without using data from historical failures. This example confirms that successful OPLS models can be made utilizing procedurally generated curves and a comprehensive training set is not required for the disclosed methods to be used to monitor column deterioration.

[00174] Overall, the disclosed methods provide a robust, sensitive method for monitoring column deterioration, which can be used to improve protein process development, including antibody process development.

[00175] The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and the accompanying figures. Such modifications are intended to fall within the scope of the appended claims.