

MIT Open Access Articles

The application of HPLC/MS analysis with a multi-enzyme digest strategy to characterize different interferon product variants produced from Pichia pastoris

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

As Published: <https://doi.org/10.1007/s00726-019-02772-9>

Publisher: Springer Vienna

Persistent URL: <https://hdl.handle.net/1721.1/131809>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



The application of HPLC/MS analysis with a multi-enzyme digest strategy to characterize different interferon product variants produced from *Pichia pastoris*

Cite this article as: Yu Wang, Di Liu, Laura E. Crowell, Kerry R. Love, Shiaw-lin Wu and William S. Hancock, The application of HPLC/MS analysis with a multi-enzyme digest strategy to characterize different interferon product variants produced from *Pichia pastoris*, Amino Acids <https://doi.org/10.1007/s00726-019-02772-9>

This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

Author accepted manuscript

The application of HPLC/MS analysis with a multi enzyme digest strategy to characterize different interferon product variants produced from *Pichia Pastoris*

Yu (Annie) Wang ¹, Di Liu ², Laura E. Crowell ², Kerry R. Love ², Shiao-lin Wu ³, William S. Hancock ^{1*}

¹ Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115, United States

² Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, 500 Main Street, Cambridge, Massachusetts, 02142, United States

³ BioAnalytix Inc., 790 Memorial Drive, Cambridge, Massachusetts 02139, United States

* Corresponding Author

wi.hancock@northeastern.edu

Abstract:

Interferons are signaling proteins that belong to the large class of cytokines and human interferons which are classified based on the type of receptor interactions: type I, II and III. IFN α 2b belongs to the type I interferon class with a major therapeutic application for the treatment of hepatitis B and C infections. A recombinant form of IFN α 2b expressed in *E.coli*, known as IntronA, has been approved by US Food and Drug Administration (FDA). IFN γ , also known as type II interferon, plays a significant role in the inhibition of viral replication. Actimmune® is a US Food and Drug Administration (FDA) approved version of IFN γ for the indication of reducing infections associated with Chronic Granulomatous Disease and Severe Malignant Osteopetrosis. In this study we have

applied advanced analytical methods for the characterization of IFN α 2b and IFN γ produced from *Pichia pastoris*. The multi-enzyme digestion approach has been developed to allow measurement of 100% sequence coverage and detailed analysis of post-translational variants and degradation products. In this manner we identified the following variants in IFN α 2b: N-terminal residual leader sequence, an amino acid substitution, oxidation of methionine residues and two sites of high-mannose N-glycosylation. In the *Pichia* IFN γ produced material our approach detected variants resulting from glycosylation, C-terminal proteolysis, oxidation of methionine residues and deamidation.

In this manner the analytical program was able to support rapid process development as well as identify product variants and degradation products in the resulting product.

Introduction:

Human interferon α 2b is a subtype of the human alpha interferon family and its potential therapeutic applications have been investigated extensively.^{1,2} Recombinant IFN α 2b is a 165-amino acid single chain polypeptide with a molecular weight 19.2 KDa. Licensed products include Intron A, first marketed by Schering Corp (Merck) in 1986,³ using *E. coli* expression as well as Shanferon, sold by Shantha in India. and manufactured in *Pichia pastoris*. Intron A is widely used for the treatment of myelomas, lymphomas, solid tumors and chronic hepatitis B and C infections. Interferon gamma (IFN γ) is the single member of the type II class of interferons and it is produced by lymphocytes activated by specific antigens. In addition to antiviral activity, IFN γ has important immunoregulatory functions.⁴ Interferon gamma-1b, marketed as Actimmune®, is a 143-amino acid protein and is produced in *E. coli* with an additional N-terminal

methionine residue. Unlike human IFN gamma, Actimmune is not glycosylated as a result of using *E. coli* as the host system.

Nowadays, there is a growing interest in the development of biosimilars which can provide patients affordable drugs and also bring financial benefits to the biopharmaceutical industry due to increased patient access. In the case of interferon biosimilars there are commercial opportunities as the patents for both Intron A (IFN $\alpha 2b$) and Actimmune (IFN γ) have expired. In fact, development of interferon biosimilars are now underway in many biopharmaceutical companies around the world. For example, Biosidus, Pharmaclon and Zydus are all developing biosimilars of IFN $\alpha 2b$.⁵ In a development and regulatory pathway, obtaining analytical data is the primary step to prove the similarity between the reference material and biosimilars. Due to the higher complexity of biosimilars than small molecules, new analytical methods need to be developed to support the comparability of the biosimilar product to the reference standard.

In the biopharmaceutical industry, there has been a growing interest in using *Pichia Pastoris* as a host system due to favorable factors such as high growth rates which allows reduced fermentation times, the use of defined instead of complex media and secretion of low levels of host cell proteins. The manufacture of therapeutics in yeast has been validated by product approvals by both the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA). In engineered strains yeast is capable for production of human-like post-translational modifications.⁶ All of these factors act to lower the barrier to effective development of lower price therapeutics. In this program we used codon-modified *Pichia pastoris* cloned into a pPICZ α -family factor

that was fused with a truncated form of the *Saccharomyces cerevisiae* α -factor secretion signal, under control of the methanol-inducible AOX1 promoter. The *Saccharomyces cerevisiae* α -factor preprosequence is one of the most widely used secretory signal in *Pichia pastoris*, but one concern that can arise is the generation of N-terminal variant(s) due to the incomplete removal of the signal peptide. It has been reported that during the biosynthetic process this sequence is cleaved into three peptides (one very long) upon secretion.⁷ The endogenous enzymes responsible for these cleavage events are Kex2 and Ste13 and the action of these enzymes generates heterogeneity in the N-terminal amino acid sequence of the secreted product due to the incomplete removal of leader sequence.⁸ For example, the most common variant observed contains an additional EAEA tetrapeptide at the N-terminus.⁹ However, the α factor vector can be designed without the EAEA repeats to minimize the amount of N-terminal heterogeneity.¹⁰ A further limitation with the use of *Pichia Pastoris* is that N-linked glycosylation is restricted to the high-mannose type, which is immunogenic¹¹ and results in poor pharmacokinetic parameters due to rapid clearance via binding to high mannose receptors.¹²

Interferon expression in *Pichia Pastoris* has previously been attempted however, with little protein characterization or product quality information.¹³ This paper presents the use of advanced analytical methods for the in-depth characterization of interferon products to produce in *Pichia Pastoris*. We used the multi-enzyme digestion strategy coupled with liquid chromatography - mass spectrometry analysis^{14,15} for the identification of product variants of interferons produced in the fermentation studies,

including residual leader sequence, an amino acid substitution, backbone cleavages, glycosylation as well as degradation reactions.

Methods and materials:

Reagents:

Ammonium bicarbonate, dithiothreitol and iodacetamide was from Sigma Aldrich. Trypsin, and GluC was purchased from Promega. Guanidine hydrochloride, formic acid optima LC/MS grade, water LC/MS grade, acetonitrile (ACN), Methanol, Sulphosalicylic acid and trichloroacetic acid, hydrochloric acid was purchased from Fisher Scientific. NuPAGE Novex 4-12% Bis-Tris gel, 1.0 mm, 10 well, NuPAGE MES SDS running buffer (20X), NuPAGE LDS Sample Buffer (4X), PageRuler Prestained 10-180 KDA Protein Ladder, Invitrogen SimpleBlue SafeStain, Novex pH 3-10 IEF Protein Gels, 1.0 mm, 12 well, Novex pH 3-10 Buffer Kit (includes Novex IEF Anode Buffer 50X, Novex IEF Cathode Buffer pH 3-10 10X and Novex IEF Sample Buffer pH 3-10 2X), Colloidal Blue Staining Kit and premium grade TCEP-HCl were purchased from Thermo Fisher Scientific. IEF standards was from Bio Rad. IFN γ standard material was purchased from Abcam (Cambridge, MA).

Equipment:

XCell SureLock Mini-Cell electrophoresis apparatus was from Thermo Fisher Scientific. A benchtop centrifuge was purchased from Corning. Ultimate 3000 nano LC pump (Dionex, Mountain View, CA). 10 μ m ID emitter (New Objective, Woburn, MA), LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA).

SDS-PAGE of interferon α 2b:

500 mL 1X NuPAGE SDS running buffer was prepared by mixing 25 mL MES running buffer with 475 mL deionized water. 10 μ l (1 μ g/ μ l) IFN α 2b standard with 3.3 μ l LDS sample (4x) were mixed and 10 μ l (1 μ g/ μ l) *Pichia* IFN with 3.3 μ l LDS sample buffer were mixed. Gel apparatus was set and 5 μ l protein ladder was loaded in lane 1, one blank lane was left between each sample and IFN standard and *Pichia* IFN were loaded. The voltage was set at 160 V and run for 50 min. Gel apparatus was disconnected and the gel was removed from the two plates, stained with SimpleBlue SafeStain for 30min and destained in deionized water overnight.

Isoelectric focusing (IEF) of interferon α 2b:

200 ml cathode buffer for upper chamber was prepared by mixing 20 mL cathode buffer with 180 mL deionized water. 500 mL anode buffer for lower chamber was prepared by mixing 10 mL anode buffer with 490 mL deionized water. 10 μ l IFN standard was mixed with 10 μ l IEF sample buffer, same as *Pichia* IFN. Gel apparatus was set and 200 ml cathode buffer was filled in the upper chamber and 500 ml anode buffer was filled in the outer chamber. 2 μ l IEF standard was loaded on lane 1, one blank lane was left on purpose to avoid the contamination of the samples to each other. IFN standard and *Pichia* IFN were loaded. The whole apparatus was put in an ice bucket. The voltage program was set as follows: 100 V for 60 min, 200 V for 60 min and 500 V for 30 min. While in the last 30 min of the run, fixing solution was prepared by adding 11.46 g trichloroacetic acid and 4.6 g sulphosalicylic acid in 90 ml water. When the voltage program finished, the gel was washed with deionized water three times and then transferred to the fixing solution. The gel was fixed for 60 min on a shaker. Staining solution was prepared by mixing 58 mL water, 20 mL methanol, 20 mL stainer A and 2

mL stainer B of the colloidal staining kit. The gel was then destained in water for overnight.

Trypsin digestion of interferon $\alpha 2b$:

20 μ L 6M guanidine hydrochloride (MW 95.53 g/mol) in 0.1 M NH_4HCO_3 was added to 20 μ g IFN $\alpha 2b$ reference material and *Pichia* IFN $\alpha 2b$. Followed by adding 2 μ L of 1 M Dithiothreitol (DTT) solution to the solution (final concentration of DTT: 5mM) and heating at 37°C for 30 minutes. Fresh 1M iodacetamide (IAA) solution was made in a foil-wrapped tube to avoid exposure to light and was added to the denatured, reduced IFN solution to final concentration of 10mM. The tube was covered with aluminum foil and incubated at room temperature (25 °C) for 1hr. After incubation, the solution was buffer exchanged to 100 mM ammonium bicarbonate pH8 using 10 KDA cutoff spin column with speed of 14000 rpm for 6 min. For buffer exchange, 100 mM ammonium bicarbonate solution was added with final volume 500 μ l in the cut-off tube and reduced to 100 μ l each time. This step was repeated three times to ensure pH 8 solution for enzyme digestion. An initial aliquot of trypsin (1 μ l at 0.5 μ g/ μ l) was added to the solution and incubated at 37°C for 4 hours. Then a second aliquot of the same amount of enzyme was added for incubation overnight (16 h) at room temperature. The digestion was stopped by adding 4 μ l 1% formic acid to final pH around 3. The peptide solution was aliquoted to a 30 μ l/vial and stored at -80 °C before LC/MS analysis.

GluC digestion of interferon $\alpha 2b$:

20 μ L 6M guanidine hydrochloride (MW 95.53 g/mol) 0.1 M ammonium bicarbonate was added to 20 μ g IFN $\alpha 2b$ reference material and *Pichia* IFN $\alpha 2b$. Followed by adding 2 μ L of 1 M Dithiothreitol (DTT) solution to the solution (final concentration: 5mM DTT)

and heated at 37°C for 30 minutes. Fresh 1M iodacetamide (IAA) solution was made in a foil-wrapped tube to avoid exposure to light and was added to the denatured, reduced IFN solution to final concentration of 10mM. The tube was covered with aluminum foil and incubated at room temperature (25 °C) for 1hr. After incubation, the solution was buffer exchanged to 100 mM ammonium bicarbonate pH 8 using 10 KDA cutoff spin column with speed of 14000 rpm for 6 min. For buffer exchange, 100 mM ammonium bicarbonate solution was added to final volume 500 µl in the cut-off tube and reduced to 100µl each time. This step was repeated three times to achieve pH 8 solution for enzyme digestion. First aliquot of GluC (1 µl at 0.5µg/µl) was added to the solution and incubated at 37 °C for 4 h. Then a second dose of the same amount of enzyme was used for incubation overnight (16 h at room temperature. The digestion was stopped by adding 4 µl 1% formic acid to make the final pH3. The peptide solution was aliquoted to a 30 µl/vial and stored at -80 °C before LC/MS analysis.

In-gel digestion of IFN γ gel:

Day 1: The gel bands were cut into small pieces using a scalpel. 200 µl of acetonitrile (ACN) and LC-MS water (1:1) was added to cover the gel pieces and shake at 800 rpm on Eppendorf incubator for 1 hour until the gel pieces are clear. The liquid was removed. 100 µl of acetonitrile was added to shrink the gel pieces and the liquid was removed afterwards. The gel pieces were then rehydrated with trypsin solution (12.5 ng/µl trypsin in 50 mM ammonium bicarbonate) at 4 °C for 30 min. After rehydration, the remaining supernatant was removed, 100 µl 50mM ammonium bicarbonate was added to cover the gel pieces followed by incubation at 37 °C overnight. Day 2: 50 µl 25mM ammonium bicarbonate was added to the gel pieces and incubated at 37 °C for 15 min

with shaking. Liquid was spun down and 100 μ l acetonitrile was added followed by incubation at 37 °C for 15 min with shaking. The supernatant was spun down and collected in another tube. The extraction steps were repeated 3 times. In the last extraction step, 50 μ l 1% formic acid was added to stop the digestion. The combined extracts were concentrated to 10 μ l using a speed vacuum concentrator to around 10 μ l.

LC-MS/MS:

An Ultimate 3000 nano LC pump (Dionex, Mountain View, CA) and self-packed C18 column (Magic C18, 200Å pore and 5 μ m particle size, 75 μ m internal diameter (ID) by 100 mm) connected to a coated 10 μ m ID emitter (New Objective, Woburn, MA) were coupled online to an LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) through a nanospray ion source (New Objective, Woburn, MA). Mobile phase A was 0.1 % formic acid in HPLC grade water and mobile phase B was 0.1 % formic acid in acetonitrile. The flow rate was 250 nL/min with 2 % B for 25 min. The flow rate was maintained at 200 nL/min during the separation. The gradient is as follows: 0-60 min 2-40 % B, 60-70 min 40-90 % B, 70-75 min isocratic at 90 %B and 75-78 min 2% B. The mass spectrometer was operated in a data dependent mode (DDA) to switch between MS and CIDMS2. Briefly, after a full-scan MS spectrum from m/z 400-2000 in the linear ion-trap, 8 CIDMS2, with 28 % normalized collision energy and activation Q at 25, activation steps were performed on the 8 most intense precursor ions from the full scan.

Results:

Detailed analytical information is critical for process development at an early stage where prompt analytical feedback will aid optimal strain development and high yield fermentation conditions, as well as characterization of the final product and the

demonstration of biosimilarity. The examination of two related but quite different interferons allowed the exploration of the development of analytical methods suitable for the identification and characterization of a range of variants that are produced in the two processes. In the case of *Pichia* IFN $\alpha 2b$, incomplete removal of N-terminal signal peptide lead to N-terminal variants, the presence of methionine residues results in the potential for oxidation, amino acid substitution was also observed. For *Pichia* IFN γ , the glycosylation motifs N-X-S/T, N98 (NYS) and N26 (NGT) in the amino acid sequence can lead to a high level of glycosylation. The similarity of the amino acid sequence in the C-terminus and the signal peptide can be the cause of C-terminal proteolysis, as the enzyme responsible for removal of signal peptide can also recognize the amino acids near the C-terminus. Methionine and asparagine residues have the potential of oxidation and deamidation.

Part A: Analysis of Interferon (IFN) $\alpha 2b$ reference material and *Pichia* IFN $\alpha 2b$

SDS-PAGE of IFN $\alpha 2b$

Molecular weight determination by SDS-PAGE can provide information about the presence of product variants such as residual leader sequence and proteolytic processing, as well as a comparison between the process sample and a reference material. As shown in **Figure S1**, both IFN $\alpha 2b$ reference material and *Pichia* $\alpha 2b$ have the molecular weight of approximately 19 KDA and a single band on SDS-PAGE. For an in-depth comparison of standard and the *Pichia-derived* material, LC/MS analysis was then performed on the digest of the two samples (see later).

IEF gel of IFN $\alpha 2b$

Isoelectric focusing can be used to detect charge variants present in a process development sample such as deamidation and changes in the degree of sialylation. In **Figure 1** we compared the observed pI of a *Pichia* IFN with the reference material. While the IFN $\alpha 2b$ reference material showed the expected pI and was consistent with the theoretical value of 5.99 (ExPASy Compute pI/MW), the *Pichia* sample showed a more acidic main band together with a faint extra acidic band. The nature of these modifications were then determined by LC-MS analysis.

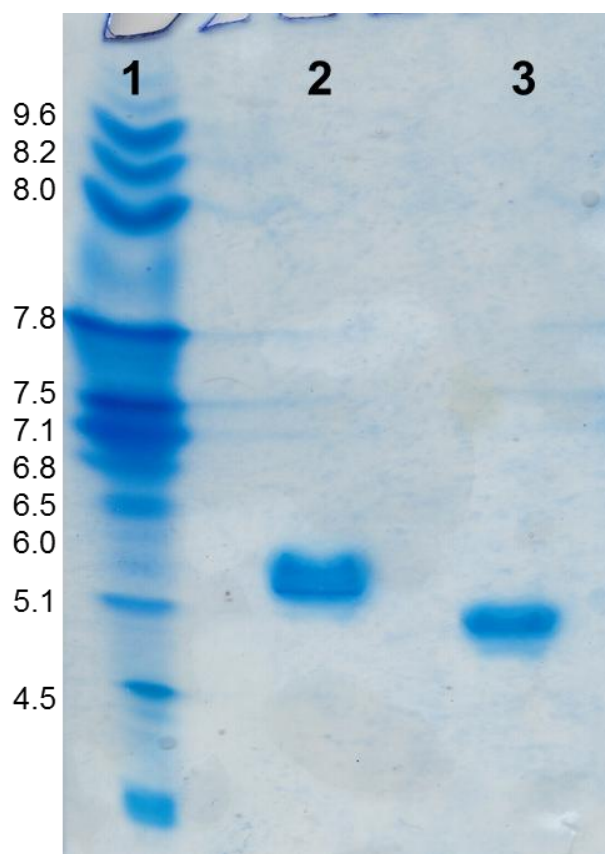


Figure 1: IEF gel of IFN $\alpha 2b$ reference material and *Pichia* IFN $\alpha 2b$.

Lane 1: IEF standards, Lane 2: IFN α 2b reference material, Lane 3: *Pichia* IFN α 2b material.

Trypsin digestion and LC/MS analysis of IFN α 2b reference material and *Pichia* sample:

Determination of amino acid sequence:

The goal of peptide mapping is to ensure that the primary structure of the drug substance is correct by obtaining 100% sequence coverage. Sometimes, one type of enzyme digestion will not give complete coverage due to factors such as limited cleavage sites in the protein with the generation of large peptides not suitable for LC-MS analysis or peptides which are too short and polar to be retained on the HPLC column or too hydrophobic for elution. The peptides identified with trypsin digestion of IFN α 2b standard and *Pichia* IFN α 2b are listed in Table 1. In this analysis, tryptic peptide T10 which contains residues 84 to 112 was not identified in the *Pichia* IFN analysis but was observed in the reference material. While the protein sequence for approved interferon α 2b remains confidential, our LC/MS analysis confirmed the DrugBank database (<https://www.drugbank.ca/>) information which is at variance with the Uniprot (<http://www.uniprot.org/>) sequence information (N vs. K at residue 112 respectively). We then confirmed that the sequence of the *Pichia* material was consistent with the sequence reported in Drugbank database by Glu-C digestion to generate a peptide length suitable for LC-MS analysis and that contains residue 112 as in TPLMNE (108-113) as in **Figure 2**.

Peptide #	AA#	Sequence	Observed m/z	Theoretical m/z	Charge	ppm	RT (min)
1	1-12	CDLPQTHSLGSR	685.826	685.828	2	2	26.4
2	13-22	RTLMLLAQMR	411.570	411.570	3	1	48.6
3	14-22	TLMLLAQMR	538.800	538.801	2	3	44.9
4	23-31	RISLFCLK	562.318	562.318	2	0	52.7
5	24-31	ISLFCLK	484.267	484.268	2	1	44.4
6	32-49	DRHDFGFPQEEFGNQFQK	1113.501	1113.504	2	2	47.7
7	50-59	AETIPVLHEM	570.290	570.292	2	3	39.2
8	60-70	IQQIFNLFSTK	669.875	669.875	2	1	50.0
9	17-83	DSSAAWDETLDDK	725.838	725.839	2	1	39.2
10	84-112	FYTELYQQLEACVIQGVGTETPLMK	1120.559	1120.556	3	3	63.9
11	113-120	EDSILAVR	451.751	451.751	2	1	30.6
12	121-125	KYFQR	371.205	371.206	2	2	25.9
14	126-131	ITLYLK	750.474	750.476	1	3	36.9
15	132-133	EK	ND	275.148	1	NA	ND
16	134-144	KYSPCAWEVVR	465.569	465.568	3	3	39.6
17	145-149	AEIMR	619.321	619.323	1	4	17.3
18	150-162	SFSLSTNLQESLR	741.382	741.383	2	1	42.4
19	163-165	SKE	ND	363.188	1	NA	ND

Table 1: Peptide mapping result for IFN α 2b standard using trypsin in-gel digestion.¹

¹ Several short peptides with two or three amino acids (T15, 2 amino acids) and T19, 3 amino acids), were too polar to retain on C18 column for LC-MS analysis. These missing amino acids were identified by Glu-C digestion to achieve 100% sequence coverage (data not shown).

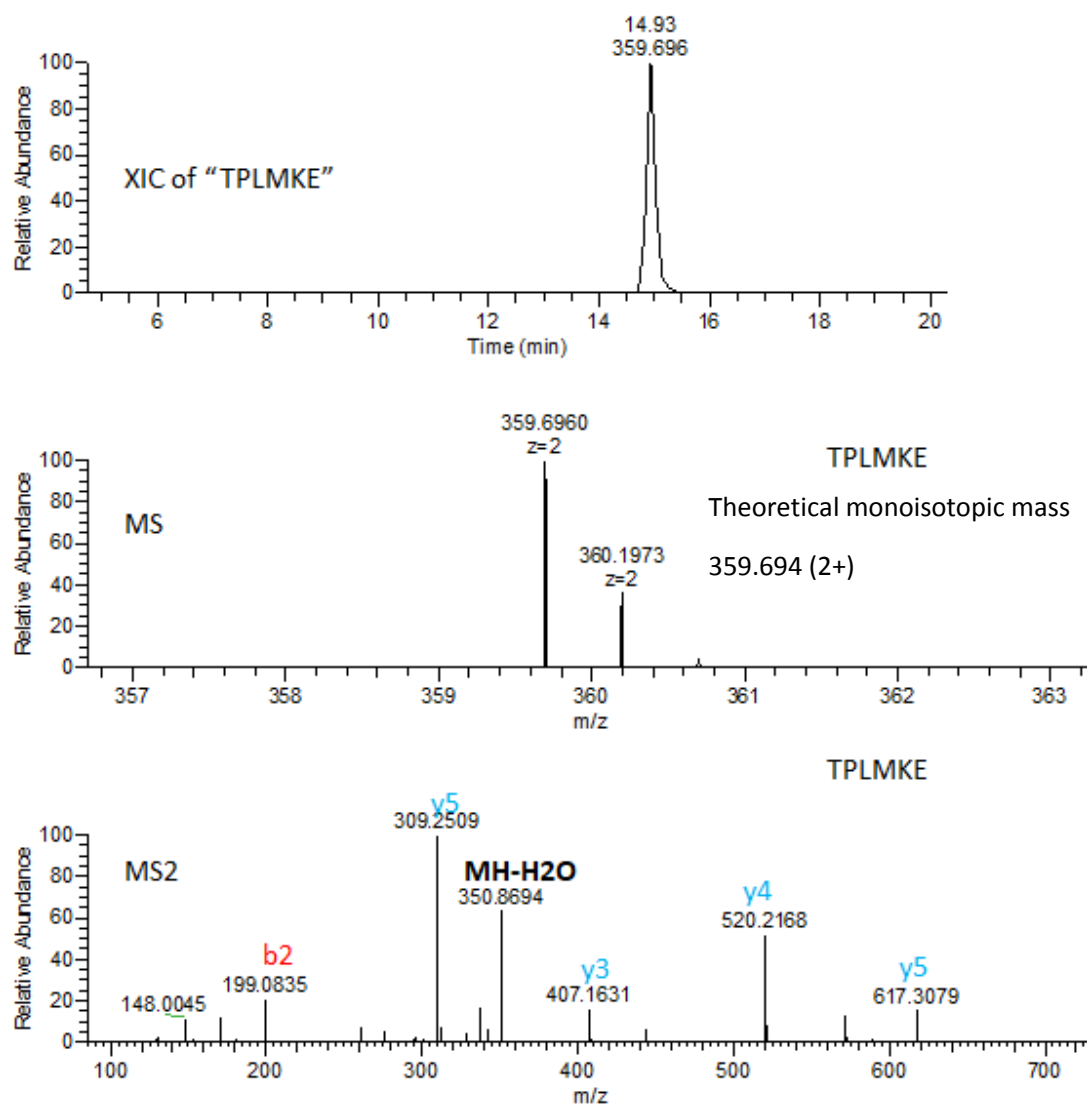


Figure 2: Confirmation of amino acid substitution K112 to N 112 of *Pichia* IFN α 2b by GluC digestion.

A: Extracted Ion Chromatogram of "TPLMNE", B: Monoisotopic mass of "TPLMNE", C: CID MS2 of "TPLMNE".

Analysis of residual leader sequence:

To examine if there is any residual leader sequence present in the process development samples we customized the expected protein sequence by combining the amino acid sequences for α factor and IFN α 2b. The trypsin digest LC/MS data was then searched against the combined FASTA file using Thermo Biopharma Finder 2.0. To effectively use Biopharma Finder for in-depth characterization of recombinant proteins there are several parameters critical for an effective database search. These parameters include S/N ratio (set at default value 20, unless the identification of a very low level of variant is needed), mass accuracy (set as 5 ppm with a minimum confidence level of 0.8 to ensure low level of false positive identifications). With this approach, we determined that “EEGVSLEKR” was still attached to the N-terminus of the *Pichia* IFN α 2b via the observation of the peptide “EEGVSLEK (see **Figure S2**) as well as “RCDLPQTHSLGSR” but these peptides were not present in the reference material (see **Figure S3**).

Oxidation analysis

The oxidation level comparison of reference IFN α 2b and *Pichia* material was shown in Table 2. The percentage of oxidation was calculated using the following equation: $\text{peak area of oxidized peak} / (\text{peak area of oxidized peak} + \text{peak area of unoxidized peak}) \times 100\%$. Except for the residue Met16 in *Pichia* material, which is oxidized at a slightly higher level than the standard material, the other four methionine residues have a similar oxidation level to the standard. The low level of oxidation observed with the reference material also indicates that our in-solution digestion protocol induced little artificial oxidation.

Site	Standard	<i>Pichia</i>
Met 16	2.3%	6.9%
Met 21	2.9%	1.9%
Met 59	3.6%	5.8%
Met 111	1.0%	2.6%
Met 159	2.7%	1.4%

Table 2: Oxidation analysis of IFN α 2b reference material and *Pichia* material.

Part B: Analysis of Interferon (IFN) γ reference material (Abchem) and *Pichia* IFN γ

In-gel digestion of SDS-PAGE gel bands IFN γ with trypsin: N-glycosylation of *Pichia* IFN γ

As stated above, for *Pichia* IFN γ , glycosylation motif N-X-S/T, N98 (NYS) and N26 (NGT) present in the amino acid sequence can lead to a high level of glycosylation. The similarity of the amino acid sequence in the C-terminal and the signal peptide is the cause of C-terminal proteolysis, as the enzyme responsible for removal of signal peptide also recognize the amino acids near the C-terminus. Methionine and asparagine residues have the potential for oxidation and deamidation.

As shown in the gel image **Figure S4**, there were two bands observed for *Pichia* IFN fermentation supernatant (MW 17 KDA and 19 KDA) and one band for the reference

material (Abcam) at 15 KDA. LC-MS analysis of the in-gel trypsin digests of these bands was performed and the raw data was searched against the amino acid sequence of IFN γ together with that of the α factor (BioPharma Finder 2.0). For the 15 KDA reference material band, except a short peptide that has “KR” repeats, all of the expected sequences were identified. While in the 17 KDA *Pichia* IFN sample, all tryptic peptides were identified except the peptide containing residues 91-108 “DDFEKLTNYSVTDLNVQR” and the C-terminal peptide 127-144 “TGKRKRSQMLFRGRRASQ”. In the 19 KDA *Pichia* IFN γ band, three peptides 15-35 “YFNAGHSDVADNGTLFLGILK”, 91-108 “DDFEKLTNYS VTDLNVQR” and C-terminal peptide “TGKRKRSQMLFRGRRASQ” were not observed. After inspecting the amino acid sequence of these two missing peptides, we found each peptide contains an N-glycosylation motif N-X-S/T, N98 (NYS) and N26 (NGT). In the 17 KDA band, N98 was fully glycosylated as shown in **Figure 3** and N26 was not glycosylated. In the 19 KDA band, both N 26 (see **Figure S5**) and N 98 were fully glycosylated. N26 had mannosylation from Man9 extending to Man16 with Man 11 the highest glycoform. N98 contained mannosylation from Man6 to Man12 with Man9 as the highest abundance glycoform. Mannosylation was also the reason for the molecular weight shift that was observed on SDS-PAGE: the mass of glycan GalNAc2Man9 is around 1865 Da, which counts for the ~2KDA shift for 17 KDA band. The mass of glycan GalNAc2Man11 is around 2188 Da, which accounts for the additional 2 KDA mass shift for 19 KDA band.

#	AA#	Sequence	Observed m/z	Theoretical m/z	charge	ppm	RT(min)
1	1-14	MQDPYVKEAENLKK	564.961	564.959	3	3	24.33
2	15-35	YFNAGHSDVADNGTLFLGILK	1126.572	1126.571	2	1	42.83
3	36-44	NWKEESDRK	397.864	397.863	3	2	17.18
4	45-53	IMQSQIVSF	526.777	526.776	2	2	30.24
5	54-56	YFK	457.245	457.245	1	1	16.52
6	57-69	LFKNFKDDQSIQK	537.625	537.623	3	4	23.25
7	70-81	SVETIKEDMNVK	696.857	696.856	2	2	19.46
8	82-88	FFDSNKK	443.228	443.227	2	2	18.20
9	89-95	KRDDFEK	313.163	313.163	3	1	16.45
10	96-108	LTNYSVTDLNVQR	761.898	761.897	2	2	29.53
11	109-118	KAIHELIQVM	591.340	591.339	2	1	31.10
12	119-126	AELSPAARK	786.436	786.436	1	1	17.57
13	127-144	TGKRKR	ND	745.479	1	NA	ND
14	133-144	SQMLFRGRRASQ	718.879	718.881	2	-2	27.23

Table 3: Peptide mapping of IFN γ reference material with trypsin digestion.

Except 127-132 “TGKRKR” that has the KRKR repeats, all the peptides were covered by trypsin digestion.

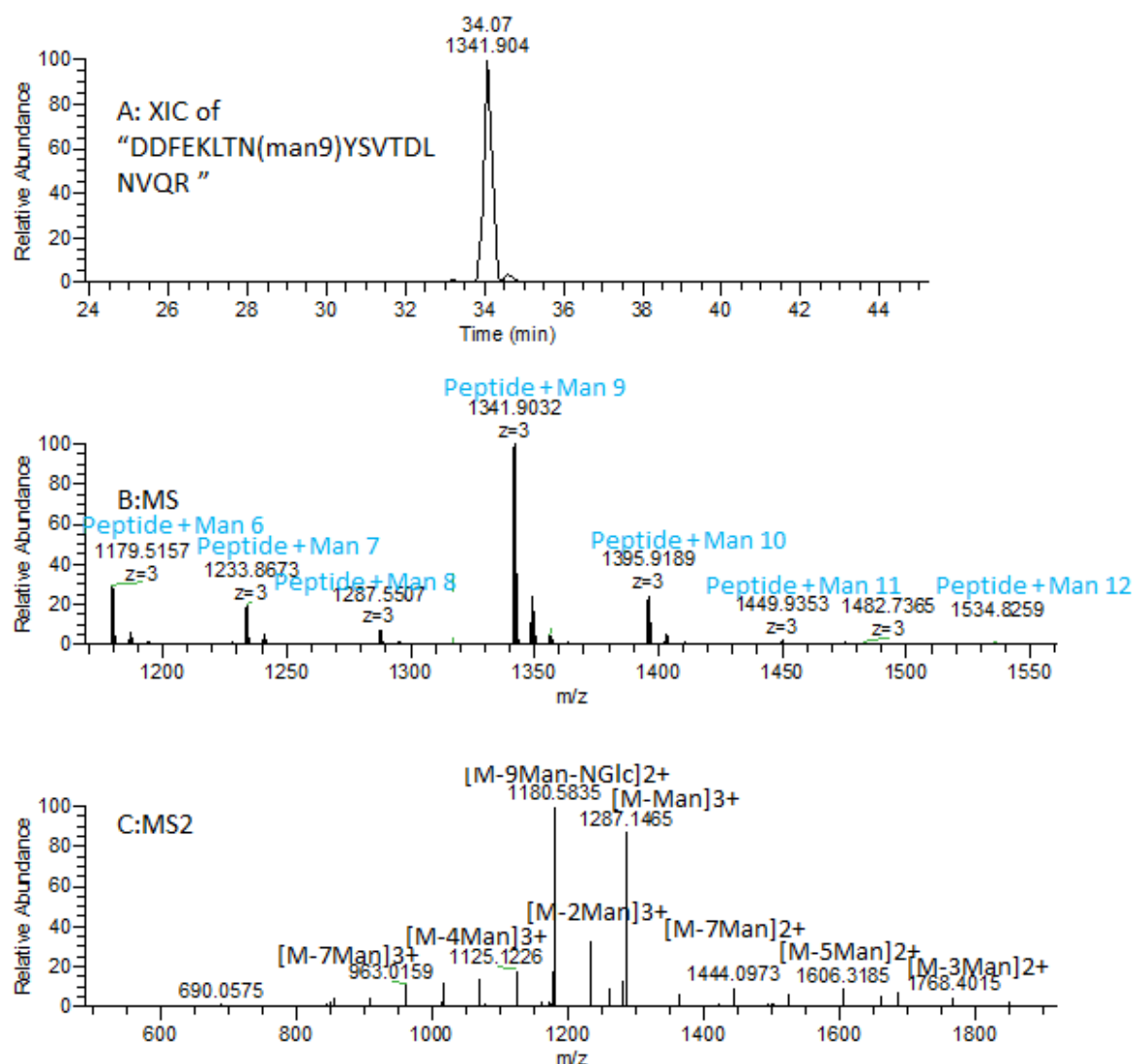


Figure 3: Identification of Man9 glycosylation at N98 in the 17 KDA band.

A: Extracted Ion Chromatogram of "DDFEKLTN (man9) YSVTDLNVQR" B:

Monoisotopic mass of mannosylation from Man6 to Man12, all the glycoforms coelute.

C: CIDMS2 of the highest glycoform Man9.

Oxidation and deamidation

The analysis of crude IFN γ process development sample requires an in-gel digestion analysis which can induce high levels of degradation products during the extensive sample preparation steps. The IFN γ amino acid sequence contains two deamidation hot-spots, NG of N26 and NS of N84. In the 17 KDa band, the deamidation level of N26 is >90% and N84 is 20%. In the 19 KDa band, as N26 is fully glycosylated, no deamidation was observed in that site. The observed amounts of oxidation and deamidation level are summarized in Table 4.

Sites	Modification	Reference	Pichia 17KDa	Pichia 19KDa
Met1	Oxidation	76.6%	87.8%	72.0%
Met46	Oxidation	81.0%	85.4%	77.2%
Met78	Oxidation	67.6%	72.4%	72.6%
Met118	Oxidation	22.2%	38.5%	21.7%
Asn26	Deamidation	>90%	>90%	NA
Asn84	Deamidation	5.2%	22.7%	6.5%

Table 4: Oxidation analysis of IFN γ reference material and *Pichia* material.

FOOTNOTE NA = Not available

In-gel digestion of IFN γ SDS-PAGE gel bands with Glu-C: C-terminal proteolysis

GluC digestion was performed on the two bands of lane 15 in the SDS-PAGE gel as well as the reference material (Lane 1). As there are only a few glutamic acid residues ("E") in the IFN γ sequence, the observed sequence coverage by Glu-C digestion was low. However, residues121-129 "LSPA AKTGK" was identified in the *Pichia* IFN γ but not

the reference material indicating the proteolysis had occurred at residue 129 (K) in the *Pichia* material (**Figure S6**).

Discussion:

Characterization of process development material at an early stage is a challenging yet important task because early analytical information can accelerate the development process. The analytical challenges include low concentrations of the product in fermentation samples, high level of interferences caused by media components and host cell proteins (HCPs) as well as unknown modifications in the product molecule. Recently, mass spectrometry based methods have enjoyed a dramatic growth in popularity, as they offer a variety approaches to efficiently gain diverse information on product quality. For biosimilar development, there is both the need for efficient process development material and a suitable reference material. For the approval process, it is necessary to characterize the protein amino acid sequence at a 100% level as well as any post-translational modifications and degradation products. Since any structural changes can have effects on drug potency and have safety effects such as generation of immune responses in a patient,¹⁶ it is necessary to have analytical methods for complete product characterization.

As discussed previously, *Pichia pastoris* has gained wide application for protein production with several successful cases of recombinant therapeutic protein production in yeast which demonstrated its unique features such as low cost, high growth rate, usage of defined media and low level of secretion of HCPs.¹⁷ For example, Biocon, a Indian biotechnology company, implement its *Pichia pastoris* platform for the production of insulin and insulin analogs and successfully commercialized the large-scale

manufacture of human insulin. Common characteristics of protein variants expressed in *Pichia pastoris* include residual leader sequence when using an α factor signal peptide and high mannose N-glycosylation structures. The complex glycosylation patterns can also impede the pharmaceutical development process and hyper mannosylation and terminal α -1, 3-mannose linkages can result in rapid in vivo clearance as well immunogenicity of the recombinant protein in clinical studies.¹⁸⁻²⁰

Since interferon α 2b and interferon γ have distinctly different amino acid sequences therefore can contain a range of product variants at various sites. Thus, multiple analytical strategies are required for the in-depth characterization of all the potential variants. As summarized in Tables 5 and 6, the variants of these two IFN product require the design of different LC/MS analytical methods where it is important to optimize the selection of the appropriate protease to generate peptides containing the variant sites that are appropriate for LC-MS analysis. In the case of the IFN γ process development, the need for the analysis of crude fermentation samples added an additional analytical challenge. In this case, we demonstrated that in-gel digestion rather than in-solution digestion is the method of choice. The SDS-PAGE approach used in this analysis has the following advantages: removal of any detergents present in the sample that can be detrimental to mass spectrometric analysis, the target protein can be separated from most HCPs, proteolysis and covalent aggregation can be observed with reduced and non-reduced gels ahead of MS analysis. One drawback of the in-gel digestion method is the high level of oxidation artifacts that can be induced during the sample preparation process. On the other hand, in-solution enzyme digestion which is

suitable for purified material as the case of IFN $\alpha 2b$ can greatly reduce the level of artifacts.

Interferon $\alpha 2b$			
residue	variants	peptides	enzyme
EEGVSLER	Leader sequence	EEGVSLER and RCDLPQTHSLGSR	trypsin
112 N	Amino acid substitution	118-123: TPLMNE	Glu-C
16 Met	Oxidation	14-22: TLMLLAQMR	trypsin
21 Met	Oxidation	14-22: TLMLLAQMR	trypsin
59 Met	Oxidation	50-59: AETIPVLHEM	trypsin
111 Met	Oxidation	118-123: TPLMNE	Glu-C
159 Met	Oxidation	145-149: AEMIR	trypsin

Table 5: Variants identified for *Pichia* IFN $\alpha 2b$.

Interferon γ			
residue	variants	peptides	enzyme
26 Asn	Glycosylation	15-35: YF N AGHSDVADNGTLFLGILK	trypsin
98 Asn	Glycosylation	96-108: LT N YSVTDLNVQR	trypsin
129 K	Proteolysis	121-129: LSPAAKTG K	Glu-C
1 Met	Oxidation	1-7: M QDPYVK	trypsin
46 Met	Oxidation	45-53: I MQSQIVSF	trypsin
78 Met	Oxidation	70-81: SVETIKED M NVK	trypsin
118 Met	Oxidation	109-118: KAIHELIQ V M	trypsin
26 Asn	Deamidation	15-35: YFNAGHSDVAD N GTLFLGILK	trypsin
84 Asn	Deamidation	82-87: FF N SNK	trypsin

Table 6: Variants identified for *Pichia* IFN γ .

The above HPLC/MS analyses generated with our multi-enzyme digest strategy was developed for the characterization of IFNs $\alpha 2b$ and γ in process development samples. For *Pichia* IFN $\alpha 2b$ the major variants that were observed were an amino acid substitution and residual leader sequence. Amino acid substitutions can occur by clonal variation in the host cell line²¹ or the use of a different cDNA sequence used in cell line construction²². Such changes may obviously have effects on protein structure and function²³ and typically can only be detected via a complete determination of the amino acid sequence, but in the case of a charged to neutral amino acid substitution such a variant can be readily screened by isoelectric focusing (IEF). In this example, the variant was a substitution of lysine to asparagine (K112N, pKa10.6, reference material) which results in the acidic shift observed in the IEF gel (see **Figure 1**). In this manner, an IEF gel can be used as a quick assessment of product variants with pKa differences that can be then used to assess process development samples.

. Every aspect of the protein structure including amino acid sequence, post-translational modifications (PTMs) is related to the characteristics of a process, including strain construction, cultivation conditions and downstream purification. The quality of early stage process development material reflects the efficacy of the manufacturing platform and an early stage analytical evaluation is critical for host system selection, media optimization and cultivation conditions. For example, a high level of oxidation in the product may be due to a high oxygen density in bioreactor, proteolysis characterization can be beneficial for selection of a suitable protease inhibitor, aggregation characterization is useful for the optimization of pH in the fermentation and suggest the need for addition of detergents to the formulation buffer.²⁴

Conclusion:

A multi-enzyme digestion approach was successfully developed for the characterization of interferon products expressed from *Pichia pastoris*. For IFN α 2b, trypsin digestion was used initially for peptide mapping, identification of leader sequence and amino acid substitution. Glu-C digestion also provided complementary data for the characterization of an amino acid substitution. An IEF gel method followed by LC/MS analysis was developed as a quick way to monitor the acidic shift of pI due to an amino acid substitution.

In-gel digestion with trypsin or Glu-C was developed for the characterization of a crude, unpurified IFN γ strain supernatant. The material was shown to have mannosylation at the two N-glycosylation sites and C-terminal proteolysis. *Pichia pastoris* has gained popularity and is attractive to the biotherapeutic industry due to its rapid growth in culture and efficient protein secretion. The method developed above could be extremely

helpful especially for process development at an early stage where prompt analytical feedback can help optimize strain development and fermentation conditions in *Pichia pastoris* strains.

Acknowledgement

This work was supported by the Defense Advanced Research Projects Agency (DARPA) and SPAWAR Systems Center Pacific (SSC Pacific) under Contract No. N66001-13-C-4025. The content is solely the responsibility of the authors and does not necessarily represent the official views of DARPA or SSC.

Author Information

* Corresponding Author

wi.hancock@northeastern.edu

Notes

The author declares no competing financial interest. And no patients or animals are involved in this research.

Reference:

- (1) Asmana Ningrum, R., Human interferon alpha-2b: a therapeutic protein for cancer treatment. *Scientifica (Cairo)* **2014**, 2014, 970315.
- (2) Parker, B. S.; Rautela, J.; Hertzog, P. J., Antitumour actions of interferons: implications for cancer therapy. *Nat Rev Cancer* **2016**, 16 (3), 131-44.
- (3) Liu, Y. H.; Wylie, D.; Zhao, J.; Cure, R.; Cutler, C.; Cannon-Carlson, S.; Yang, X.; Nagabhushan, T. L.; Pramanik, B. N., Mass spectrometric characterization of the isoforms in *Escherichia coli* recombinant DNA-derived interferon alpha-2b. *Anal Biochem* **2011**, 408 (1), 105-17.

- (4) Schroder, K.; Hertzog, P. J.; Ravasi, T.; Hume, D. A., Interferon-gamma: an overview of signals, mechanisms and functions. *J Leukoc Biol* **2004**, 75 (2), 163-89.
- (5) <https://adisinsight.springer.com/drugs/800038646>
- (6) Ahmad, M.; Hirz, M.; Pichler, H.; Schwab, H., Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol* **2014**, 98 (12), 5301-17.
- (7) Waters, M. G.; Evans, E. A.; Blobel, G., Prepro-alpha-factor has a cleavable signal sequence. *J Biol Chem* **1988**, 263 (13), 6209-14.
- (8) Cereghino, J. L.; Cregg, J. M., Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* **2000**, 24 (1), 45-66.
- (9) He, Z.; Huang, Y.; Qin, Y.; Liu, Z.; Mo, D.; Cong, P.; Chen, Y., Comparison of alpha-factor preprosequence and a classical mammalian signal peptide for secretion of recombinant xylanase xynB from yeast *Pichia pastoris*. *J Microbiol Biotechnol* **2012**, 22 (4), 479-83.
- (10) Fitzgerald, I.; Glick, B. S., Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. *Microb Cell Fact* **2014**, 13 (1), 125.
- (11) Gemmill, T. R.; Trimble, R. B., Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim Biophys Acta* **1999**, 1426 (2), 227-37.
- (12) Hamilton, S. R.; Davidson, R. C.; Sethuraman, N.; Nett, J. H.; Jiang, Y.; Rios, S.; Bobrowicz, P.; Stadheim, T. A.; Li, H.; Choi, B. K.; Hopkins, D.; Wischnewski, H.; Roser, J.; Mitchell, T.; Strawbridge, R. R.; Hoopes, J.; Wildt, S.; Gerngross, T. U.,

Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **2006**, 313 (5792), 1441-3.

(13) Wang, D.; Ren, H.; Xu, J. W.; Sun, P. D.; Fang, X. D., Expression, purification and characterization of human interferon- γ in *Pichia pastoris*. *Mol Med Rep* **2014**, 9 (2), 715-9.

(14) Choudhary, G.; Wu, S. L.; Shieh, P.; Hancock, W. S., Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *J Proteome Res* **2003**, 2 (1), 59-67.

(15) Wang, Y. A.; Wu, D.; Auclair, J. R.; Salisbury, J. P.; Sarin, R.; Tang, Y.; Mozdierz, N. J.; Shah, K.; Zhang, A. F.; Wu, S. L.; Agar, J. N.; Love, J. C.; Love, K. R.; Hancock, W. S., Integrated Bottom-Up and Top-Down Liquid Chromatography-Mass Spectrometry for Characterization of Recombinant Human Growth Hormone Degradation Products. *Anal Chem* **2017**, 89 (23), 12771-12777.

(16) Moussa, E. M.; Panchal, J. P.; Moorthy, B. S.; Blum, J. S.; Joubert, M. K.; Narhi, L. O.; Topp, E. M., Immunogenicity of Therapeutic Protein Aggregates. *J Pharm Sci* **2016**, 105 (2), 417-430.

(17) Bollok, M.; Resina, D.; Valero, F.; Ferrer, P., Recent patents on the *Pichia pastoris* expression system: expanding the toolbox for recombinant protein production. *Recent Pat Biotechnol* **2009**, 3 (3), 192-201.

(18) Kalidas, C.; Joshi, L.; Batt, C., Characterization of glycosylated variants of beta-lactoglobulin expressed in *Pichia pastoris*. *Protein Eng* **2001**, 14 (3), 201-7. (19) Teh, S. H.; Fong, M. Y.; Mohamed, Z., Expression and analysis of the glycosylation properties

of recombinant human erythropoietin expressed in *Pichia pastoris*. *Genet Mol Biol* **2011**, 34 (3), 464-70.

(20) Dai, M.; Yu, C.; Fang, T.; Fu, L.; Wang, J.; Zhang, J.; Ren, J.; Xu, J.; Zhang, X.; Chen, W., Identification and Functional Characterization of Glycosylation of Recombinant Human Platelet-Derived Growth Factor-BB in *Pichia pastoris*. *PLoS One* **2015**, 10 (12), e0145419.

(21) Asare, E.; Mugavero, J.; Jiang, P.; Wimmer, E.; Paul, A. V., A Single Amino Acid Substitution in Poliovirus Nonstructural Protein 2CATPase Causes Conditional Defects in Encapsidation and Uncoating. *J Virol* **2016**, 90 (14), 6174-6186.

(22) Ruiz-Jarabo, C. M.; Pariente, N.; Baranowski, E.; Dávila, M.; Gómez-Mariano, G.; Domingo, E., Expansion of host-cell tropism of foot-and-mouth disease virus despite replication in a constant environment. *J Gen Virol* **2004**, 85 (Pt 8), 2289-97.

(23) Ng, P. C.; Henikoff, S., Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **2006**, 7, 61-80.

(24) Crowell, L. E.; Lu, A. E.; Love, K. R.; Stockdale, A.; Timmick, S. M.; Wu, D.; Wang, Y. A.; Doherty, W.; Bonnyman, A.; Vecchiarello, N.; Goodwine, C.; Bradbury, L.; Brady, J. R.; Clark, J. J.; Colant, N. A.; Cvetkovic, A.; Dalvie, N. C.; Liu, D.; Liu, Y.; Mascarenhas, C. A.; Matthews, C. B.; Mozdierz, N. J.; Shah, K. A.; Wu, S. L.; Hancock, W. S.; Braatz, R. D.; Cramer, S. M.; Love, J. C., On-demand manufacturing of clinical-quality biopharmaceuticals. *Nat Biotechnol* **2018**. 36, 988–995.