

# Holistic Analytical Characterization and Risk Assessment of Residual Host Cell Protein Impurities in an Active Pharmaceutical Ingredient (API) Synthesized by Biocatalysts

Fengqiang Wang<sup>1</sup>, Xuanwen LI<sup>1</sup>, Michael Swanson<sup>1</sup>, Erik Guetschow<sup>1</sup>, Matthew Winston<sup>1</sup>, Joseph P. Smith<sup>1</sup>, Erik Hoyt<sup>1</sup>, Zhijian Liu<sup>1</sup>, Douglas Richardson<sup>1</sup>, Xiaodong Bu<sup>1</sup>, Vibha Jawa<sup>1</sup>, and Narayan Variankaval<sup>1</sup>

<sup>1</sup>Merck Research Laboratories

January 30, 2022

## Abstract

Host cell proteins (HCPs) are a significant class of process-related impurities commonly associated with the manufacturing of biopharmaceuticals. However, due to the increased use of crude enzymes as biocatalysts for modern organic synthesis, HCPs can also be introduced as a new class of impurities in chemical drugs. In both cases, residual HCPs need to be adequately removed to ensure product purity, quality, and patient safety. Although a lot of attentions have been focused on defining a universally acceptable limit for such impurities, the risks associated with residual HCPs on product quality, safety, and efficacy often need to be determined on a case-by-case basis taken into consideration of residual HCP profile in the product, the dose, dosage form, and administration route etc. Here we describe the unique challenges for residual HCP control presented by the biocatalytic synthesis of a Merck investigational stimulator of interferon genes protein (STING) agonist, MK-1454, which is a cyclic dinucleotide synthesized using *E. coli* cell lysate overexpressing cyclic GMP-AMP synthase (cGAS) as a biocatalyst. In this study, a holistic characterization of residual protein impurities using a variety of analytical tools, together with in silico immunogenicity prediction of identified proteins, facilitated risk assessment and guided process development to achieve adequate removal of residual protein impurities in MK-1454 API.

## Holistic Analytical Characterization and Risk Assessment of Residual Host Cell Protein Impurities in an Active Pharmaceutical Ingredient (API) Synthesized by Biocatalysts

Fengqiang Wang<sup>1, 2\*</sup>, Xuanwen Li<sup>3</sup>, Michael Swanson<sup>4</sup>, Erik Guetschow<sup>1</sup>, Matthew Winston<sup>5</sup>, Joseph P. Smith<sup>3</sup>, Erik Hoyt<sup>1</sup>, Zhijian Liu<sup>5</sup>, Douglas Richardson<sup>3</sup>, Xiaodong Bu<sup>1</sup>, Vibha Jawa<sup>4</sup>, Narayan Variankaval<sup>1\*</sup>

<sup>1</sup>Small Molecular Analytical Research & Development, Rahway, NJ

<sup>2</sup>Biologics Analytical Research & Development, Kenilworth, NJ

<sup>3</sup>Analytical Enabling Technologies, Analytical Research & Development, Kenilworth, NJ

<sup>4</sup>Pharmacokinetics, Pharmacodynamics & Drug Metabolism (PPDM), West Point, PA

<sup>5</sup>Small Molecular Process Research & Development, Rahway, NJ

Merck & Co. Inc, 2000 Galloping Hill Rd., Kenilworth, NJ 07033

\*To whom correspondence should be addressed to:

Email: Fengqiang.wang@merck.com or Narayan.variankaval@merck.com

## Abstract:

Host cell proteins (HCPs) are a significant class of process-related impurities commonly associated with the manufacturing of biopharmaceuticals. However, due to the increased use of crude enzymes as biocatalysts for modern organic synthesis, HCPs can also be introduced as a new class of impurities in chemical drugs. In both cases, residual HCPs need to be adequately removed to ensure product purity, quality, and patient safety. Although a lot of attentions have been focused on defining a universally acceptable limit for such impurities, the risks associated with residual HCPs on product quality, safety, and efficacy often need to be determined on a case-by-case basis taken into consideration of residual HCP profile in the product, the dose, dosage form, and administration route etc. Here we describe the unique challenges for residual HCP control presented by the biocatalytic synthesis of a Merck investigational stimulator of interferon genes protein (STING) agonist, MK-1454, which is a cyclic dinucleotide synthesized using *E. coli* cell lysate overexpressing cyclic GMP-AMP synthase (cGAS) as a biocatalyst. In this study, a holistic characterization of residual protein impurities using a variety of analytical tools, together with in silico immunogenicity prediction of identified proteins, facilitated risk assessment and guided process development to achieve adequate removal of residual protein impurities in MK-1454 API.

**Keywords:** Residual host cell proteins, biocatalysis, immunogenicity, MK-1454, cGAS, analytical control strategy, risk assessment

## Introduction

Host cell proteins (HCPs) are a significant class of process-related impurities that need to be monitored and adequately removed during bioprocess development (Hogwood, Bracewell, & Smales; X. Wang, Hunter, & Mozier, 2009). Residual HCP impurity in biological drug substance or drug product are generally considered as a critical quality attribute (CQA) due to its potential impact on product quality, safety, and efficacy (FDA, 1997; Guideline, 1999). The impacts can be classified into four main categories: 1) immunogenicity; 2) adjuvant effects; 3) biological activities; and 4) enzymatic activities (Jones et al., 2021; Vanderlaan et al., 2018; X. Wang et al., 2009). Since most of biologics are produced using non-human cell lines, immunogenicity is one of the major concerns caused by residual HCP impurities in biological products if not adequately removed. Immunogenicity can lead to various degrees of adverse effects in patients ranging from more severe cytokine storm, chronic inflammation, hypersensitivity to mild injection site reactions (Reijers et al., 2019). Although the majority of commercialized biologics have proven to be safe with the presence of only trace amount of residual HCPs, there are a few cases where the presence of high level of immunogenic HCPs delayed product development and approval (Vanderlaan et al., 2018). IB1001, a recombinant coagulation factor IX, caused anti-CHO HCP immune response in 30% patients receiving treatment at the time of its clinical development, which resulted in the suspension of two clinical trials of IB1001 by FDA in 2012. Despite the lack of clear correlation to adverse events, the manufacturing process was modified with the addition of hydrophobic interaction chromatography (HIC) step to reduce the level of residual HCPs in IB1001 to < 26 ng/mg from 58500 ng/mg as detected in former process, and the product was later approved by FDA in 2015 with the trade name IXINITY® (Cheung et al., 2016). Lebrikizumab, a humanized IgG4 monoclonal antibody targeting IL-13, was found to contain 242-328 ng/mg CHO phospholipase B-Like 2 (PLBL2) in the clinical material used for a pivotal phase III Lute and Verse clinical trials. Anti-PLBL2 immune response was observed in 76-90% of subjects enrolled in a Pivotal phase III Lute and Verse clinical trials. The trials were later converted to phase IIb studies although the observed immune response has not shown a direct link with clinical adverse effect or an impact to the efficacy of lebrikizumab. Phase III clinical trials done with substantially reduced levels of PLBL2 (0.2-0.4 ng/mg) showed significantly less and dose-dependent frequency of immune response to PLBL2 (Fischer et al.; Hanania et al., 2015). Other than eliciting immune response against themselves, the presence of residual HCPs in biological product can also act as adjuvant. One example is Somatropin Sandoz powder for injection (Covance), a biosimilar recombinant human growth hormone(hGH) of Pfizer's Genotropin® derived from *E. coli*. During its development, ~60% of patients enrolled developed anti-hGH antibodies during clinical trials, potentially due to the presence of high levels (~1400 ppm) of ribose phosphate isomerase (RPI) (Vanderlaan et al., 2018). The manufacturing process was later changed to further remove *E. coli* HCPs and the drug was later approved in 2006 and marketed as Omnitrope® (Romer et al., 2007; Saenger, 2009; Vanderlaan et al., 2018).

Apart from immunogenicity concerns, residual HCPs can also cause serious adverse effects due to their biological activity. For example, the presence of monocyte chemoattractant protein-1 (MCP-1) in a CTLA4-IgG1 fusion protein led to a clinical hold due to serious adverse effects related to histamine release in patients (Vanderlaan et al., 2018). Similarly, presence of an *E. coli* protein Flagellin in a biological product caused acute toxicity mediated by Toll-like receptor 5 (TRL-5) and resulted in a clinical hold of the product during development (Vanderlaan et al., 2018). In contrast to the relatively rare immunogenicity and biological activity problems related to residual HCPs, residual enzymatic activity caused by insufficient removal of HCPs appears to be quite common. Many HCPs with protease activity can cause product degradation, some of the examples including Adam 19 and Furin (Clarke et al., 2019), Cathepsin D (Bee et al.), and Cathepsin L (Luo et al., 2019). Other than degrading the drug product, certain enzymes could also degrade excipients used in the formulation of biological drugs. Some notable residual enzymes in this category include PLBL2, group XV lysosomal phospholipase A2 (LPLA2), and lipoprotein lipase (LPL) (Chiu et al.; Dixit, Salamat-Miller, Salinas, Taylor, & Basu; Hall, Sandefur, Frye, Tuley, & Huang).

As a CQA, the level of residual HCPs present in final drug substance and drug product often needs to be tested for batch release. It is both an industry-wide common understanding and a regulatory requirement to remove HCPs to acceptable low levels that will not affect product quality, safety, and efficacy (FDA, 1997; Hogwood et al.).

Unlike biologics, small molecule drugs are usually chemically synthesized, and the manufacturing process is typically free of protein residuals. However, with the increased use of biocatalysis in organic synthesis, recombinant enzymes used to catalyze chemical reactions can be introduced into small molecule chemical drugs as a new class of impurities, along with host cell components including HCPs, especially when whole-cell lysate is used as a catalyst (Reetz, 2013). Biocatalysis contributes to a greener pharmaceutical process by 1) use of highly selective enzymes so that protection and deprotection steps associated with chemical synthesis can be reduced or eliminated, hence, reducing the number of process steps and associated E-factors (Kg of waste produced per Kg of product) when compared to chemical synthesis; 2) use of mild conditions and aqueous solutions that reduce the use of hazardous reagents and organic solvents in the reactions; 3) use of enzymes with high selectivity and activity to achieve an excellent stereochemical purity and high conversion rate (Patel, 2006; Reetz, 2013; Woodley, 2008). Recent successful examples include the use of enzyme cascades to produce opioids in yeast utilizing 21-23 enzymes from plants (de Maria & Hollmann, 2015; Galanie, Thodey, Trenchard, Interrante, & Smolke, 2015) and the use of biocatalytic cascade for the manufacture of islatravir, an investigational HIV treatment (Huffman et al., 2019). Although isolated or immobilized enzymes are the ideal candidates for biocatalysis, at times, whole-cell lysate or partially purified enzymes have also been used to maintain enzymatic activity and reduce operational cost.

The analytical control strategy for enzyme and total residual proteins in biocatalytic synthesis of APIs for oral delivery is well summarized in the two papers published by Wells et al. (Wells, Finch, Michels, & Wong, 2012; Wells et al., 2016). However, when the whole-cell lysate is charged in the reaction and the API is used for parenteral dosing, there is an increased risk of immunogenicity if the enzyme and/or HCPs were not adequately removed during the isolation process. One recent example is the use of evolved cyclic GMP-AMP synthase (cGAS) to convert ATP and GTP derivatives to a cyclic dinucleotide API, MK-1454 (John A. McIntosh<sup>1\*</sup>, 2022; Novotna et al., 2019), which binds to Stimulator of Interferon Genes (STING) to initiate a downstream transcription cascade and type I interferon signaling to stimulate the immune surveillance against tumor cells (Lama et al., 2019; Zhou et al., 2018). Since cGAS needs the presence of DNA and other cofactors to be active, enzyme purification or immobilization yields very limited activity. Therefore, whole-cell lysate overexpressing cGAS is used for biocatalysis, which creates additional challenges in process development to remove cell debris, endotoxin, DNA (genomic and plasmid), and host cell proteins. Particularly, traditional chemical approaches using liquid-liquid extraction, solid phase extraction, crystallization, and recrystallization for API isolation and purification are not effective in removing residual HCPs especially when a bulk volume of cell lysate was added to the reaction mixture. As mentioned previously, the presence of residual proteins in small molecule chemical drugs can have a detrimental impact on product quality, safety, and/or efficacy, with a major concern being the immunogenicity of proteins of microbial origin. This

is challenging not only in process development but also in analytical development since the methods used for measuring trace amount of residual proteins are not typically used for oral dosage form drugs synthesized by biocatalytic routes. In this article, we will use the residual protein measurement and control in MK-1454 as an example to illustrate the analytical challenges and control strategies for residual HCP control in parenteral dosage form API synthesized using a biocatalytic route.

## Materials and Methods

### API isolation

End of reaction mixture was treated with sodium phosphate and filtered to obtain a homogenous solution with 55-63% purity as analyzed using a liquid chromatography method with liquid chromatography area percentage (LCAP). The homogeneous solution was then mixed with 18 wt% sodium carbonate and a phase cut was performed to obtain a homogenous solution with 70-80% LCAP. A pH swing recrystallization was then performed to obtain the free acid form with >99% LCAP (Figure 1a). A process change from tandem reactions to one pot reaction was later developed and the API was purified utilizing organic extraction after  $\text{Na}_2\text{HPO}_4$  treatment and filtration to extract MK-1454 API into organic phase followed by several aqueous washes and back-extraction of the API into the aqueous phase with the addition of NaOH solution. The pH was then neutralized by the addition of HCl solution to crystallize the API (Figure 1b).

### Enzyme linked immunosorbent assay (ELISA)

Enzyme linked immunosorbent assay was performed using a commercially available sandwich ELISA kit for *E. coli* HCP testing from Cygnus Technologies (Southport, NC). Both the standard included in the kit (kit standard) and an *E. coli* cell lysate supernatant were used to calibrate the amount of residual *E. coli* HCPs in the final API samples. In addition, the immunoreactivity of several other enzymes used in the reaction was also tested by the commercial kit to determine if the kit antibody recognizes those enzymes.

### One-dimensional SDS-Polyacrylamide Gel Electrophoresis (1D SDS-PAGE)

Samples at different dilutions or BSA standards at different concentrations were mixed with 5 mM DTT and NuPAGE LDS sample loading buffer (4×) to have a final concentration of 0.5-1 mM DTT and 1× sample loading buffer. The mixture was then heated at 85°C for at least 5 mins to denature proteins. Electrophoresis was performed using a NuPAGE 4-12% Bis-Tris gel (1.0 mm×15 well, Invitrogen) according to the manufacturer's instructions and the gels were run using 1×MES running buffer from a 20× stock (Thermo Fisher Scientific, Somerset, NJ). PageRuler Plus Pre-stained Protein Ladder from Thermo Fisher Scientific (Somerset, NJ) was used as molecular weight marker. Gel was run at 200 V for 35 min and later stained either with Imperial staining solution or Mass Spectrometry compatible silver staining kit (Pierce, Thermo Scientific, Somerset, NJ) according to the manuals provided in the kit.

### Western blot

Western blot was performed on API samples and ELISA kit standards using the *E. coli* HCP WB Kit from Cygnus Technologies (Southport, NC). Various amounts of standards and *E. coli* control antigens from the kit was loaded onto NuPAGE 4-12% Bis-Tris gel (1.0 mm ×15 well, Invitrogen) after mixing with DTT and 4× LDS sample loading buffer. The proteins were then resolved on the gel by running at 200 V for 35 min. Upon the completion of gel electrophoresis, one gel was stained with Imperial stain solution to visualize protein bands, and a duplicate gel was transferred to PVDF membranes using iBlot (Invitrogen, Carlsbad, CA) apparatus. After transfer, the membrane was blotted using the *E. coli* HCP WB Kit from Cygnus Technologies (Southport, NC) following the kit's instructions.

### Gel bands excising and in-gel digestion

Protein bands visualized on 1D SDS-PAGE gel by Imperial Staining solution were excised and cut with a clean razor blade into 1×1 to 2×2 mm pieces. Gel pieces were placed into a 1.5 ml low-binding Eppendorf microcentrifuge tube and 200 µL of destaining buffer (40% Acetonitrile) was added. Samples were incubated at 37°C for 30 min with shaking. Destaining buffer was then discarded and gel pieces were washed with

destaining buffer another time until the Coomassie dye was no longer visible in the gel. In-gel tryptic digestion was performed according to a previously published protocol (Paulo, 2016).

### NanoLC-MS/MS for protein identification

The extracted peptides from in-gel digestion were concentrated with a Speed-Vac concentrator. Four  $\mu\text{L}$  of peptide digest were injected via an auto-sampler and separated with a reversed-phase liquid chromatography (RP-LC) C18 column (100Å, 2  $\mu\text{m}$ , 75  $\mu\text{m} \times 500 \text{ mm}$ ) at a flow rate of 250 nL/min using an EASY-nLC<sup>TM</sup> 1200 system. Mobile phase A was water with 0.1% formic acid (FA), and mobile phase B was 80% acetonitrile with 0.1% FA. Peptides were eluted using a linear gradient with increased concentration of mobile phase B from 0 to 2% for 2 min, 2-40% for 70 min, 40-80% for 5 min, and then 80-100% for 10 min.

Peptides were acquired with a Q Exactive<sup>TM</sup> HF-X hybrid Quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific, Somerset, NJ) controlled by Xcalibur with full scan MS spectra from 300 to 2000 m/z. MS was run in data dependent mode with the parent ion being analyzed in the FTMS and the top 15 most abundant ions being selected for fragmentation and analysis. Tandem mass spectrometric data was analyzed using the Proteome Discoverer v 2.2 search algorithm (Thermo Fisher Scientific, Somerset, NJ) against the customized UniProt database ([www.uniprot.org](http://www.uniprot.org); Proteome ID UP000000625 and UP000002032) with sequences from evolved enzymes and co-factors used in the biocatalytic process combined with proteome from *Escherichia coli* (*E. coli*) strain K12. The MS data was also searched against the proteome of *E. coli* BL21-DE3 strain, from which cGAS was produced, to maximize *E. coli* HCP detection. The precursor mass tolerance was 15 ppm and fragment mass tolerance was 0.02 Da. Proteins were considered positively identified if two or more unique peptide sequences were identified with targeted false discovery rate for both peptide and protein at 1%.

### Immunogenicity risk assessment and in silico tools

Multiple factors can lead to immunogenicity. These include intrinsic factors like the sequence and structure of a given protein and the post translational modifications that can occur during the process (Jiskoot, Rispen, & Kijanka, 2019). In addition, risk factors also include those from unwanted impurities that can carryover from purification during drug manufacturing, liabilities due to formulations and excipients and devices as well as how the drug gets administered (Jiskoot et al., 2019). To understand the sequence-based risk of residual proteins associated with the API, multiple algorithm-based tools were used. The EpiMatrix algorithm evaluate the ability of processed peptides (9-mer sequences) in a protein to bind with the 8 most prevalent MHC alleles that represent over 98% of the human population (Bailey-Kellogg et al.; Goey, Bell, & Kontoravdi, 2018; Jawa et al.). Multiple high-density T-cell epitopes or clusters were assessed using the ClustiMer tool. Additionally, the epitopes that could be processed and presented were further analyzed in the JanusMatrix to determine which predicted epitopes may be cross-react with epitopes derived from the human genome on the basis of conservation of T cell receptor (TCR)-facing residues. ClustiMers with JanusMatrix scores greater than 1 were excluded from the analysis based on the assumption that the auto-reactive TCR containing cells were eliminated during T cell development (Bailey-Kellogg et al.).

## Results

### ELISA measurement of residual *E. coli* HCPs in Prep. Lab batch MK-1454 API.

A commercial *E. coli* HCP kit from Cygnus Technologies was used to estimate the total amount of residual *E. coli* HCPs in MK-1454 early process API samples (Prep. Lab. batch). Given that the reaction mixture contains not only *E. coli* HCPs but also evolved recombinant enzyme along with chaperone proteins and antibiotics resistance genes co-expressed with the enzyme and that process-specific antibodies targeting those proteins are not available, an attempt was made to use an in-house produced *E. coli* cell lysate supernatant expressing cGAS to represent the proteins used in the synthesis with the assumption that all proteins used in the process are non-specifically purged with a similar purging factor through the purification process. A standard curve was plotted using the absorbance signals ( $A_{450-650}$ , y-axis) generated from this in-house standard at different concentrations against its concentrations (ng/mL, x-axis) to calibrate the amount of

proteins remaining in the API. As shown in Figure 2, the binding affinity as indicated by the  $EC_{50}$  value (or C value) of the 4-parameter non-linear logistic fitting curve with the in-house standard ( $EC_{50} = 5199$ ) is much lower than that with the Cygnus kit *E. coli* HCP standard ( $EC_{50} = 170.5$ ). When using the Cygnus kit standard to calibrate the immune-equivalent amount of proteins included in the in-house standard, the standard series have a % recovery ranging from 0.83% to 11.49% of its nominal value (Table 1). Differences in binding affinity to the anti-HCP antibodies in ELISA between the two standard series suggest the composition differences between the kit standard and in-house standard and that a portion of proteins included in the in-house standard may not be reactive to the antibodies used in the kit. Not surprisingly, the amount of proteins in Prep. Lab. MK-1454 API samples at different concentrations measured by using each standard curve differs significantly, with the numbers obtained using in-house standard being ~16-78 times higher than those obtained using the kit standard (Figure 3a). When plotting the dilution curves of API (mg/mL, blue) along with the kit standard (ng/mL, green) and in-house standard (ng/mL, red) using the absorbance values generated by ELISA against its concentrations, we noticed the differences between API response curve and the two standard curves, indicating protein composition differences among all three of them (Figure 3b). As shown in Figure 3c, the HCP to API mass ratio (ng/mg) increases along with the dilution of API samples from 50 mg/mL to 1.28 mg/mL, regardless of using which standards to back-calculate residual HCP concentrations, then stayed relatively consistent between 1.28 mg/mL to 0.082 mg/mL and increased again along with the dilution of API samples from 0.082 to 0.013 mg/mL. This dilutional non-linearity often indicates the presence of matrix interference and/or co-purifying HCPs that have antigen excess to the antibodies used in the assay (Zhu-Shimoni et al., 2014). It also shows the complexity of protein measurement by immunoassay, where matrix interference, specificity, sensitivity, sample linearity, precision and accuracy of measurement must be carefully assessed to qualify and validate the assay for its intended purpose (FDA, 1997; Guideline, 1999). Given that multiple enzymes have been over-expressed and added to the biocatalytic route for MK-1454 synthesis (John A. McIntosh<sup>1\*</sup>, 2022), the commercial kit, with its antibody raised against generic *E. coli* HCPs, is not expected to accurately quantify the recombinant enzymes, the associated chaperones and co-expressed antibiotic resistant proteins. The inability of kit antibody to detect cGAS was confirmed by Western blot, where purified cGAS was loaded onto the SDS-PAGE gel and blotted with anti-*E. coli* HCP antibody obtained from the commercial kit (Figure 4a and 4b). The immunoreactivity of other enzymes added to the process prior to the last reaction step was also tested by the commercial ELISA kit and shown as the baseline (Figure 4c). No concentration-dependent responses were observed for each of these three purified enzymes, enzyme A (AK, adenylate kinase), B (GK, guanylate kinase), and C (AcK, acetate kinase), indicating the lack of antibody coverage to those evolved enzymes (Figure 4c). Therefore, before a process specific ELISA can be developed, orthogonal methods are required to monitor the clearance of total residual proteins including those evolved enzymes.

### **Immunoreactivity of Cygnus *E. coli* HCP kit antibody to the standards used and to the residual proteins in MK-1454 Prep. Lab. batch API**

To find out the composition difference between Cygnus kit standard and the in-house lysate supernatant standard, different concentrations of *E. coli* control antigen used as the kit standard and lysate supernatant standard (concentration determined by micro BCA protein quantitation kit) were loaded and resolved on duplicate 4-12% SDS-PAGE gels as described in the Materials and methods. One of the gels was then Coomassie blue stained and the other transferred to PVDF membrane using iBlot and blotted with Cygnus anti-*E. coli* HCP antibodies provided in the kit. The protein profile differences between the two standards are shown in Figure 4a, and their immunoreactivity to the Cygnus antibody are shown in Figure 4b. Cygnus antibodies show stronger immunoreactivity to the Cygnus kit standard than the in-house standard in the Western blot results (Figure 4b). The Cygnus standard also shows a broader representation of *E. coli* HCPs ranging from high molecular weight to low molecular weight (10-180 kDa) while the in-house lysate supernatant standard has relatively less amount of low molecular weight ([?]40 kDa) portion of proteins (Figure 4a and 4b). A densitometry analysis shows higher number of bands and higher total volume of proteins in *E. coli* control antigen than in the in-house standard detected both on the Coomassie blue stained gel and the Western blotted membrane. Cygnus antibody is also able to detect certain residual

proteins in Prep. Lab batch API, with the immunoreactivity decreases along with the decrease of API loading concentration (Figure 4d). Comparing the bands detected by Cygnus antibody in Prep. Lab batch API samples (Figure 4d) to those detected by Coomassie blue staining at the high API concentration (Figure 5a) indicates that not all proteins detected by Coomassie blue staining is detectable by the Cygnus antibody, further indicating the need of using orthogonal methods for total residual protein quantitation.

### 1D SDS-PAGE with Coomassie blue and silver stain to detect total residual proteins in API samples

Unlike ELISA that relies on specific antibody-antigen binding to quantify protein amount, SDS-PAGE resolves all proteins according to their molecular weight, and its detection sensitivity depends on the staining method used, with Coomassie blue staining has a detection sensitivity of  $\sim 10$ -100 ng/band and silver staining has a detection sensitivity of  $\sim 0.1$ -1 ng/band (Weiss W., 2009). To estimate the amount of residual protein levels in the API samples, Prep. Lab batch API at different concentrations, along with a bovine serum albumin (BSA) standard obtained from micro BCA protein assay kit (Pierce Biotechnology, Rockford, IL) at different concentrations, were loaded and separated on 4-12% Bis-Tris gel as described in Materials and methods. Duplicate gels were either stained with Coomassie blue (Imperial stain) or silver stain. As shown in Figure 5, Coomassie blue stained gel (Figure 5a) can barely detect BSA at  $\sim 0.4$   $\mu\text{g/mL}$  load ( $\sim 4.8$  ng) and silver stained gel (Figure 5b) is able to detect BSA at as low as  $0.05$   $\mu\text{g/mL}$  ( $\sim 0.6$  ng). In  $1$  mg/mL API, no protein band is detectable by Coomassie blue staining and a few bands are barely detectable with silver staining. When increasing the concentration of API, more proteins become detectable in both Coomassie blue and silver-stained gels. Although different proteins can have different response factors when stained with Coomassie blue (Congdon, Muth, & Splittgerber, 1993; "The European Agency for the Evaluation of Medicinal Products Human Medicines Evaluation Unit

", 1997; Tal, Silberstein, & Nusser, 1985), the amount of proteins detected in the API samples on the gel can be estimated using the intensity of protein bands observed and BSA as a reference standard (data analysis performed using the ImageQuant IQTL 8.0 software, GE healthcare (Piscataway, NJ). Since there is barely any protein band detectable in API at  $1$  and  $2$  mg/mL load, and BSA band is detectable at  $0.4$   $\mu\text{g/mL}$ , assuming all proteins have similar detection sensitivity as BSA, then no individual protein in the API is above  $200$  ng/mg ( $0.4$   $\mu\text{g/mL}/2$  mg/mL). The bands observed in  $4$ ,  $10$ , and  $30$  mg/mL API were also used to estimate the total protein amount in those bands according to the calibration curve generated by BSA standard using IQTL8.0 software. A second gel was run and Coomassie blue stained to assess the variability of this estimate (gel image not shown). The ng/mg difference calculated between the two gels are included in Table 2. Across three different API concentrations, the average protein amount is estimated to be  $764$  ng/mg, ranging from  $621$  to  $884$  ng/mg. Of note, since Coomassie blue non-specifically binds all proteins through ionic interactions between dye sulfonic acid groups and positive protein amine groups as well as through Van der Waals attractions, in this method, all proteins over the detection limit should be visible on the gel. The increased appearance of protein bands along with the increase of loading concentration indicates that many protein species is at an abundance level lower than the detection limit when API is loaded at low concentrations, while the dominant proteins were shown as  $6$  major bands in Figure 5 and 6. Although using SDS-PAGE gel with Coomassie blue staining for protein quantitation can have high variations due to gel loading volume variability, the differences in response factors of staining between BSA and other proteins, the time difference on staining and destaining steps etc., the results from gel analysis indicates the presence of proteins at levels much higher than that quantified by Cygnus kit, which is also indicated in our previous publication using ImageJ software coupled with densitometry analysis on 1D SDS-PAGE gel (Smith et al., 2022). To further identify those proteins and investigate whether they can be detected by the Cygnus kit antibody, we excised those bands and did in-gel digestion coupled with nanoLC-MS/MS protein identification.

### In-gel digestion and LC-MS/MS identification of protein bands detected on the gel

As shown in Figure 6,  $6$  bands were excised and a blank between band B1 and B2 were also excised as background band (BB). All seven bands were in-gel digested and proteins in those band were identified by

LC-MS/MS as described in Materials and Methods. A total of 526 proteins were identified in all bands excised with 277 proteins were identified in B1 bands, 272 proteins were identified in B2 bands, 330 in B3 bands, 380 in B4 bands, 264 in B5 bands, and 301 proteins in B6 bands, and 325 proteins in BB. The mass spectrometry identification covers 5 magnitude of order in dynamic range, with a few abundant proteins accounting for a higher % of total protein identified and more proteins are at much lower abundance. The distribution of proteins identified in each band along with their accumulative abundance were plotted in Figure 7. Proteins accounting for more than 1% of total proteins identified in that band based on the total peptide ion intensity are ranked and their accumulative abundance plotted in Figure 7 (based on K12 database). Based on this ranking and a research of BL21 database and the band intensity quantitation based on Coomassie blue stained gel, the top abundant proteins that have estimated total amount over 10 ppm (ng/mg) were listed in Table 3, along with their pI, MW and in silico immunogenicity score and subjective risk.

### **In silico immunogenicity assessment of *E. coli* derived HCPs**

The immunogenicity of most abundant *E. coli* derived HCPs that were estimated to have an abundance levels of at least 10 ng/mg based on SDS-PAGE and LC-MS data were assessed using the in silico tools described earlier (Bailey-Kellogg et al.). The proteins where multiple high affinity epitopes (top 5%) and associated with a cluster were identified and shown in Table 3. A relative ranking of the epitopes on the immunogenicity scale is displayed in Figure 8. The sequence-based risk (subjective risk) of the most abundant six HCPs (levels estimated to be > 20 ng/mg) is considered low with few non-autoreactive Clustimers displaying relatively low Clustimer scores on the scale. A few other *E. coli* HCPs with a higher predicted risk as evident by higher number of high binding and high affinity eiptopes were present at much lower levels and hence considered low risk (de Zafra, Quarmby, Francissen, Vanderlaan, & Zhu-Shimoni; F. Q. Wang, Richardson, & Shameem, 2015). It should be noted that the residual proteins are relatively small with MW ranges from 7.4 kDa to 48.2 kDa, and the pI ranges from 5.17 to 7.24 (top 6). Although considered low risk, this information was supplied to process chemists to further develop a purification method to de-risk the potential impact from these residual proteins.

### **Residual protein analysis in workup APIs using organic extraction.**

As described in Materials and Methods, a workup procedure was developed to further remove residual proteins by using organic extraction and aqueous wash followed by a pH swing and aqueous back extraction (Figure 1b). The organic extraction efficiently extracts API to MeTHF using TGDE as a co-solvent and most residual proteins were removed in the rag layer and aqueous layer. The final API prepared from this procedure was analyzed using Cygnus kit ELISA, a silver-stained gel (Figure 9a), as well as a Coomassie blue stained gel (Figure 9b) with bands excising coupled with nano-LC-MS/MS (Table 4) again to analyze the residual protein level. No detectable amount of residual proteins was reactive to the Cygnus kit antibody and no proteins were detectable even at 20 mg/mL API load on a silver-stained SDS-PAGE gel (Figure 9a). Only five proteins were identified by LC-MS/MS from the bands excised from Coomassie blue-stained gel (Figure 9b), each with amount lower than 10 ng/mg. Similar results were obtained from three representative batches of APIs isolated by this new procedure (Table 4).

### **Conclusions**

Residual HCP control and their risk assessment has been an industry-wide challenge for biotechnology companies. This is mostly due to the complexity and heterogeneity of HCP makeup in the upstream and downstream process and the relative low abundance of HCP in final drug substance. The latter makes analyzing residual HCPs like finding needles in a haystack and thus requires highly sensitive and specific detection methods that can detect and distinguish HCPs from a dominant matrix of therapeutic proteins. A sandwich ELISA utilizing polyclonal antibodies that can recognize and capture residual HCPs is often used as the workhorse for residual HCP measurement. However, since the ELISA relies on the antibodies to detect and quantify residual HCP amount in samples, the ability of such antibodies to detect potential HCPs that can reside in the process intermediates and final drug substance need to be demonstrated and



the method validated for its precision, accuracy, linearity, range, LOQ and robustness (Gunawan et al., 2018). In bioprocess development, a generic or platform HCP ELISA that is commercially available or developed in-house is often used in the early phase development up to the stage of process validation with appropriate assay qualification to gain insight on the HCP clearance trend and batch consistency. From phase III and beyond, either a platform assay or upstream process specific assay is preferred to mitigate the risk of inadequate coverage of HCPs specific to the manufacturing process by a generic HCP ELISA (USP Monograph Chapter 1132). However, even with a well-validated process specific assay, chances are that not all residual proteins are quantified accurately given the difficulty to achieve 100% coverage and to find relevant standards to quantify residual HCPs in all process intermediates. More commonly, an upstream process mock HCP culture from a null cell line is used as the calibration standard for the ELISA assay. This often leads to quantitation error when certain HCPs are enriched during this process, especially when the amount of HCPs present is in excess of the antibodies available to capture and detect the HCPs (Zhu-Shimoni et al., 2014). As indicated in Figure 3, the use of different calibration standards can lead to significantly different measurement of protein values, despite being uncommon to see such a large extent of difference with well-qualified assay standard. Therefore, orthogonal methods are often needed to supplement the results obtained from ELISA testing to evaluate the overall risk of residual proteins while the ELISA method needs demonstrate its fitness for its intended purpose in early phase development and fully validated at late phase development to demonstrate its precision, accuracy, linearity, range, sensitivity, specificity, and robustness.

Unlike biologics, residual HCPs in small molecule APIs often have distinct biochemical properties and can be easily separated from API by SEC-HPLC or Tangential flow filtration (TFF). However, the use of TFF and column-based separation is not desired in small molecule process development. Process chemists tend to use phase cut and crystallization as the main means of isolation (Wells et al., 2012; Wells et al., 2016). In the case of MK-1454 discussed in this manuscript, an initial isolation of API using these traditional process chemistry techniques achieved high API yield and purity comparable to chemical synthesis. However, *E. coli* proteins along with the enzymes added to the process were not completely removed or polished, leaving a large pool of proteins present at trace levels in the Prep. Lab batch API. The amount of residual *E. coli* proteins were estimated using the commercial *E. coli* HCP kit while the enzymes used in the reaction are not reactive to the kit antibody (Figure 4b and 4c). Although efforts have been made and some success has been achieved to use the total input level of proteins (reactive or not) to estimate the residual amount of proteins in the API, quantification by this approach have the risk of over-estimating *E. coli* proteins if contamination occurs during the process. To overcome these challenges, 1D SDS-PAGE gel with silver stain and LC-MS was used to estimate the total protein amount in API and assess the risks associated with those proteins by their relative abundance level and in silico predicted immunogenicity. Although the proteins present in the Prep. Lab. batch API are not considered to pose significant immunogenicity risk, these materials haven't been assessed for immunotoxicity in animal studies or clinical trials. Instead, the chemically synthesized API was used in early clinical study and the biocatalytic route is developed for commercial chemistry. To minimize the potential immunogenicity risk and allow a direct use of biocatalytic route synthesized API in clinical trials, further reduction of residual proteins is achieved by process optimization. The new workup process has barely any detectable level of proteins as analyzed by ELISA, 1D SDS-PAGE with silver stain and proteomic LC-MS/MS. This case study demonstrates the importance of a holistic analytical control strategy in HCP characterization for biocatalytic route synthesized API. This holistic analytical control strategy allows process chemistry to design new commercial manufacturing process to remove residual proteins (HCP and enzymes) to insignificant levels (<10 ng/mg) in three representative batches of API. With a robust process and holistic analytical characterization, a process-specific ELISA using antibody reagents developed for matching cell lysate used for MK-1454 biocatalysis may not be needed in late phase development weighing in the time and resources investment in developing such an assay, the API comparability and the low demand in MK-1454 quantity in commercial manufacturing. However, the holistic analytical characterization presented here, together with the API stability monitoring, will be essential to reduce patient safety and product quality risk associated with the presence of residual *E. coli* HCPs and enzymes.

## Acknowledgement:

The authors thank Drs. Sachin Lohani and Peter Salmon for their critical review of this manuscript. The authors would also like to thank Dr. Feng Peng, Nelo Rivera, Steve Castro, Marc Poirier, John McIntosh, Taylor Stewart, and Stephanie Davis for their support in the biocatalytic process chemistry development, and analytical testing of MK-1454 API.

## References:

- Bailey-Kellogg, C., Gutiérrez, A. H., Moise, L., Terry, F., Martin, W. D., & De Groot, A. S. CHOPPI: A web tool for the analysis of immunogenicity risk from host cell proteins in CHO-based protein production. *Biotechnology and Bioengineering*, *111* (11), 2170-2182.
- Bee, J. S., Tie, L., Johnson, D., Dimitrova, M. N., Jusino, K. C., & Afdahl, C. D. Trace levels of the CHO host cell protease cathepsin D caused particle formation in a monoclonal antibody product. *Biotechnology Progress*, *31* (5), 1360-1369.
- Cheung, P., Emanuel, A., Heward, J., Maddalena, J., Toth, D., Saward, L., & Kodihalli, S. (2016). Reduced immunogenic response to residual CHO cell protein in recombinant factor IX (IB1001) drug product in normal healthy rabbits. *Haemophilia*, *22* (3), E220-E222.
- Chiu, J., Valente, K. N., Levy, N. E., Min, L., Lenhoff, A. M., & Lee, K. H. Knockout of a difficult-to-remove CHO host cell protein, lipoprotein lipase, for improved polysorbate stability in monoclonal antibody formulations. *Biotechnology and Bioengineering*, *114* (5), 1006-1015.
- Clarke, C., Gallagher, C., Kelly, R. M., Henry, M., Meleady, P., Frye, C. C., . . . Clynes, M. (2019). Transcriptomic analysis of IgG4 Fc-fusion protein degradation in a panel of clonally-derived CHO cell lines using RNASeq. *116* (6), 1556-1562. doi:10.1002/bit.26958
- Congdon, R. W., Muth, G. W., & Splittgerber, A. G. (1993). The Binding Interaction of Coomassie Blue with Proteins. *Anal Biochem*, *213* (2), 407-413.
- de Maria, P. D., & Hollmann, F. (2015). On the (Un)greenness of Biocatalysis: Some Challenging Figures and Some Promising Options. *Frontiers in Microbiology*, *6* .
- de Zafra, C. L., Quarmby, V., Francissen, K., Vanderlaan, M., & Zhu-Shimoni, J. Host cell proteins in biotechnology-derived products: A risk assessment framework. *Biotechnology and Bioengineering*, *112* (11), 2284-2291.
- Dixit, N., Salamat-Miller, N., Salinas, P. A., Taylor, K. D., & Basu, S. K. Residual Host Cell Protein Promotes Polysorbate 20 Degradation in a Sulfatase Drug Product Leading to Free Fatty Acid Particles. *J Pharm Sci*, *105* (5), 1657-1666.
- The European Agency for the Evaluation of Medicinal Products Human Medicines Evaluation Unit (1997). *CPMP/BWP/382/97* .
- FDA. (1997). Points to Consider in the Manufacture & Testing of Monoclonal Products for Human Use. In.
- Fischer, S. K., Cheu, M., Peng, K., Lowe, J., Araujo, J., Murray, E., . . . Song, A. Specific Immune Response to Phospholipase B-Like 2 Protein, a Host Cell Impurity in Lebrikizumab Clinical Material. *Aaps J*, *19* (1), 254-263.
- Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F., & Smolke, C. D. (2015). SYNTHETIC BIOLOGY Complete biosynthesis of opioids in yeast. *Science*, *349* (6252), 1095-1100.
- Goey, C. H., Bell, D., & Kontoravdi, C. (2018). Mild hypothermic culture conditions affect residual host cell protein composition post-Protein A chromatography. *MAbs*, *10* (3), 476-487.
- Guideline, I. (1999). Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products Q6B. *International Conference on Harmonisation of Technical Requirement for Registration of Pharmaceuticals for Human Use* .

- Gunawan, F., Nishihara, J., Liu, P., Sandoval, W., Vanderlaan, M., Zhang, H. D., & Krawitz, D. (2018). Comparison of platform host cell protein ELISA to process-specific host cell protein ELISA. *Biotechnology and Bioengineering*, 115 (2), 382-389.
- Hall, T., Sandefur, S. L., Frye, C. C., Tuley, T. L., & Huang, L. Polysorbates 20 and 80 Degradation by Group XV Lysosomal Phospholipase A2 Isomer X1 in Monoclonal Antibody Formulations. *J Pharm Sci*, 105 (5), 1633-1642.
- Hanania, N. A., Noonan, M., Corren, J., Korenblat, P., Zheng, Y. A., Fischer, S. K., . . . Yen, K. (2015). Lebrikizumab in moderate-to-severe asthma: pooled data from two randomised placebo-controlled studies. *Thorax*, 70 (8), 748-756.
- Hogwood, C. E., Bracewell, D. G., & Smales, C. M. Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Curr Opin Biotechnol*, 30 , 153-160.
- Huffman, M. A., Fryszkowska, A., Alvizo, O., Borra-Garske, M., Campos, K. R., Canada, K. A., . . . Yang, H. (2019). Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science*, 366 (6470), 1255-1259. doi:10.1126/science.aay8484
- Jawa, V., Joubert, M. K., Zhang, Q., Deshpande, M., Hapuarachchi, S., Hall, M. P., & Flynn, G. C. Evaluating Immunogenicity Risk Due to Host Cell Protein Impurities in Antibody-Based Biotherapeutics. *Aaps J*, 18 (6), 1439-1452.
- Jiskoot, W., Rispens, T., & Kijanka, G. (2019). Immunogenicity of Therapeutic Proteins. In (pp. 139-150).
- John A. McIntosh<sup>1\*</sup>, Z. L., Brian M. Andresen<sup>2\*</sup>, Nastaran Salehi Marzijarani<sup>1</sup>, Jeffrey C. Moore<sup>1</sup>, Nicholas M. Marshall<sup>1</sup>, Margie Borra-Garske<sup>3</sup>, Jennifer V. Obligacion<sup>1</sup>, Patrick S. Fier<sup>1</sup>, Feng Peng<sup>1</sup>, Jacob H. Forstater<sup>1</sup>, Matthew S. Winston<sup>1</sup>, Chihui An<sup>1</sup>, Wonsuk Chang<sup>2</sup>, Jongwon Lim<sup>2</sup>, Mark A. Huffman<sup>1</sup>, Steven P. Miller<sup>1</sup>, Fuh-Rong Tsay<sup>1</sup>, Michael D. Altman<sup>2</sup>, Charles A. Lesburg<sup>2</sup>, Dietrich Steinhuebel<sup>1</sup>, B. Wesley Trotter<sup>2</sup>, Jared N. Cumming<sup>2</sup>, Alan Northrup<sup>2</sup>, Xiaodong Bu<sup>1</sup>, Benjamin F. Mann<sup>1</sup>, Mirinda Biba<sup>1</sup>, Kaori Hiraga<sup>1</sup>, Grant S. Murphy<sup>1</sup>, Joshua N. Kolev<sup>1</sup>, Amanda Makarewicz<sup>1</sup>, Weilan Pan<sup>1</sup>, Iman Farasat<sup>1</sup>, Rachel S. Bade<sup>1</sup>, Kevin Stone<sup>1</sup>, Duan Da<sup>3</sup>, Oscar Alviso<sup>3</sup>, Donovan Adpressa<sup>2</sup>, Erik Guetschow<sup>1</sup>, Erik Hoyt<sup>1</sup>, Erik L. Regalado<sup>1</sup>, Steve Castro<sup>1</sup>, Nelo Rivera<sup>1</sup>, Joseph P. Smith<sup>1</sup>, Fengqiang Wang<sup>1</sup>, Alejandro Crespo<sup>1</sup>, Deeptak Verma<sup>1</sup>, Stephanus Axnanda<sup>1</sup>, Zachary Dance<sup>1</sup>, Paul N. Devine<sup>1</sup>, David Tschaen<sup>1</sup>, Keith A. Canada<sup>1</sup>, Paul G. Bulger<sup>1</sup>, Benjamin D. Sherry<sup>1</sup>, Matthew D. Truppo<sup>1</sup>, Rebecca T. Ruck<sup>1</sup>, Louis-Charles Campeau<sup>1</sup>, David Jonathan Bennett<sup>2</sup>, Guy R. Humphrey<sup>1</sup>, Kevin R. Campos<sup>1</sup>, Matthew L. Maddess<sup>1</sup>. (2022). A kinase-cGAS cascade to synthesize a therapeutic STING activator *Nature*
- Jones, M., Palackal, N., Wang, F., Gaza-Bulsecos, G., Hurkmans, K., Zhao, Y., . . . Connolly, T. (2021). "High-risk" host cell proteins (HCPs): A multi-company collaborative view. *Biotechnology and Bioengineering*, 118 (8), 2870-2885. doi:https://doi.org/10.1002/bit.27808
- Lama, L., Adura, C., Xie, W., Tomita, D., Kamei, T., Kuryavyy, V., . . . Tuschl, T. (2019). Development of human cGAS-specific small-molecule inhibitors for repression of dsDNA-triggered interferon expression. *Nature Communications*, 10 (1), 2261. doi:10.1038/s41467-019-08620-4
- Luo, H. B., Tie, L., Cao, M. Y., Hunter, A. K., Pabst, T. M., Du, J. L., . . . Wang, W. K. (2019). Cathepsin L Causes Proteolytic Cleavage of Chinese-Hamster-Ovary Cell Expressed Proteins During Processing and Storage: Identification, Characterization, and Mitigation. *Biotechnology Progress*, 35 (1).
- Novotna, B., Vanekova, L., Zavrel, M., Budesinsky, M., Dejmek, M., Smola, M., . . . Birkus, G. (2019). Enzymatic Preparation of 2'-5',3'-5'-Cyclic Dinucleotides, Their Binding Properties to Stimulator of Interferon Genes Adaptor Protein, and Structure/Activity Correlations. *J Med Chem*, 62 (23), 10676-10690. doi:10.1021/acs.jmedchem.9b01062
- Patel, R. N. (2006). Biocatalysis: Synthesis of chiral intermediates for drugs. *Current Opinion in Drug Discovery & Development*, 9 (6), 741-764.

- Paulo, J. A. (2016). Sample preparation for proteomic analysis using a GeLC-MS/MS strategy. *Journal of biological methods*, 3 (3), e45. doi:10.14440/jbm.2016.106
- Reetz, M. T. (2013). Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *Journal of the American Chemical Society*, 135 (34), 12480-12496.
- Reijers, J. A. A., Malone, K. E., Bajramovic, J. J., Verbeek, R., Burggraaf, J., & Moerland, M. (2019). Adverse immunostimulation caused by impurities: The dark side of biopharmaceuticals. *British Journal of Clinical Pharmacology*, 85 (7), 1418-1426.
- Romer, T., Peter, F., Saenger, P., Starzyk, J., Koehler, B., Korman, E., . . . Berghout, A. J. J. o. E. I. (2007). Efficacy and safety of a new ready-to-use recombinant human growth hormone solution. 30 (7), 578-589. doi:10.1007/bf03346352
- Saenger, P. (2009). Current status of biosimilar growth hormone. *International journal of pediatric endocrinology*, 2009 , 370329-370329. doi:10.1155/2009/370329
- Smith, J. P., Ralbovsky, N. M., Lauro, M. L., Hoyt, E., Guetschow, E. D., Wang, F., . . . Bu, X. (2022). Quantitation and speciation of residual protein within active pharmaceutical ingredients using image analysis with SDS-PAGE. *J Pharm Biomed Anal*, 207 , 114393. doi:10.1016/j.jpba.2021.114393
- Tal, M., Silberstein, A., & Nusser, E. (1985). Why Does Coomassie Brilliant Blue-R Interact Differently with Different Proteins - a Partial Answer. *Journal of Biological Chemistry*, 260 (18), 9976-9980.
- Vanderlaan, M., Zhu-Shimoni, J., Lin, S. S., Gunawan, F., Waerner, T., & Van Cott, K. E. (2018). Experience with Host Cell Protein Impurities in Biopharmaceuticals. *Biotechnology Progress*, 34 (4), 828-837.
- Wang, F. Q., Richardson, D., & Shameem, M. (2015). Host-Cell Protein Measurement and Control. *Biopharm International*, 28 (6), 32-+.
- Wang, X., Hunter, A. K., & Mozier, N. M. (2009). Host cell proteins in biologics development: Identification, quantitation and risk assessment. *Biotechnology and Bioengineering*, 103 (3), 446-458.
- Weiss W., W. F., Görg A. (2009). Protein Detection and Quantitation Technologies for Gel-Based Proteome Analysis. In S. A. Reinders J. (Ed.), *Proteomics. Methods in Molecular Biology (Methods and Protocols)* . [https://doi.org/10.1007/978-1-60761-157-8\\_4](https://doi.org/10.1007/978-1-60761-157-8_4): Humana Press
- Wells, A. S., Finch, G. L., Michels, P. C., & Wong, J. W. (2012). Use of Enzymes in the Manufacture of Active Pharmaceutical Ingredients-A Science and Safety-Based Approach To Ensure Patient Safety and Drug Quality. *Organic Process Research & Development*, 16 (12), 1986-1993.
- Wells, A. S., Wong, J. W., Michels, P. C., Entwistle, D. A., Fandrick, K., Finch, G. L., . . . Watson, T. J. (2016). Case Studies Illustrating a Science and Risk-Based Approach to Ensuring Drug Quality When Using Enzymes in the Manufacture of Active Pharmaceuticals Ingredients for Oral Dosage Form. *Organic Process Research & Development*, 20 (3), 594-601.
- Woodley, J. M. (2008). New opportunities for biocatalysis: making pharmaceutical processes greener. *Trends in Biotechnology*, 26 (6), 321-327.
- Zhou, W., Whiteley, A. T., Mann, C. C. D., Morehouse, B. R., Nowak, R. P., Fischer, E. S., . . . Kranzusch, P. J. (2018). Structure of the Human cGAS-DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell*, 174 (2), 300-+.
- Zhu-Shimoni, J., Yu, C., Nishihara, J., Wong, R. M., Gunawan, F., Lin, M., . . . Vanderlaan, M. (2014). Host Cell Protein Testing by ELISAs and the Use of Orthogonal Methods. *Biotechnology and Bioengineering*, 111 (12), 2367-2379.

## Figure legends:

**Figure 1.** The initial (a) and improved (b) workflow for MK-1454 biocatalysis and API isolation to remove residual proteins.

**Figure 2.** Standard curves plotted using the ELISA signal generated from kit standard (green curve) and an in-house lysate supernatant standard (red curve) against their respective standard concentrations. The A, B, C, D values of the curve are listed on the right.

**Figure 3.** The dilutional non-linearity of prep-lab batch MK-1454 and the differences in HCP (ng/mg) measurement by using different standards. a., the ng/mL protein concentration (y-axis) measured by using an in-house standard (red curve) and Cygnus kit standard (green curve) at different concentrations of MK-1454 API (mg/mL, x-axis); b., The 4-parameter fit binding curve of plotted using the mean absorbance at 450 nm (y-axis) against Cygnus kit standard (ng/mL, green curve), in-house lysate standard (ng/mL, red curve), and MK-1454 API (mg/mL, blue curve) at various concentrations on x-axis.; c., the relative protein concentration (ng/mg, y-axis) measured by using an in-house standard (red curve) and Cygnus kit standard (green curve) respectively at different concentrations (mg/mL) of MK-1454 API.

**Figure 4.** SDS-PAGE comparing the protein profile between Cygnus kit standard and in-house lysate supernatant standard and their immunoreactivity to the kit antibody. a, SDS-PAGE gel with different concentrations of each standard and Coomassie blue staining. b, Western blot of a membrane transferred from a duplicate SDS-PAGE gel showing the immunoreactivity of Cygnus kit antibody to the proteins in each standard. c, Immunoreactivity detected by Cygnus kit ELISA on enzymes A, B, and C, which were used in the early steps of MK-1454 biocatalysis prior to the last reaction step. d, Immunoreactivity of proteins detected in Prep. Lab, API to Cygnus kit antibody on Western blot.

**Figure 5.** SDS-PAGE resolved proteins with Coomassie blue staining (a) or silver staining (b) in different concentrations of Prep Lab API loaded together with different amount of BSA standard.

**Figure 6.** SDS-PAGE gel loaded with 30 mg/mL and stained with Coomassie blue showing the 6 major bands detected and excised for in-gel digestion and LC-MS.

**Figure 7.** Proteins identified in each band of SDS-PAGE gel resolved proteins by LC-MS/MS.

**Figure 8.** EpiMatrix ClustiMer immunogenicity scale. Clustimers derived from residual proteins in API vary in their calculated immunogenicity risk. Clustimers with low human homology and their scores were identified and calculated in the EpiMatrix program. Clustimers were binned according to their score and are displayed on the right side of the scale. Number in parenthesis indicates multiple peptides from a given protein are included in the range. Foreign proteins and human CLIP are shown on the left side of the scale to give relative risk. Immunogenicity risk increases from green (minimal) to red (high). Abbreviations: HCV= Hepatitis C virus, CLIP= Class II-associated invariant chain peptide. EBV=Epstein-Barr virus. TRPCIF=Transcriptional Regulatory Protein CpxR IcIR Family, AHRC= Alkyl Hydroperoxide Reductase C, 30SRPS10= 30S Ribosomal Protein S10, CspC= Cold Shock-like Protein C.

**Figure 9.** MK-1454 API isolated and purified with the new process resolved on SDS-PAGE with silver stain (a), Coomassie blue stain (b).BSA, bovine serum albumin, MK-1454 157, 165, 166 are three small scale biocatalytic API isolated with the new workup procedure. C-10 mg/mL, represents a chemically synthesized MK-1454 control that is free of protein impurities.

**Figure 1.**

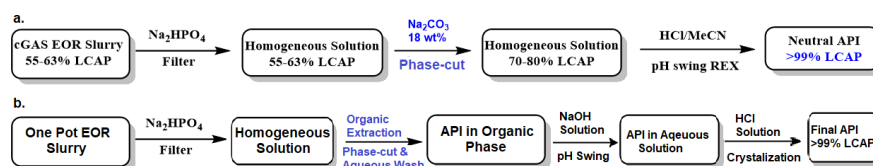


Figure 2.

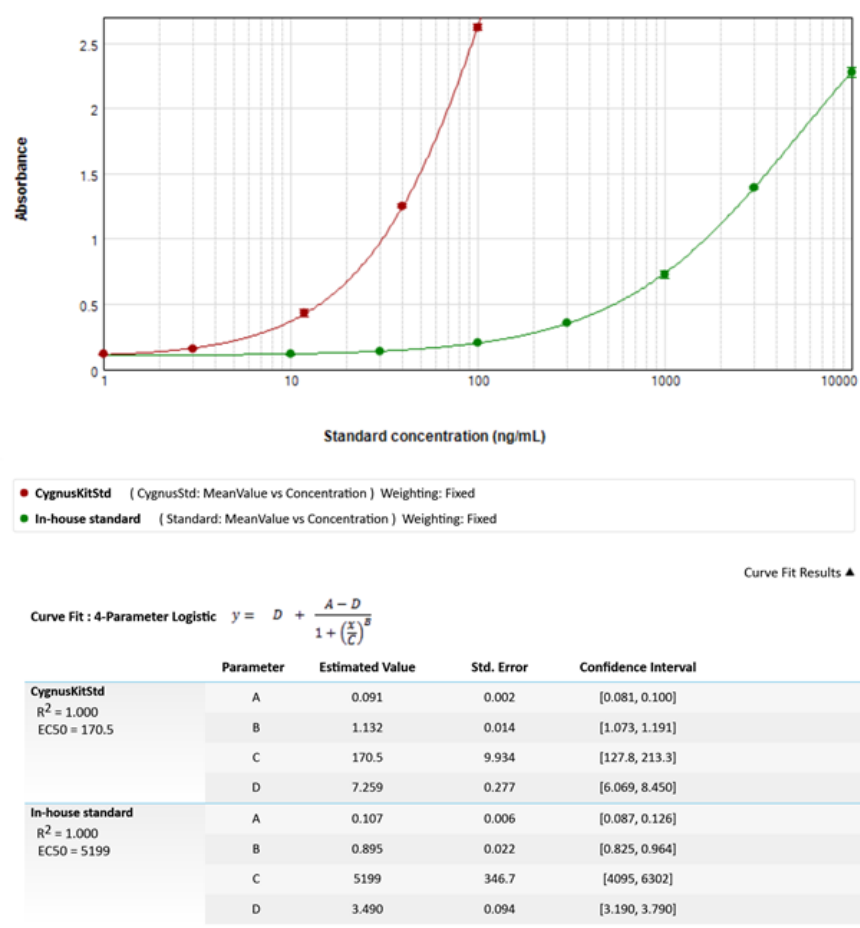


Figure 3.

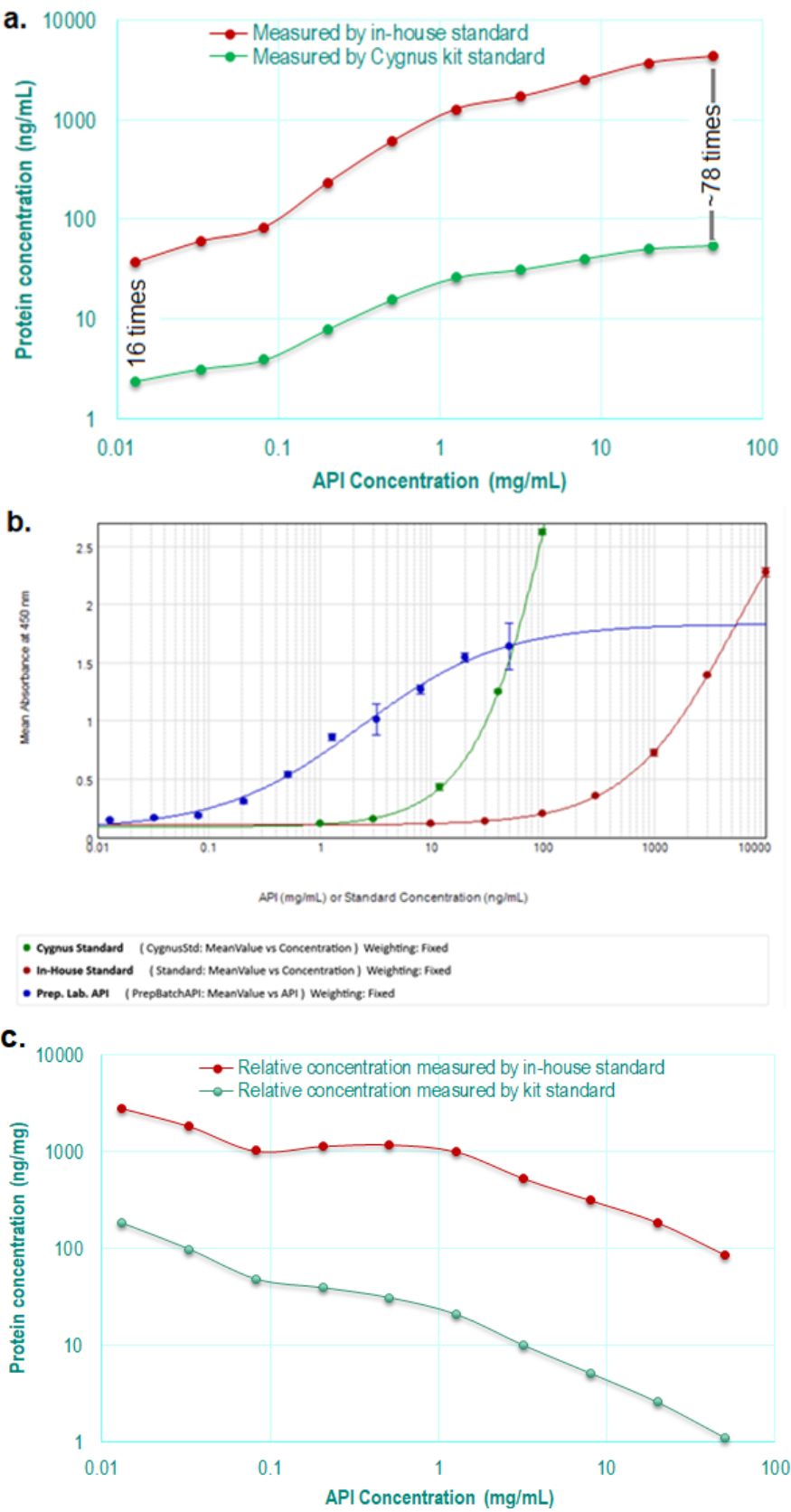


Figure 4.

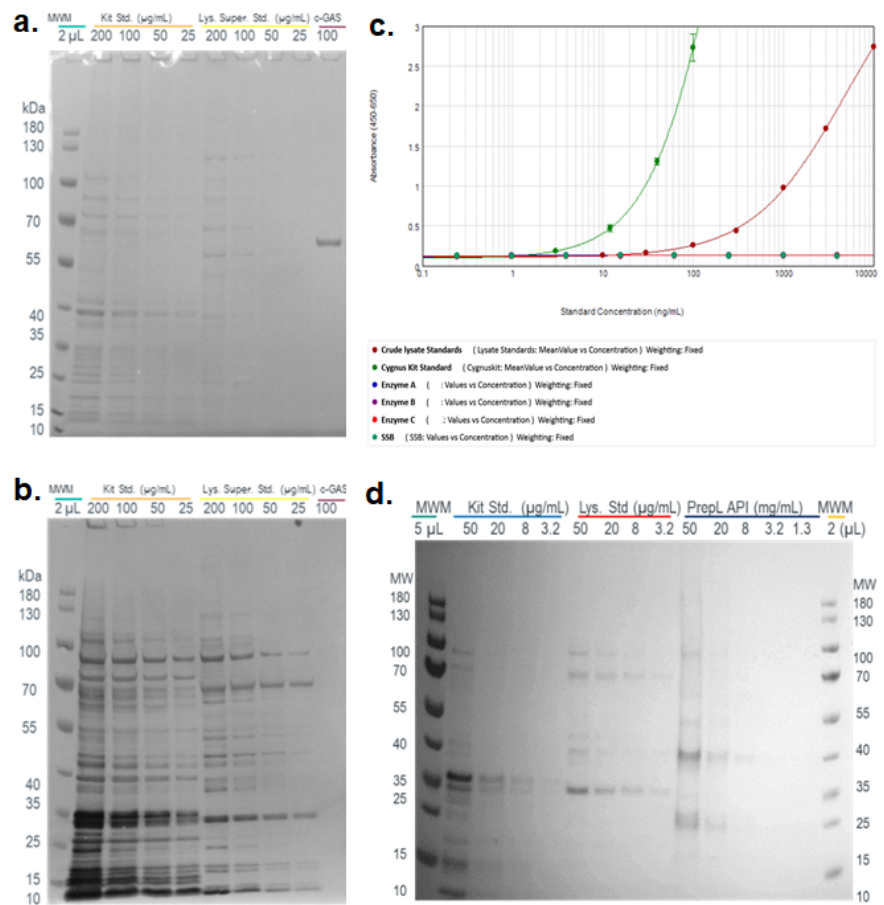


Figure 5

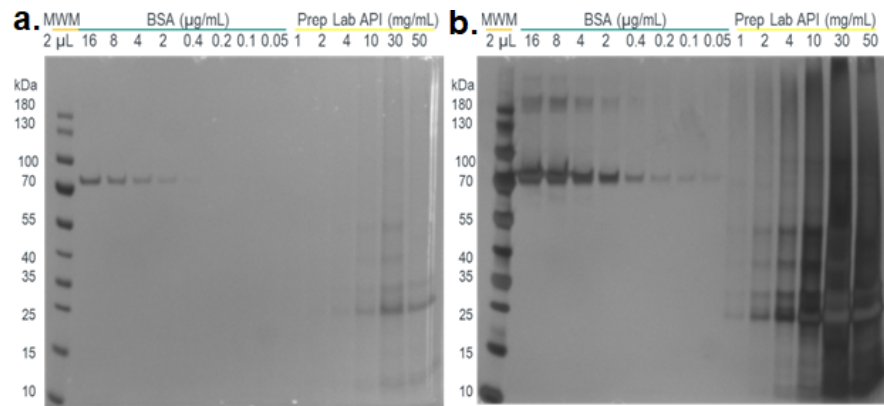


Figure 6.



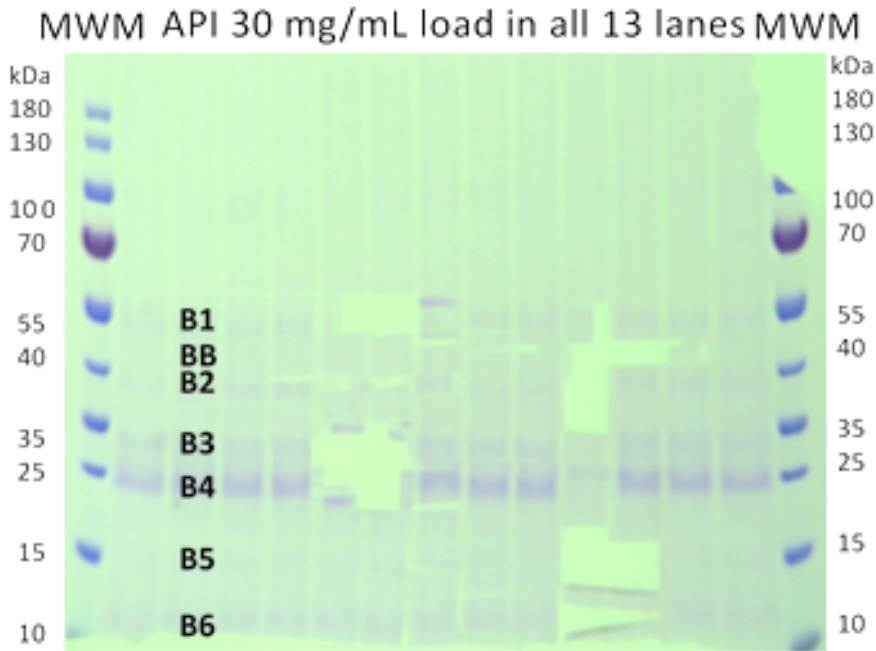


Figure 7.

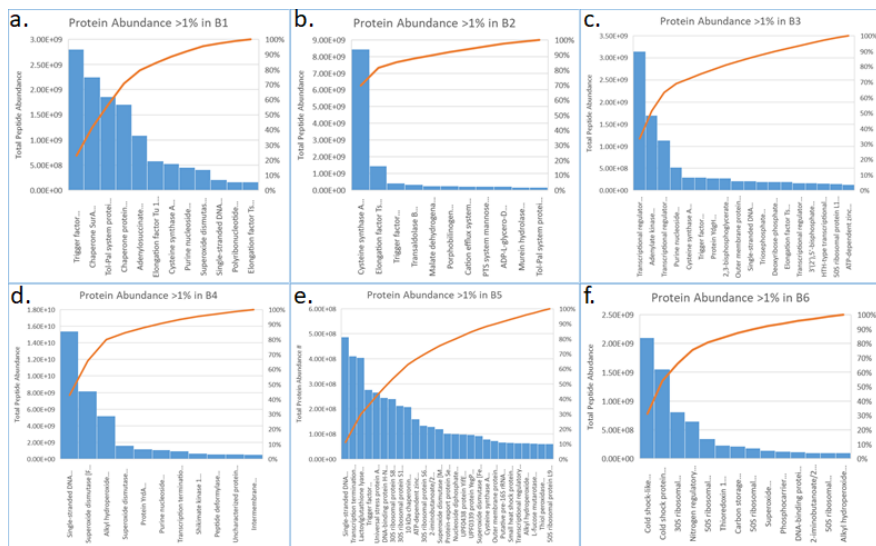


Figure 8

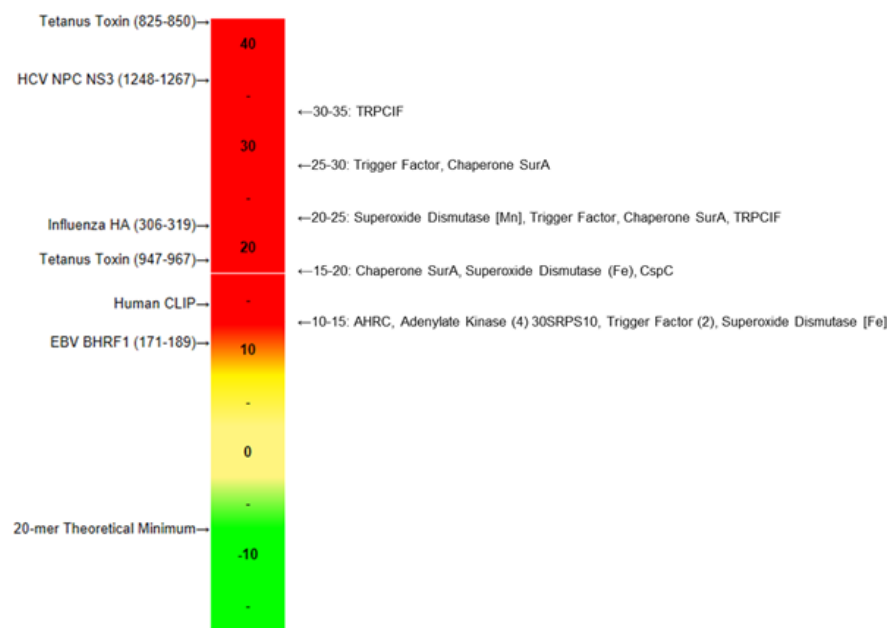


Figure 9.

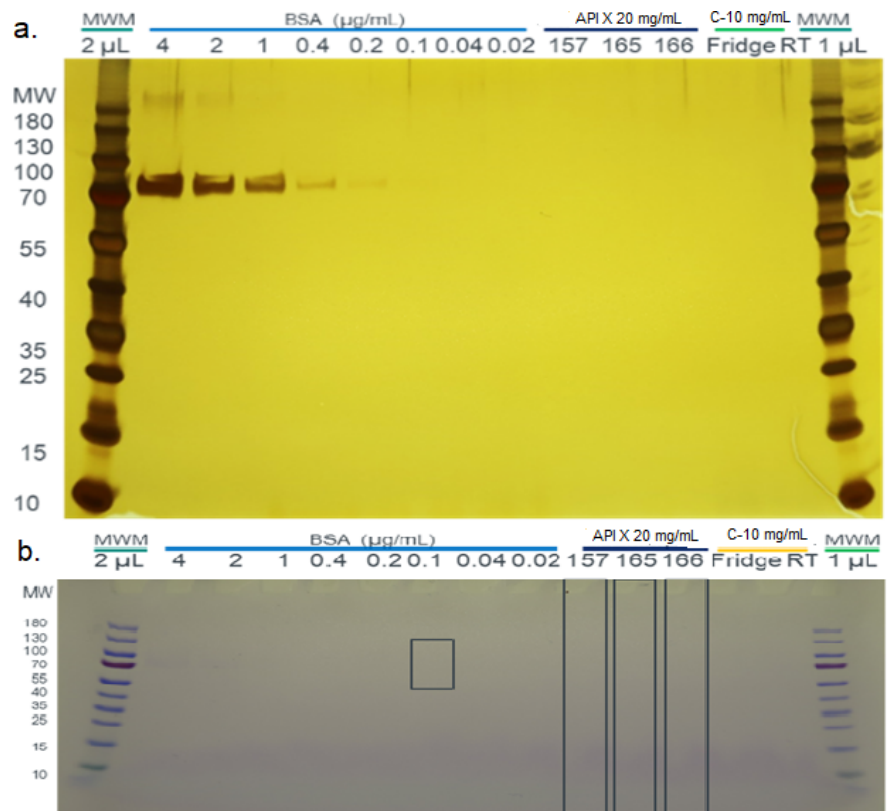


Table 1. The immune-equivalent amount of in-house lysate standard measured by Cygnus *E. coli* HCP ELISA kit as compared to the protein amount measured by BCA assay.

In-house standard concentration by BCA assay (ng/mL)	Mean Result tested by Cygnus kit (ng/mL) (n=3)
10,000	82.53
3,000	45.20
1,000	21.86
300	9.64
100	4.52
30	2.07
10	1.15

**Table 2.** Densitometry analysis of residual protein amount in MK-1454 Prep. Lab API loaded at different concentrations with BSA as a loading standard on duplicate SDS-PAGE gels.

Lane #	API loading concentration (mg/mL)	Residual Protein (ng/mg)	Residual Protein (ng/mg)	Mean (ng/mg)
		<b>Gel 1</b>	<b>Gel 2</b>	
10	1	<400	<400	<400
11	2	<200	<200	<200
12	4	1000	242	621
13	10	1036	732	884
14	30	662	913	788
Mean (12-14)				<b>764</b>
% RSD				<b>17%</b>

**Table 3.** Top 13 *E. coli* proteins estimated to be above 10 ppm among all protein identified by LC-MS/MS in Prep. Lab API and their predictive immunogenicity.

Protein Name	MW	pI	Estimated Amount (ng/mg)	Number of Clus-tiMers*	ClustiMer Score	Subjective Risk
Single-stranded DNA binding protein	19	5.58	~103	0	0	Low
Superoxide Dismutase (Fe)	21.3	5.95	~64	2	17.09; 14.81	Low
Cysteine Synthase	34.5	6.06	~40	0	0	Low
Alkyl hydroperoxide reductase C	20.7	5.17	~35	1	10.39	Low
Cold shock-like protein CspC	7.4	7.24	~34	1	17.21	Low

Protein Name	MW	pI	Estimated Amount (ng/mg)	Number of Clus-tiMers*	ClustiMer Score	Subjective Risk
Cold shock-like protein CspA	7.4	5.95	~25	0	0	Low
Transcriptional 30 regulatory protein CpxR		5.58	~19	3	34.98; 21.24; 17.95	High
IcIR family 30S ribosomal protein S10	11.7	9.69	~13	1	12.57	Low
Trigger Factor	48.2	4.88	~13	4	25.01; 21.6; 12.76; 12.65	High
Superoxide dismutase [Mn]	23.1	6.96	~11	1	20.68	High
Nitrogen regulatory protein P-II	12.4	5.34	~11	0	0	Low
Adenylate Kinase	23.6	5.76	~10	4	14.54; 11.93; 10.8; 10.02	Low
Chaperone SurA	47.3	6.98	~10	3	27.12; 22.97; 15.19	High

**Table 4. Residual proteins identified in MK-1454 API isolated with the new process as indicated in Figure 1b.**

Protein identified	# Unique peptides	MW (kDa)	p
Bifunctional aspartokinase/homoserine dehydrogenase	12	88.9	5
Aspartokinase	4	48.5	5
Bifunctional glutamine synthetase adenylyltransferase/adenylyl-removing enzyme	4	108.3	5
4-hydroxy-tetrahydrodipicolinate reductase	4	28.7	5
3-dehydroquinate synthase	2	38.8	6
Bovine serum albumin (BSA)	16	69.2	6

\*, the estimate is based on the total peak area of detectable peptides in the identified proteins as compared to the total peak area of detectable peptides in the 0.1 µg/mL load of BSA standard on the Coomassie blue stained gel (molar ratio). BSA has a total peak area of  $3.93 \times 10^7$ , the molar ratio was then converted to mass ratio (ng/mg) based on the MW difference. The detection of 0.1 µg/mL of BSA in 20 mg/mL MK-1454

API is equivalent to 5 ng/mg API.