nature research

Corresponding author(s): J. Christopher Love

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

The experiments presented in this work were representative of greater than 35 independent operational replicates performed under similar conditions to determine optimal process parameters for production and their impact on quality attributes of proteins. Sample sizes were not predetermined. Instead, replicate experiments (min n=3) were initially performed for each condition or process. Variations or quality deviations were noted. Process conditions were adjusted accordingly and experiments were performed again in replicate to confirm new conditions. Iterations were performed until variation and quality was acceptable. Sample size for non-clinical studies was selected to be the minimum number of animals required to obtain statistically significant results.

2. Data exclusions

Describe any data exclusions.

Technical outliers were excluded from the cell-based potency assays for biologic activity of hGH (as determined by a third-party CRO Bioassay GmbH based on their Quality Assurance protocols for data generated in these assays and described in Methods). No data were excluded from analysis in other experiments reported.

3. Replication

Describe whether the experimental findings were reliably reproduced.

The production and characterization of G-CSF, IFNalpha-2b, and hGH presented in this work were representative of more than 35 independent operational replicates performed with similar parameters to map optimal production conditions. Quality attributes of materials were consistent with expectations (for example, higher dissolved oxygen led to higher levels of oxidation of protein). Materials produced in this work were regularly generated in triplicate using three independent automated production systems.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

For the non-clinical studies presented in this work randomization was performed to allocate animals into experimental groups using Research Randomizer software, version 4.0.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

For the non-clinical studies presented in this work blinding was not performed. The study employed methodology to minimize uncertainty and to control bias for data collection and analysis, however, which included but was not limited to: concurrent control data, system suitability assessment, randomization, and method controls such as blanks and replicates.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Wonderware 2014 R2 was used as a human-machine interface to the integrated system and as a data historian. GraphPad Prism v7 was used to plot and analyze all data in the manuscript and supplemental material, except for the following. MATLAB 2017a was used for plotting and analysis of the fermentograms and UV traces presented. LightCycler software release 1.5.0SP4 was used for qPCR analysis. Empower 3 was used for control in chromatographic analyses. Thermo BioPharmaFinder 2.0 was used for analysis of LCMS data. Provantis 9.3.1 and WinNonlin 6.3 was used for analysis in the non clinical studies.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Custom materials used for operation of the InSCyT system are available from the authors or from the companies noted in the methods section.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

The antibodies used for the G-CSF ELISA include Biolegend BVD13-3A5 (Cat# 502102, Lot# B129744), Biolegend BVD11-37G10 (Cat# 506702, Lot# B191907), and Abcam Streptavidin-HRP ab7403 (Lot# GR305788-2). The antibodies used for the IFNalpha-2b ELISA include AssayPro 31168-05121 (Lot# IB071910405) and Abcam Streptavidin-HRP ab7403 (Lot# GR305788-2). The antibodies used for the G-CSF ELISA in the PK studies were from the commercial Quantikine kit (R&D Systems Cat# DCS50). For Biolegend products, each lot of antibody was quality validated by ELISA using recombinant G-CSF. For Abcam products, each lot was validated against Biotinylated IgG in a standard capture ELISA using a peroxidase substrate. For AssayPro products, each lot of antibody was validated by ELISA using biotinylated recombinant IFNaplha-2b. The Quantikine kit was validated by ELISA using recombinant human G-CSF.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Komagataella phaffii NRRL Y-11430 was obtained from ATCC (Catalog number 76273).

b. Describe the method of cell line authentication used.

K. phaffii NRRL Y-11430 and derivatives described in the methods were authenticated by genome sequencing as reported in Love, K. R. et al. BMC Genomics (2016) 17:550.

c. Report whether the cell lines were tested for mycoplasma contamination.

Only yeast were used in this study and all cultivations used animal-free components; mycoplasma contamination/testing is not applicable and was not performed.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Thirty-nine Sprague Dawley rats (male) were used in PK/PD studies. All rats were at least 5 weeks old and at least 200g. Thirty Sprague Dawley rats (15 male and 15 female) were used in the repeated-dose study. Weights ranged from 232.7 - 328.0g. All females were non-pregnant and nulliparous.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No human research participants were involved in this study.