TU Delft

Delft University of Technology

Characterisation of the E. coli HMS174 and BLR host cell proteome to guide purification process development

Disela, Roxana; Bussy, Olivier Le; Geldhof, Geoffroy; Pabst, Martin; Ottens, Marcel

DOI
10.1002/biot.202300068

Publication date
2023

Document Version
Final published version

Published in
Biotechnology Journal

Citation (APA)
Disela, R., Bussy, O. L., Geldhof, G., Pabst, M., & Ottens, M. (2023). Characterisation of the E. coli HMS174 and BLR host cell proteome to guide purification process development. *Biotechnology Journal*, *18*(9), Article 2300068. https://doi.org/10.1002/biot.202300068

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Received: 12 February 2023 | Revised: 11 May 2023 | Accepted: 17 May 2023

DOI: 10.1002/biot.202300068

Biotechnology Journal

BIOTECH METHOD

Characterisation of the *E. coli* HMS174 and BLR host cell proteome to guide purification process development

Roxana Disela[1] | **Olivier Le Bussy**[2] | **Geoffroy Geldhof**[2] | **Martin Pabst**[1] | **Marcel Ottens**[1]

[1]Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

[2]GSK, Technical Research & Development, Rixensart, Belgium

Correspondence
Martin Pabst, Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands.
Email: m.pabst@tudelft.nl

Funding information
GlaxoSmithKline Biologicals S.A

Abstract

Mass-spectrometry-based proteomics is increasingly employed to monitor purification processes or to detect critical host cell proteins in the final drug substance. This approach is inherently unbiased and can be used to identify individual host cell proteins without prior knowledge. In process development for the purification of new biopharmaceuticals, such as protein subunit vaccines, a broader knowledge of the host cell proteome could promote a more rational process design. Proteomics can establish qualitative and quantitative information on the complete host cell proteome before purification (i.e., protein abundances and physicochemical properties). Such information allows for a more rational design of the purification strategy and accelerates purification process development. In this study, we present an extensive proteomic characterisation of two *E. coli* host cell strains widely employed in academia and industry to produce therapeutic proteins, BLR and HMS174. The established database contains the observed abundance of each identified protein, information relating to their hydrophobicity, the isoelectric point, molecular weight, and toxicity. These physicochemical properties were plotted on proteome property maps to showcase the selection of suitable purification strategies. Furthermore, sequence alignment allowed integration of subunit information and occurrences of post-translational modifications from the well-studied *E. coli* K12 strain.

KEYWORDS
bioprocess engineering, chromatography, e. coli, host cell proteomics, proteomics

Abbreviations: ABC, ammonium bicarbonate; AGC, automatic gain control; BLR, *E. coli* strain BLR(DE3); BSA, bovine serum albumin; CHO, Chinese hamster ovary; DDA, data-dependent acquisition; DTT, dithiothreitol; ELISA, enzyme-linked immunosorbent assays; emPAI, exponentially modified protein abundance index; FDR, false discovery rate; GRAVY, grand average of hydropathy; HCP, host cell protein; HIC, hydrophobic interaction chromatography; HMS, *E. coli* strain HMS174(DE3); IAA, iodoacetamide; IEX, ion-exchange chromatography; IT, injection time; IPTG, Isopropyl β-D-1-thiogalactopyranoside; mAb, monoclonal antibody; MW, molecular weight; NCE, normalized collision energy; PAI, protein abundance index; pI, isoelectric point; PTM, post-translational modification; TCA, trichloroacetic acid; TFA, trifluoroacetic acid; QSPR, quantitative structure–property relationship.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Biotechnology Journal* published by Wiley-VCH GmbH.

1 | INTRODUCTION

Throughout the history of biopharmaceutical production, the effective removal and detection of host cell protein (HCP) impurities from the final drug product have been the subject of intensive research and development.[1–3] The presence of such impurities can have adverse effects on patient safety or product stability when present in the final drug product. For example, significant amounts of HCP impurities could be linked to strong side effects of the recent ChAdOx1 nCoV-19 vaccine.[4] To minimise impacts on patients and improve product quality, the effective removal of such impurities is of utmost importance.[3] At the same time, the pressure to accelerate the process development of biopharmaceuticals, especially the downstream processing,[5–7] is high. Vaccines in particular require accelerated development to ensure timely responses to emerging pandemics, which has only recently become evident with the COVID-19 outbreak and pandemic.

Impurities can originate from the process or the product itself (e.g., the degraded or aggregated form of the product). Process-related impurities originate from the host cell expression system used to produce the protein therapeutic. When host cells are disrupted to obtain the intracellular or periplasmic products, impurities from the host such as HCPs, DNA, RNA, and endotoxins are released. Therefore, extensive purification must be performed, where, the HCP content is reduced in every purification step until the target quality is reached (Figure 1). The structural and physicochemical properties of HCPs may closely resemble those of the protein therapeutic produced such that the elimination of such HCPs poses a significant challenge and is, therefore, the subject of extensive analytical development.

The acceptable levels of HCPs in vaccines are defined on a case-by-case basis by regulatory authorities.[8] For example, Zhu et al.[9] investigated a malaria vaccine candidate expressed in *E. coli*. The total HCP concentration was specified to be 90 ng or <1100 ppm per dose in this case.[10] Tolerated HCP levels for vaccines are generally higher compared to those of drugs for chronic diseases (<100 ppm).[8]

Jones et al. identified high-risk, immunogenic, biologically active, or enzymatically active HCPs, which showed the potential to degrade either the product molecules or the excipients in the formulation.[2] Using this knowledge, Chiu et al. furthermore knocked out genes from CHO cells to prevent the expression of high-risk and difficult-to-remove HCPs.[11] The types of persistent HCP(s), however, not only depend on the employed host cell expression system, but also on the produced protein therapeutics. These may have very different physicochemical properties and therefore different critical HCPs than previously purified products.

Monitoring the purification process and measuring residual HCPs are the focus of intensive analytical development. Anti-HCP enzyme-linked immunosorbent assays (ELISAs) are the gold standard for determining overall HCP content to detection levels as low as 1 ng mL^{-1}.[10,12] However, the ELISA technique can only detect proteins against which it is developed, and total protein ELISAs do not provide information on individual proteins present in the drug substance or product. Therefore, the use of orthogonal methods to support process development and validation is recommended.[13]

Significant advancements in high-resolution mass spectrometry in recent decades have enabled large-scale proteomics with greater accuracy, sensitivity and throughput. Mass-spectrometry-based proteomics has emerged as a powerful alternative to identify and quantify HCPs to detection limits of up to 5 ppm for known and unknown components.[14] Consequently, host cell proteomics have been increasingly employed to monitor purification progress and to confirm the absence of specific HCPs in the final drug substance or product.[8,9,13–16]

When a new purification process is designed, suitable chromatography resins and buffer conditions have to be identified. Three main process steps are commonly used in protein purification.[15] The first is the "capture step", which serves as the gross purification step. The bulk of the impurities is removed, thereby concentrating the protein product. The subsequent intermediate purification steps use various chromatographic resins to further reduce impurities. Finally, the polishing step removes low-abundance and minor impurities.[15] Frequently applied chromatographic separation techniques are ion exchange, hydrophobic interaction, mixed mode, size-exclusion, or affinity-based chromatography, where packed bed resins are currently state-of-the-art.[6]

Identifying the most effective technique for the removal of HCPs is difficult without extensive experimental and predictive data. In particular, anticipating the presence of critical HCPs that are difficult to remove or that are retained by the product during processing remains challenging.[12] Currently, the development of new processes still requires expert knowledge and high-throughput screening approaches to identify suitable conditions for the development of effective purification steps.[15,17] Advanced process development tools are needed that use a more rational and systematic approach.[12,18] In previous work, mechanistic models have been used to describe the binding behaviour of HCP on several chromatographic columns.[19–21] Isotherm parameters of HCP were determined from the chromatographic separations. Alternatively, the affinity of process-related impurities (including HCPs) to a library of resins was described.[22,23]

Notably, extensive data are available on model organisms commonly employed in clinical and medical studies, such as *E. coli* K12, CHO cells or *Pichia pastoris*. Conversely, limited studies have been conducted on the proteomes of strains developed and optimised for biotechnological applications, including the widely employed host strains of *E. coli* BLR and HMS174. The advantages of comprehensively analysing the proteome present in the harvest before the capture step is often overlooked. Knowledge of protein impurities, including their abundance and characteristics relative to the expressed protein therapeutic, can facilitate the development of an effective purification strategy.

In this study, we characterise the complete host cell proteome of two widely employed *E. coli* strains BLR and HMS174, using state-of-the-art Orbitrap mass spectrometry. The established proteomic data were further used to construct a database resource containing information regarding observed expression levels, hydrophobicity, isoelectric points (pI), molecular weights (MW), subunit information, possible post-translational modifications (PTMs), and toxicity for every possible gene product. The properties of the expressed protein therapeutics can

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

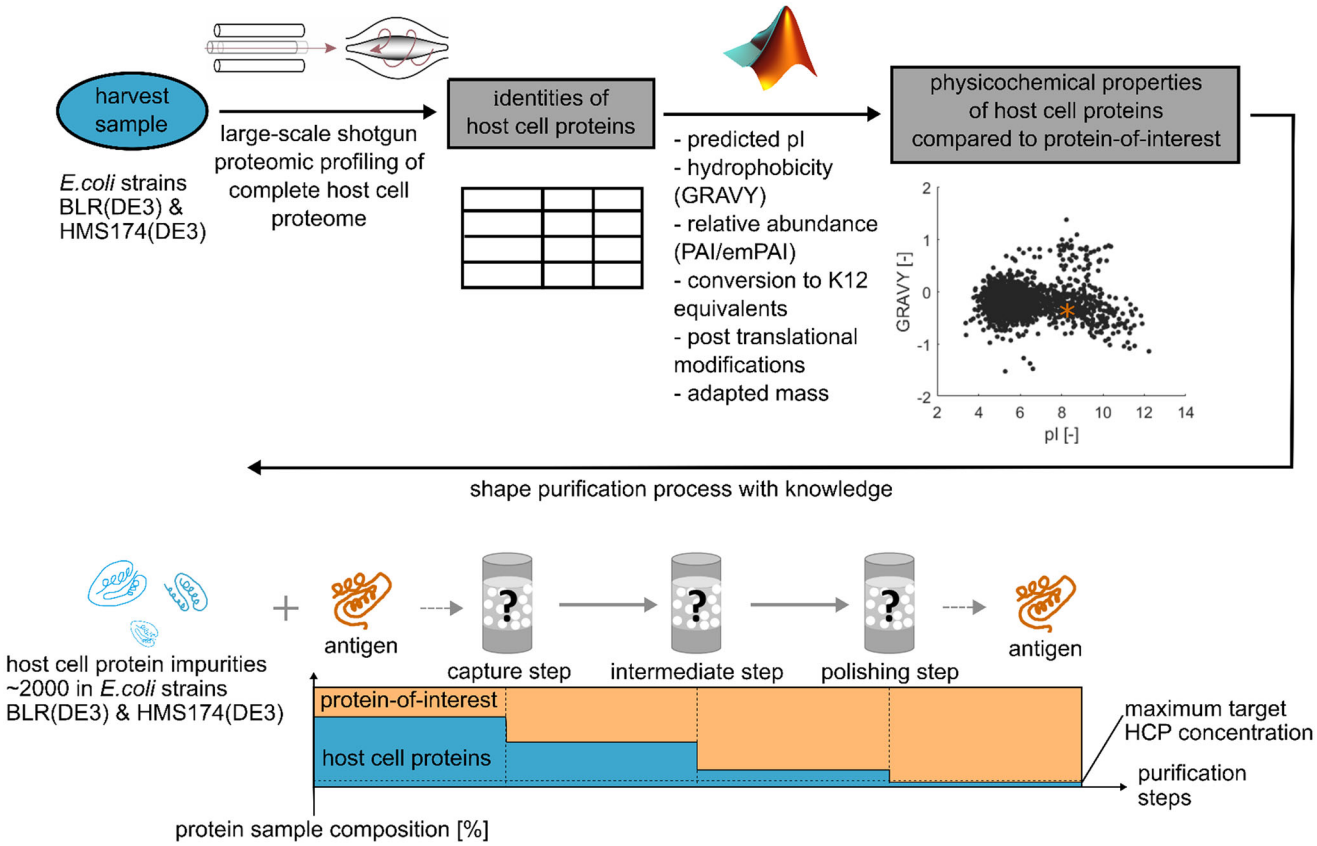

FIGURE 1 Schematic overview of the process development approach guided by large-scale host cell proteomics described in this study. The clarified harvest sample from the fermentation process was analysed using mass-spectrometry-based proteomics to identify all detectable HCPs. Further, a range of physicochemical properties was calculated for every possible gene product. One can guide the selection of the most suitable purification process by comparing the properties of protein therapeutics with those in the established database resource.

then be evaluated in the context of the complete host cell proteome. This extensive resource generated by mass spectrometry analysis of the host cell proteome, therefore, leads to a more rational and accelerated purification process development. Furthermore, we exemplify the use of the database resource for purification process development of the capture step for two model antigens used in a protein subunit vaccine produced with the *E. coli* strains BLR and HMS174.

2 | MATERIALS AND METHODS

2.1 | *E. coli* fermentation and harvest sample

The cultivation was performed as a standard fed-batch process using semi-synthetic media. Working seed for the pre-culture was first amplified in a shake flask until it reached an OD_{650} of about 2.0. Then ca. 20 mL of pre-culture is added in a 20 L fermenter filled with 9 L of culture medium. In the first part of the fermentation bacterial biomass was produced in fed-batch mode taking approximately 18 h to reach a volume of 12 L. Afterwards in the second phase of the fermentation, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce the production of the model antigen (same procedure in null plasmid strains). After 24 h the fermentation harvest was obtained and clarified.

The harvest samples were derived from the *E. coli* strains BLR(DE3) and HMS174(DE3) (further called BLR and HMS174). For both strains a fermentation was conducted using an empty plasmid cassette which did not encode the gene of the antigen. These two samples from null plasmid cell lines were frozen at −80°C before the clarification step. The third sample was obtained from the *E. coli* strain BLR producing the model antigen recombinantly. In the clarification, the *E. coli* cells in all samples were disrupted by homogenisation with a French pressure cell (*Sim Aminco* Spectronic Instruments), to obtain the intracellular, soluble products. In the further clarification, the samples were centrifuged for 45 min at 15,000 × *g* and filtered with a 0.2 μm PES filter. All harvest material for the analysis of the host cell proteome was provided by GSK (Rixensart, Belgium).

2.2 | Sample preparation for host cell proteomic analysis

The *E. coli* host cell proteome samples from BLR and HMS174 were prepared in accordance with recently published protocols by den

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ridder et al.[24] A detailed description is provided in the supplementary information of this manuscript.

2.3 | Shotgun host cell proteomics

The shotgun proteomics experiments are described in the supplementary information in detail. Briefly, HCPs were identified using a nano-liquid-chromatography separation system consisting of an EASY-nLC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column and a QE plus Orbitrap mass spectrometer (Thermo Scientific, Germany). The Orbitrap was operated in data-dependent acquisition (DDA) mode. The mass spectrometry proteomics raw data for the null plasmid cell lines of *E. coli* strains BLR and HMS174, have been deposited in the ProteomeXchange consortium database with the dataset identifier PXD035590.

2.4 | Processing of mass spectrometric raw data

Mass spectrometric raw data were analysed using PEAKS Studio X (Bioinformatics Solutions Inc., Canada) as described in more detail in the supplementary information. The mass spectrometric raw data were further analysed using strain specific proteome sequence databases obtained from NCBI (*E. coli* BLR: BioProject PRJNA379778 and *E. coli* HMS174 BioProject PRJEB6353) and the GPM crap contaminant proteins sequences (https://www.thegpm.org/crap/). Relative protein abundances (or content) were estimated using the protein abundance index (PAI) and the exponentially modified PAI (emPAI) according to Ishihama et al.[25] Label free quantification of protein abundance changes between the null plasmid *E. coli* strain and the corresponding antigen producing strain was performed using the PEAKSQ module.[26]

2.5 | Construction of host cell proteome property databases, for *E. coli* BLR and HMS174

The two databases accessible in the supplementary data were based on the mass spectrometry measurement of the clarified harvest samples originating from null plasmid cell lines from the *E. coli* strains BLR and HMS174. Each protein has a protein group, protein ID and accession assigned. The average mass, area and coverage is determined via the MS measurement and proteins are ranked according to their spectral count. PAI is defined as the number of sequenced peptides (fragmentation spectra assigned with significant score and as the top match to an individual identified protein) divided by the number of its calculated, observable peptides.[27] This value was used as the abundance measure in the comparison between proteomes. Furthermore, the PAI was converted to the emPAI, equal to 10^{PAI} minus one as described by Ishihama et al.[25] With help of the emPAI, the protein content was calculated in molar percent and weight percent[25] Each individual protein was assigned its calculated physicochemical properties. Calculated pI, calculated charge and grand average of hydropathy (GRAVY) as a measure of hydrophobicity were chosen as properties that define the most useful separation mechanism. For this purpose an in-house Matlab program was written that sorted the proteins according to their accession and assigned the physicochemical parameter predicted based on the amino acid sequence. The isoelectric point was predicted using the Matlab function "isoelectric" and the "Isoelectric Point Calculator 2.0″ software[28], that predicts the pI based on 21 different models. The average pI of the different calculation methods was used in the plotted graphs thereafter. The charge of the proteins was calculated in Matlab with the function "isoelectric" based on the amino acid sequence of the protein. The hydrophobicity was extracted in form of the GRAVY based on the amino acid sequence of the HCP (http://www.gravy-calculator.de/). A GRAVY value below 0 describes a hydrophilic protein, while scores above 0 are describing hydrophobic proteins. The sum of GRAVY values of the amino acids in the protein sequence divided by the number of amino acids is used as the GRAVY value of the protein. The toxicity is predicted using the ToxinPred2 tool[29]. Selected machine learning technique was hybrid (RF + BLAST + MERCI) with a threshold value of 0.6. Protein subunit information and knowledge about possible occurrence of PTMs for *E. coli* BLR and HMS174 were inferred from the *E. coli* K12 strain, which proteome sequence was obtained from Uniprot reference proteome sequence database (UP000000625_83333). The alignment of sequences was performed for this purpose using the Diamond sequence aligner [30] where the quality of the match was assessed by considering sequence identity and e-values.

2.6 | Codes and functions used for visualisation of host cell proteome properties

In-house Matlab scripts were used to plot the physicochemical properties of the identified proteins into property maps using scatter plots. The database including all proteins identified in the sample was used as input and the abundance was plotted over the mass, pI and GRAVY of the identified proteins. In the next step, the pI was plotted over the GRAVY and the charge at pH 7.0 over the GRAVY values. This analysis was conducted for all identified proteins, the 20 top abundant HCPs and the antigen properties.

3 | RESULTS AND DISCUSSION

3.1 | A comprehensive host cell proteome database for *E. coli* BLR and HMS174

Characterising the host cell proteome (i.e., protein abundances and predicted properties) is expected to streamline the development of purification processes significantly. Hence, we performed a proteomic characterisation of the widely employed *E. coli* BLR and HMS174 strains and predicted the physicochemical properties

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for all possible gene products. For example, differences in pI and hydrophobicity (GRAVY) affect the selection of the most common chromatographic methods, which are ion-exchange chromatography (IEX) and hydrophobic interaction chromatography (HIC). The proteome database was further expanded with parameters such as protein coverage, area, and protein content indices (protein abundance index PAI and the exponentially modified protein abundance index emPAI). The most abundant proteins in the database for the BLR and HMS174 strains are presented in Table 1 and Table 2, respectively. The complete database for both strains is available in the supplementary material. From the 4295 proteins of the complete proteome of *E. coli* BLR, 1993 HCPs were detected in the null plasmid strain, and 2006 were identified when additionally expressing the model antigen. In *E. coli* HMS174, 4216 proteins are found in the theoretical proteome, of which 1886 were detected in the null plasmid strain. Most of the abundant proteins have functions in biosynthesis or are ribosomal proteins. The most abundant protein in both strains, appeared to be the ATP synthase F1 subunit epsilon. This protein generates ATP from ADP in the presence of a proton gradient across the membranes. However, this protein is relatively small and has only one theoretically observable peptide (in the considered mass range 800–2400 Da) according to the original definition of PAI.[31] Therefore, the observed peptides divided by the number of theoretically observable peptides provides disproportionally high PAI values. Furthermore, we linked all protein sequences to homologous counterparts of the well investigated model organism *E. coli* K12 using sequence alignment. This enabled inferring information about possible complex formation and occurrence of PTMs. The latter could alter the protein size and net charge. For *E. coli* BLR, 224 PTMs are listed in the database, while 221 PTMs are listed for *E. coli* HMS174.

3.2 | HCP differences between *E. coli* strains, null plasmid, and antigen expressing strains

Furthermore, we compared the proteome and expression pattern between the BLR and HMS174 (null plasmid) strains. Out of all the identified proteins, approximately 80% (1590 proteins) were detected in both strains. A correlation graph using the abundance values (expressed by the PAI metric) provided for a linear regression with an R^2 of 0.69 (Figure 2A). This overlap shows that the bulk amount of HCPs is comparable even between different *E. coli* strains. Furthermore, we compared the identified proteins and abundances between the BLR null plasmid and the corresponding antigen-expressing strain. Here, approximately 90% (1779) of the identified proteins were identical in both samples. After plotting the abundances of the observed proteins, an R^2 value of 0.81 was obtained (Figure 2B). Differences in the abundances, however, may also be partly due to slight differences in the sample preparation procedure (e.g., the antigen-containing harvest was exposed to one freeze/thaw cycle before the clarification step). Nevertheless, the expression of the antigen is expected to have some impact on the observed host cell proteome. The differences may be minor and the findings from the null strain can be applied to determine a purification strategy for the antigen-producing strain.

3.3 | Visualizing the host cell proteome using global property maps

The properties of the host cell proteomes were further visualised using proteome property maps. The use of global property maps can be an effective tool for designing an optimal purification strategy.[19] For example, differences in the properties between the most abundant HCPs (or critical HCPs) and the antigen allow identification of the most promising resins for the first purification step. In the following subsequent sections various property maps (abundance vs. pI/GRAVY; pI vs. GRAVY; and net charge vs. GRAVY) are discussed. The data of two model antigens are shown and possible purification strategies for the capture step are discussed based on differences between the antigens and the most abundant proteins.

3.4 | Abundances versus molecular weight (MW), pI, and hydrophobicity (GRAVY)

The (null plasmid) BLR and HMS174 strains were compared based on properties such as MW (mass), pI, and GRAVY. Utilizing this approach enabled the search for conditions in which the majority of the HCPs differ from the expressed protein therapeutics (in this case, antigens). The properties of "antigen 1" expressed in BLR and "antigen 2" expressed in HMS174 are shown in the graph in relation to the properties of the HCPs (Figure 3) to define a purification strategy. Both strains show similar distributions of abundances compared to their protein properties, which is unsurprising, as a large number of proteins are identified in both strains with relatively similar abundances.

The MWs of the HCPs vary between 2 and 250 kDa, with the majority of proteins having a MW < 50 kDa (Figure 3A and 3D). The high-abundance proteins are in the lower MW range. Antigen 1 has a MW of 59 kDa, while antigen 2 (28 kDa) is comparatively small. Separating the antigens from the HCPs with a separation mechanism based on the size of the molecules, for example, size exclusion chromatography (SEC), seems to be suitable for later purification steps.[15] The discrepancy between mass of abundant HCP to the antigens might be a poor separation property for the capture step.

The pI spectrum of the identified HCPs ranges from pH 3.4–12.2, where the majority of the proteins are acidic (Figure 3B and 3E). A valley with fewer proteins is visible between a pI of 7 and 8. This valley can be explained by the intracellular pH for *E. coli* (approx. pH = 7.5) that would decrease the stability of proteins with a similar pI. Antigen 1 is located at the lower end of the pI spectrum with a pI of 4.4, while antigen 2 is close to the valley with fewer identified proteins with a pI of 8.4. Both antigens have pIs that are significantly different from the HCPs. One could consider a separation based on charge, such as IEX, as a promising capture step.

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 The top 20 HCPs (according to the PAI values) observed in the null plasmid fermentation using the *E. coli* strain BLR and their physicochemical properties. The complete database is provided as SI table.

Protein accession	Protein name	Avg. mass [Da]	Area BLR [-]	PAI [-]	emPAI [-]	Protein content [mol %]	Protein content [weight %]	GRAVY [-]	Net charge pH 7.0 Matlab [-]	Average pI [-]	Accession closest *E. coli* K12 analogue
ARH99394.1_3613	F1 sector of membrane-bound ATP synthase epsilon subunit	15,068	3.41E+08	5.00	99,999	73.03	63.65	−0.095	−4.6	5.4	P0A6E6
ARH98063.1_2282	phosphohistidinoprotein-hexose phosphotransferasecomponent of PTS system (Hpr)	9,119	1.04E+09	4.00	9999	7.30	3.85	−0.166	−1.5	5.6	P0AA04
ARH98931.1_3150	50S ribosomal subunit protein L22	12,226	6.06E+09	3.25	1777	1.30	0.92	−0.349	10.9	10.4	P61175
ARH99273.1_3492	50S ribosomal subunit protein L33	6,372	2.29E+09	3.00	999	0.73	0.27	−0.804	10.5	10.5	P0A7N9
ARH97584.1_1803	CopC family protein	13,410	1.34E+08	3.00	999	0.73	0.57	−0.262	5.4	9.5	P0AA57
ARH97040.1_1259	putative uncharacterized protein YciS	11,351	1.69E+07	3.00	999	0.73	0.48	0.680	4.2	9.6	P0ACV4
ARH99646.1_3865	50S ribosomal subunit protein L7/L12	12,295	6.16E+09	2.75	561	0.41	0.29	0.295	−8.0	4.5	P0A7K2
ARH98925.1_3144	50S ribosomal subunit protein L24	11,316	8.61E+09	2.75	561	0.41	0.27	−0.381	11.2	10.5	P60624
ARH96394.1_613	glutamate/aspartate ABC transporter substrate-binding protein	33,375	7.26E+09	2.73	532	0.39	0.75	−0.445	3.4	8.4	P37902
ARH98288.1_2507	glycine betaine ABC transporter substrate-binding protein	36,046	1.28E+09	2.57	372	0.27	0.57	−0.320	−2.9	6.0	P0AFM2
ARH98851.1_3070	50S ribosomal subunit protein L13	16,019	8.02E+09	2.50	315	0.23	0.21	−0.540	11.7	10.1	P0AA10
ARH98514.1_2733	fructose-bisphosphate aldolase class II	39,147	4.59E+09	2.50	315	0.23	0.52	−0.224	−9.0	5.5	P0AB71
ARH96740.1_959	methylglyoxal synthase	16,919	6.30E+07	2.50	315	0.23	0.23	0.038	−1.4	6.2	P0A731
ARH99864.1_4083	30S ribosomal subunit protein S6	15,173	6.41E+09	2.40	250	0.18	0.16	−0.745	−6.6	5.1	P02358
ARH96687.1_906	30S ribosomal protein S1	61,158	1.64E+10	2.33	214	0.16	0.55	−0.300	−27.1	4.7	P0AG67
ARH98856.1_3075	malate dehydrogenase NAD(P)-binding	32,337	7.68E+09	2.33	214	0.16	0.29	0.194	−2.6	5.5	P61889
ARH98791.1_3010	30S ribosomal subunit protein S15	10,269	5.80E+09	2.33	214	0.16	0.09	−0.673	8.2	10.6	P0ADZ4
ARH97768.1_1987	galactitol-specific enzyme IIB component of PTS	10,270	1.70E+09	2.33	214	0.16	0.09	0.291	−1.4	5.8	P37188
ARH99054.1_3273	aspartate-semialdehyde dehydrogenase NAD(P)-binding	40,034	9.87E+08	2.33	214	0.16	0.36	−0.040	−5.9	5.2	P0A9Q9
ARH98920.1_3139	50S ribosomal subunit protein L18	12,770	3.64E+09	2.25	177	0.13	0.10	−0.395	10.7	10.6	P0C018

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 The top 20 most abundant HCPs (according to the PAI values) identified in the null plasmid fermentation of the *E. coli* strain HMS174 and their physicochemical properties. The complete database is provided as SI table.

Protein accession	Protein name	Avg. mass [Da]	Area HMS [-]	PAI [-]	emPAI [-]	Protein content [mol %]	pPotein content [weight %]	GRAVY [-]	Net charge pH 7.0 Matlab [-]	Average pI [-]	Accession closest *E.coli* K12 analogue
CDY62850.1	ATP synthase F1 complex-epsilon subunit of ATP synthase F1 complex	15,068	3.10E+08	6.00	999,999	95.69	93.53	−0.095	−4.6	5.4	P0A6E6
CDY60193.1	phosphohistidinoprotein-hexose phosphotransferasecomponent of PTS system (Hpr)	9119	4.81E+09	4.00	9,999	0.96	0.57	−0.166	−1.5	5.6	P0AA04
CDY64403.1	major type 1 subunit fimbrin (pilin) subunit of fimbrial complex	18,111	7.18E+07	4.00	9,999	0.96	1.12	0.310	−2.6	4.9	P04128
CDY56811.1	methylglyoxal synthase	16,919	3.63E+08	3.50	3,161	0.30	0.33	0.038	−1.4	6.2	P0A731
CDY58853.1	conserved protein	13,410	4.22E+07	3.00	999	0.10	0.08	−0.262	5.4	9.5	P0AA57
CDY56069.1	conserved protein involved in translation	17,526	1.06E+07	3.00	999	0.10	0.11	−0.311	−22.6	4.0	P0A898
CDY63707.1	50S ribosomal subunit protein L22 subunit of ribosome	12,226	1.39E+09	2.75	561	0.05	0.04	−0.349	10.9	10.4	P61175
CDY56061.1	glutamate/aspartate ABC transporter-periplasmic binding protein subunit of GltIJKL glutamate ABC transporter	33,420	3.56E+09	2.73	533	0.05	0.11	−0.472	3.4	8.4	P37902
CDY61370.1	fructose bisphosphate aldolase class II	39,147	5.41E+09	2.50	315	0.03	0.08	−0.224	−9.0	5.5	P0AB71

(Continues)

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Protein accession	Protein name	Avg. mass [Da]	Area HMS [-]	PAI [-]	emPAI [-]	Protein content [mol %]	pPotein content [weight %]	GRAVY [-]	Net charge pH 7.0 Matlab [-]	Average pI [-]	Accession closest *E.coli* K12 analogue
CDY63356.1	universal stress global stress response regulator	16,066	4.35E+09	2.50	315	0.03	0.03	−0.0556	−7.4	4.9	P0AED0
CDY57666.1	lipid hydroperoxide peroxidase	17,835	6.28E+09	2.40	250	0.02	0.03	0.256	−4.9	4.6	P0A862
CDY56210.1	SucB subunit of dihydrolipoyltranssuccinylase and 2-oxoglutarate dehydrogenase complex	44,011	5.43E+09	2.36	230	0.02	0.06	−0.217	−6.3	5.5	P0AFG6
CDY62241.1	malate dehydrogenase	32,337	1.47E+10	2.33	214	0.02	0.04	0.194	−2.6	5.5	P61889
CDY62031.1	30S ribosomal subunit protein S15 subunit of ribosome	10,269	1.78E+09	2.33	214	0.02	0.01	−0.673	8.2	10.6	P0ADZ4
CDY62519.1	superoxide dismutase (Mn)	23,097	2.17E+09	2.29	192	0.02	0.03	−0.429	−0.1	6.5	P00448
CDY57747.1	nucleotide binding filament protein	16,017	5.19E+08	2.20	157	0.02	0.02	0.022	−4.4	5.6	P37903
CDY60191.1	CysK subunit of cysteine synthase A and bifunctional CysEK cysteine biosynthesis complex	34,490	1.55E+10	2.17	146	0.01	0.03	−0.078	−2.3	5.8	P0ABK5
CDY62230.1	50S ribosomal subunit protein L13 subunit of ribosome	16,019	2.02E+09	2.17	146	0.01	0.01	−0.540	11.7	10.1	P0AA10
CDY60673.1	phage lambda replication; host DNA synthesis; heat shock protein; protein repair subunit of DnaJ/DnaK/GrpE	21,798	3.17E+08	2.17	146	0.01	0.02	−0.372	−14.3	4.5	P09372
CDY57250.1	isocitrate dehydrogenase	45,785	2.30E+10	2.16	143	0.01	0.04	−0.154	−11.0	5.0	P08200

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

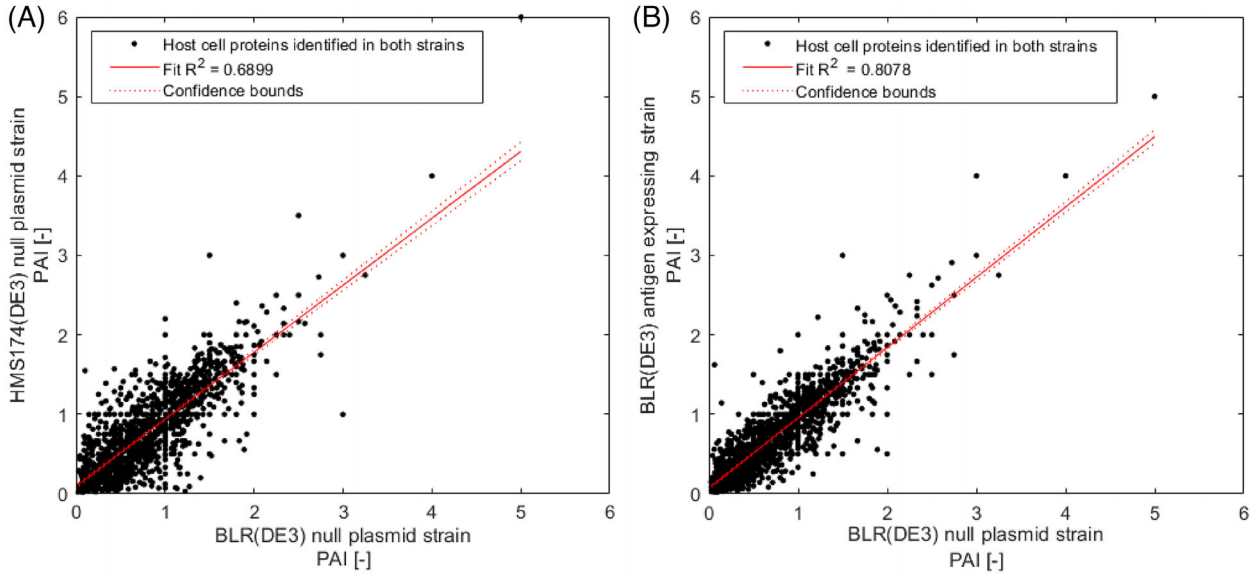

FIGURE 2 Scatterplots of the HCPs identified in the investigated E. coli strains. (A) presents a comparison of null plasmid *E. coli* strains BLR and HMS174. The correlation of 1590 proteins that were common to both strains resulted in an R^2 value of 0.6899. In (B), the null plasmid BLR was compared to the corresponding antigen-expressing strain; 1779 proteins were common to both samples, resulting in an R^2 value of 0.8078.

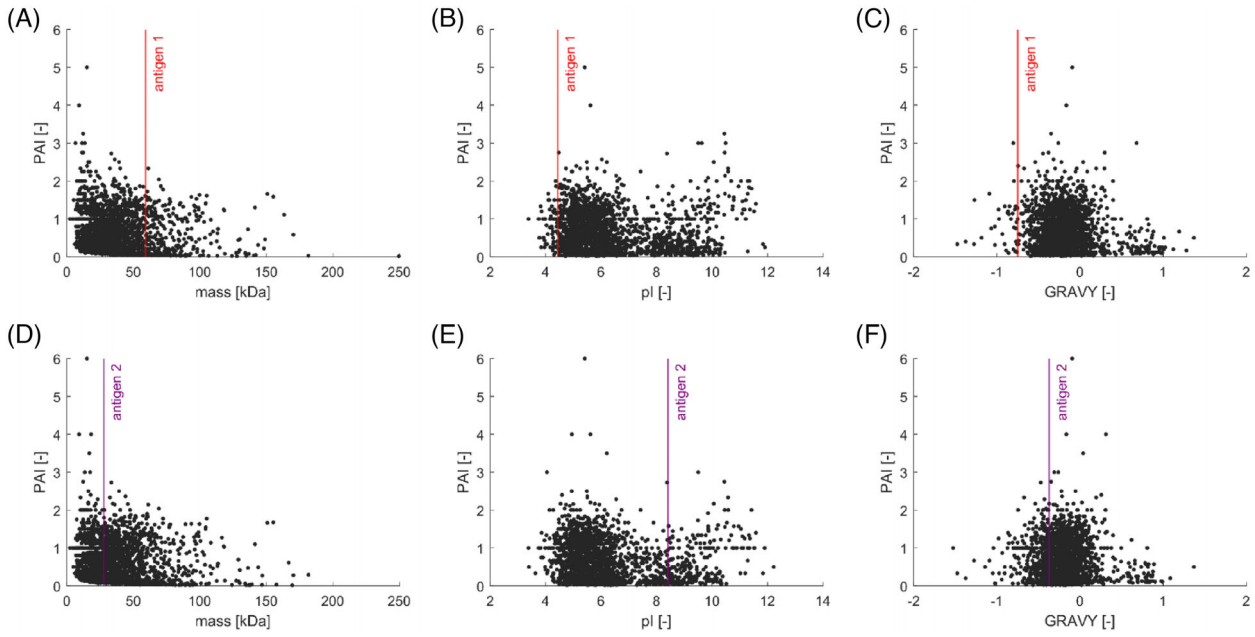

FIGURE 3 Abundances of the detected HCPs from null plasmid fermentations of the *E. coli* strains BLR (A–C) and HMS174 (D–F) are compared: (A and D) the mass of the proteins according to mass spectrometric measurements, (B and E) the average predicted isoelectric points (pI), and (C and F) hydrophobicity (GRAVY). Positive and negative GRAVY values describe hydrophobic and hydrophilic proteins, respectively. The model antigens 1 and 2 are indicated in red and purple. The abundances are expressed by the PAI parameter.

The estimated GRAVY values of the proteins range from −1.526 to +1.369. Most of the identified proteins have a slightly negative GRAVY value and are, hence, slightly hydrophilic (Figure 3C and 3F). Antigen 1 has a GRAVY of −0.749, which is relatively different from the values obtained for most HCPs. For antigen 1, a separation based on hydrophobicity (e.g., using HIC) therefore appears highly promising as a capture step.

3.5 | pI versus hydrophobicity (GRAVY)

We furthermore plotted the predicted pI against the hydrophobicity (GRAVY) of the identified host cell proteome, as shown in Figures 4A and 4C. Additionally, we generated a plot for the 20 most abundant HCPs and model antigens (Figures 4B and 4D, also listed in Table 1 and Table 2). In this example case, it was chosen to focus on the most

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

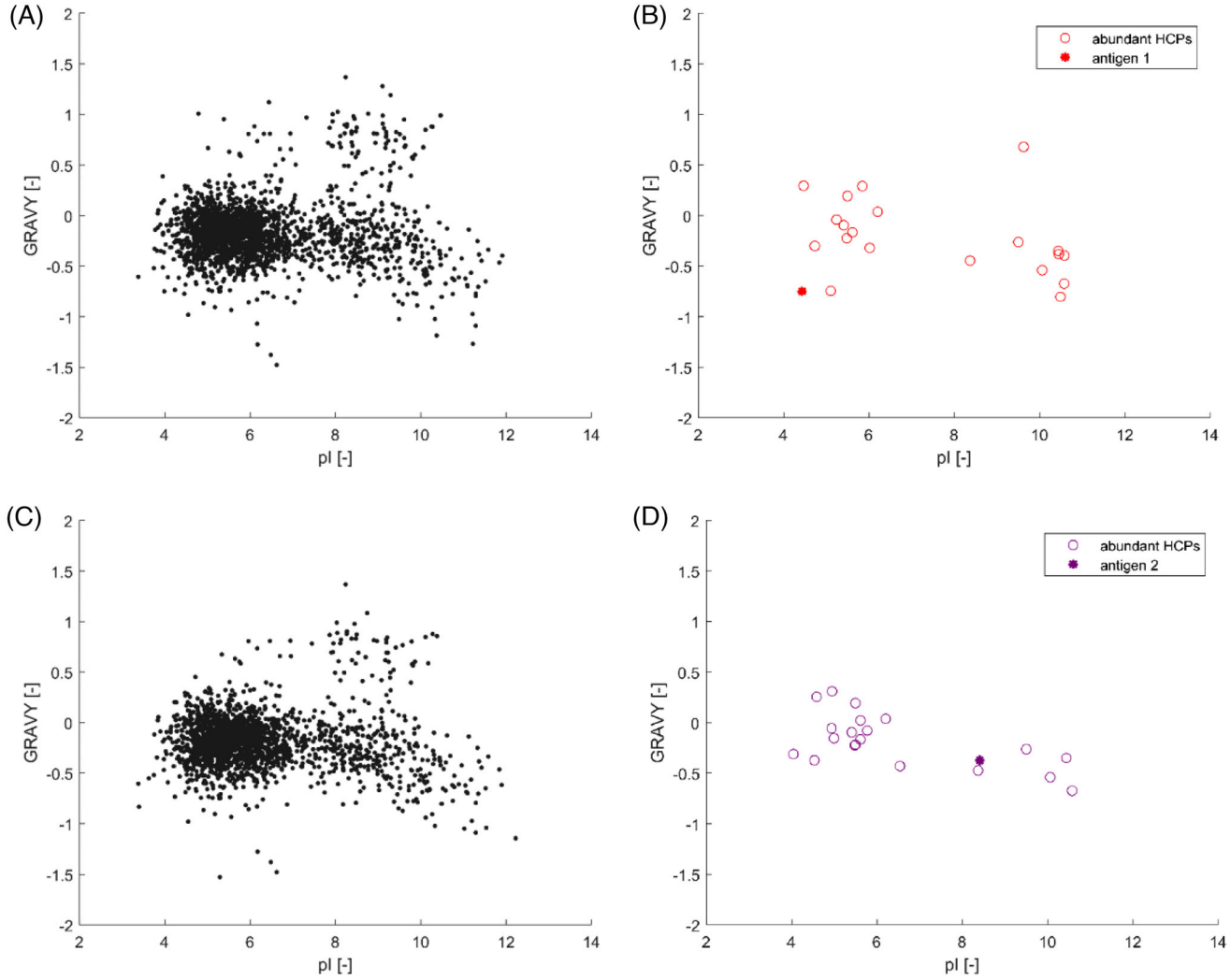

FIGURE 4 Comparison of HCPs from the null plasmid fermentations of the *E. coli* strains BLR (A and B) and HMS174 (C and D). The predicted hydrophobicity (GRAVY) is plotted against the predicted isoelectric point (pI). Displayed are (A) the properties of the complete, identified host cell proteome of BLR; (B) the properties of the most abundant HCPs and antigen 1 in BLR; (C) the properties of the complete, identified host cell proteome of HMS174; and (D) the properties of the most abundant HCPs and antigen 2 in HMS174.

abundant HCPs in the sample to design a capture step targeting the removal of the main HCP impurities. Antigen 1 has a low pI and GRAVY value compared to the most abundant HCPs. IEX together with HIC, or their combination in mixed mode chromatography, appear highly suitable for purifying this antigen.

Antigen 2 is located in close proximity to the centre of the pI spectrum of the HCPs. However, apart from the glutamate/aspartate ABC transporter periplasmic binding protein, the most abundant HCPs have a significantly different pI. IEX appears to be a suitable purification method. The GRAVY value of antigen 2, on the other hand, is not significantly different from the values of the most abundant HCPs.

3.6 | Net charge versus hydrophobicity (GRAVY)

The net charge of a protein depends on the pI and the pH value of the environment (solvent or buffer). Therefore, knowing the net charge of the HCPs at different pH values helps in selecting the most suitable conditions when using IEX. Plots at a pH of 7.0 were generated so that typically no buffer exchange (or pH adjustment) is required before the capture step, thus reducing time and costs, for example, for titration. We calculated the net charge of the HCPs at pH 7.0 and we plotted them against the predicted GRAVY values, which is shown in Figure 5. Net charges for a range of different pH conditions are furthermore included in the database resource of the supplementary information material.

In the case of BLR, 11 of the 20 most abundant proteins have a negative net charge at pH 7.0. Antigen 1 has a predicted net charge of −46.78, which is low compared to that of the other HCPs. Considering a bind-and-elute mode, anion-exchange chromatography at pH 7.0 seems highly suitable for the capture step. The other abundant HCPs with a positive net charge would be repelled by the ligands and would not bind to the resin under the identified conditions. The 11 negatively charged HCPs would bind to the resin at pH 7.0 but could be eluted

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

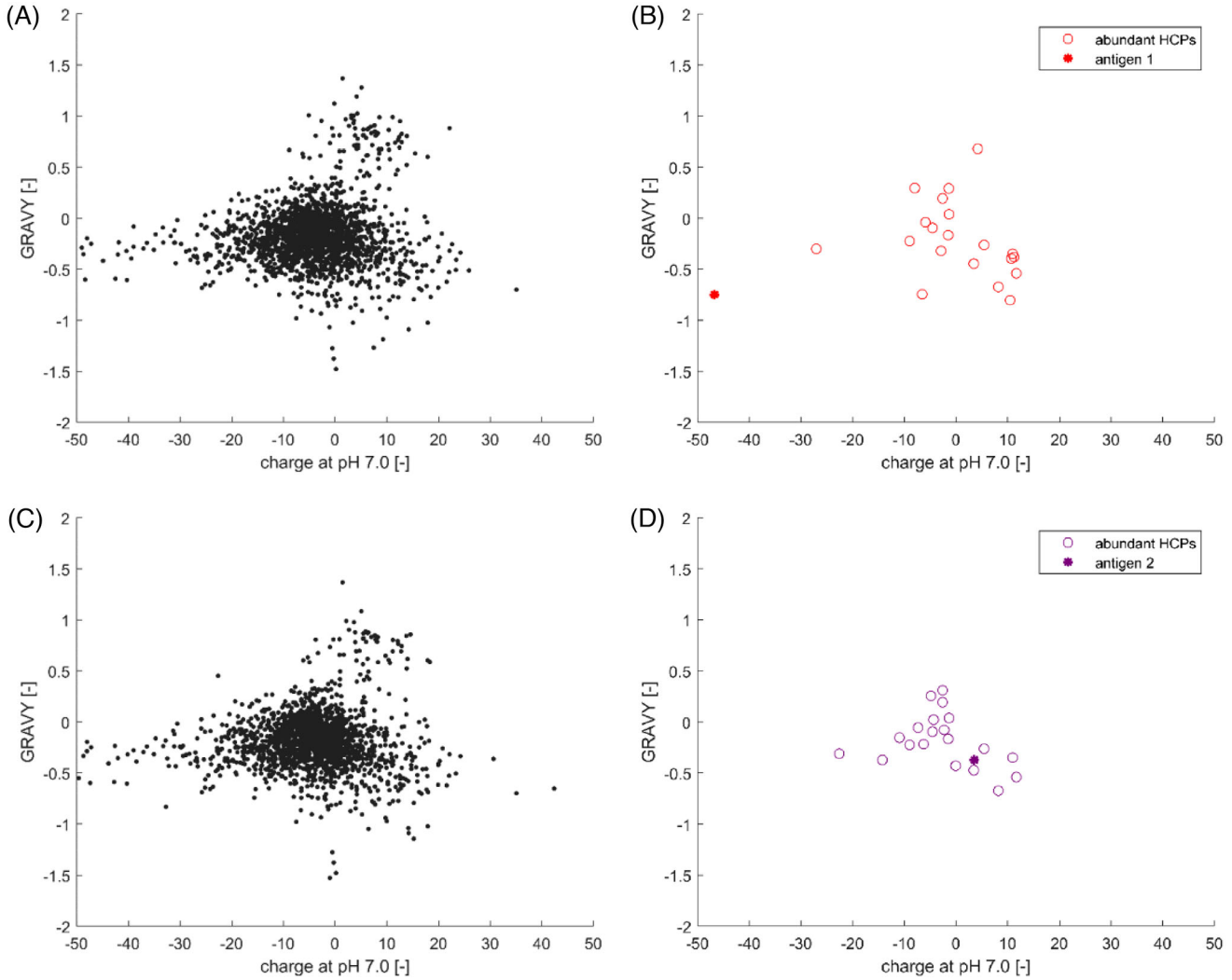

FIGURE 5 Comparison of HCPs from the null plasmid fermentations of the *E. coli* strains BLR (A and B) and HMS174 (C and D). The predicted hydrophobicity (GRAVY) is plotted against the predicted net charge at pH 7.0. Displayed are (A) the properties of the complete, identified host cell proteome of BLR; (B) the properties of the most abundant HCPs and antigen 1 in BLR; (C) the properties of the complete, identified host cell proteome of HMS174; and (D) the properties of the most abundant HCPs and antigen 2 in HMS174.

earlier using (low) salt-washing steps. A flow-through mode, on the other hand, seems suboptimal for this antigen at the specified pH. However, this approach might be suitable at pH values lower than the antigen pI. The majority of the abundant proteins in HMS174 – 15 out of 20 proteins – are negatively charged at pH 7.0. Antigen 2, on the other hand, has a slightly positive charge. In the case of a bind-and-elute mode, a cation-exchange step, combined with a salt elution step (at low ionic strength) to elute the antigen, could be suitable. Another option would be the use of an anion exchange resin in flow-through mode.

4 | CONCLUSION

The avoidance and removal of HCP impurities when purifying protein targets is particularly challenging. Characterising protein abundances and physicochemical properties enables a more rational, systematic, and accelerated development of the purification process. In this study, we performed a comprehensive characterisation of the complete host cell proteome for the widely employed *E. coli* strains BLR and HMS174. Furthermore, we constructed an extensive proteome property resource by integrating physicochemical properties such as hydrophobicity (GRAVY), calculated pI, and the predicted net charge at different pH values. Additionally, we determined PAI and emPAI parameters to estimate protein abundances and relative protein content. We then linked proteins with homologues of the well-investigated *E. coli* K12 strain shedding light on possible PTMs and complex formation. Furthermore, the protein abundances of null plasmid and antigen-expressing strains were compared, which demonstrated high similarity for the most abundant proteins.

We demonstrated the use of the established proteome resource database by creating global proteome property maps to support the design of new purification processes (or in particular to select the most

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

promising capture step). This avoids extensive trial-and-error studies and sole expert-knowledge-dependent choices.

AUTHOR CONTRIBUTIONS

Roxana Disela: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft; Writing – review & editing. Olivier Le Bussy: Conceptualization; Resources; Writing – review & editing. Geoffroy Geldhof: Conceptualization; Resources; Writing – review & editing. Martin Pabst: Conceptualization; Data curation; Methodology; Project administration; Software; Supervision; Writing – review & editing. Marcel Ottens: Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing – review & editing

ACKNOWLEDGMENTS

This study was funded by GlaxoSmithKline Biologicals S.A. under a cooperative research and development agreement between GlaxoSmithKline Biologicals S.A. (GSK, Belgium) and Delft University of Technology (TUD, The Netherlands). The authors are grateful to colleagues from GSK and the department of Biotechnology from TUD for valuable discussions.

CONFLICT OF INTEREST STATEMENT

All authors have declared the following interests: Geoffroy Geldhof and Olivier Le Bussy are employees of the GSK group of companies. Geoffroy Geldhof reports ownership of shares from the GSK group of companies. Roxana Disela reports that her PhD was financed by the GSK group of companies. The other authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics raw data for the null plasmid cell lines of *E. coli* strains BLR and HMS174, have been deposited in the ProteomeXchange consortium database with the dataset identifier PXD035590.

ORCID

Roxana Disela https://orcid.org/0000-0002-8178-5684
Martin Pabst https://orcid.org/0000-0001-9897-0723

REFERENCES

1. Bracewell, D. G., Francis, R., & Smales, C. M. (2015). The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnology and Bioengineering*, *112*, 1727–1737.
2. Jones, M., Palackal, N., Wang, F., Gaza-Bulseco, G., Hurkmans, K., Zhao, Y., Chitikila, C., Clavier, S., Liu, S., Menesale, E., Schonenbach, N. S., Sharma, S., Valax, P., Waerner, T., Zhang, L., & Connolly, T. (2021). "High-risk" host cell proteins (HCPs): A multi-company collaborative view. *Biotechnology and Bioengineering*, *118*, 2870–2885.
3. Vanderlaan, M., Zhu-Shimoni, J., Lin, S., Gunawan, F., Waerner, T., & Van Cott, K. E. (2018). Experience with host cell protein impurities in biopharmaceuticals. *Biotechnology Progress*, *34*, 828–837.
4. Krutzke, L., Roesler, R., & Wiese, S. (2021). *Research Square*, DOI 10.21203/rs.3.rs-477964/v1
5. Hanke, A. T., & Ottens, M. (2014). Purifying biopharmaceuticals: Knowledge-based chromatographic process development. *Trends in Biotechnology*, *32*, 210–220.
6. Gottschalk, U., Brorson, K., & Shukla, A. A. (2012). The need for innovation in biomanufacturing. *Nature Biotechnology*, *30*, 489–492.
7. Narayanan, H., Luna, M. F., von Stosch, M., Cruz Bournazou, M. N., Polotti, G., Morbidelli, M., Butté, A., & Sokolov, M. (2020). Bioprocessing in the digital age: The role of process models. *Biotechnology Journal*, *15*, 1–10.
8. Reiter, K., Suzuki, M., Olano, L. R., & Narum, D. L. (2019). Host cell protein quantification of an optimized purification method by mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, *174*, 650–654.
9. Zhu, D., Saul, A. J., & Miles, A. P. (2005). A quantitative slot blot assay for host cell protein impurities in recombinant proteins expressed in *E. coli*. *Journal of Immunological Methods*, *306*, 40–50.
10. Tscheliessnig, A. L., Konrath, J., Bates, R., & Jungbauer, A. (2013). Host cell protein analysis in therapeutic protein bioprocessing – Methods and applications. *Biotechnology Journal*, *8*, 655–670.
11. Chiu, J., Valente, K. N., Levy, N. E., Min, L., Lenhoff, A. M., & Lee, K. H. (2017). Knockout of a difficult-to-remove CHO host cell protein, lipoprotein lipase, for improved polysorbate stability in monoclonal antibody formulations. *Biotechnology and Bioengineering*, *114*, 1006–1015.
12. Hogwood, C. E. M., Bracewell, D. G., & Smales, C. M. (2014). Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Current Opinion in Biotechnology*, *30*, 153–160.
13. Falkenberg, H., Waldera-Lupa, D. M., Vanderlaan, M., Schwab, T., Krapfenbauer, K., Studts, J. M., Flad, T., & Waerner, T. (2019). Mass spectrometric evaluation of upstream and downstream process influences on host cell protein patterns in biopharmaceutical products. *Biotechnology Progress*, *35*, e2788.
14. Eliuk, S., & Makarov, A. (2015). Evolution of orbitrap mass spectrometry instrumentation. *The Annual Review of Analytical Chemistry*, *8*, 61–80.
15. Wen, E., & Ellis, R. (2015). N. S. Pujar, Eds., *Vaccine development and manufacturing*. Wiley.
16. Reisinger, V., Toll, H., Ernst, R., Visser, J., & Wolschin, F. (2014). A mass spectrometry-based approach to host cell protein identification and its application in a comparability exercise. *Analytical Biochemistry*, *463*, 1–6.
17. Wierling, P. S., Bogumil, R., Knieps-Grünhagen, E., & Hubbuch, J. (2007). High-throughput screening of packed-bed chromatography coupled with SELDI-TOF MS analysis: Monoclonal antibodies versus host cell protein. *Biotechnology and Bioengineering*, *98*, 440–450.
18. Keulen, D., Geldhof, G., Le Bussy, O., Pabst, M., & Ottens, M. (2022). Recent advances to accelerate purification process development: A review with a focus on vaccines. *Journal of Chromatography A*, *1676*, 463195.
19. Nfor, B. K., Ahamed, T., Pinkse, M. W. H., van der Wielen, L. A. M., Verhaert, P. D. E. M., van Dedem, G. W. K., Eppink, M. H. M., van de Sandt, E. J. A. X., & Ottens, M. (2022). Multi-dimensional fractionation and characterization of crude protein mixtures: Toward establishment of a database of protein purification process development parameters. *Biotechnology and Bioengineering*, *109*, 3070–3083.
20. Hanke, A. T., Tsintavi, E., Ramirez Vazquez, P., van der Wielen, L. A. M., Verhaert, P. D. E. M., Eppink, M. H. M., van de Sandt, E. J. A. X., & Ottens, M. (2016). 3D-liquid chromatography as a complex mixture characterization tool for knowledge-based downstream process development. *Biotechnology Progress*, *32*, 1283–1291.
21. Pirrung, S. M., Parruca da Cruz, D., Hanke, A. T., Berends, C., Van Beckhoven, R. F. W. C., Eppink, M. H. M., & Ottens, M. (2018). Chromatographic parameter determination for complex biological feedstocks. *Biotechnology Progress*, *34*, 1006–1018.
22. Timmick, S. M., Vecchiarello, N., Goodwine, C., Crowell, L. E., Love, K. R., Love, J. C., & Cramer, S. M. (2018). An impurity characterization

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

based approach for the rapid development of integrated downstream purification processes. *Biotechnology and Bioengineering*, *115*, 2048–2060.
23. Vecchiarello, N., Timmick, S. M., Goodwine, C., Crowell, L. E., Love, K. R., Love, J. C., & Cramer, S. M. (2019). A combined screening and in silico strategy for the rapid design of integrated downstream processes for process and product-related impurity removal. *Biotechnology and Bioengineering*, *116*, 2178–2190.
24. den Ridder, M., Knibbe, E., van den Brandeler, W., Daran-Lapujade, P., & Pabst, M. (2022). A systematic evaluation of yeast sample preparation protocols for spectral identifications, proteome coverage and post-isolation modifications. *Journal of Proteomics*, *261*, 104576.
25. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., & Mann, M. (2005). *Molecular & Cellular Proteomics*, *4*, 1265–1272.
26. den Ridder, M., Daran-Lapujade, P., & Pabst, M. (2020). *Fems Yeast Research*, *20*, 1–9.
27. Rappsilber, J., Ryder, U., Lamond, A. I., & Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Research*, *12*, 1231–1245.
28. Kozlowski, L. P. (2021). IPC 2.0: Prediction of isoelectric point and p K a dissociation constants. *Nucleic Acids Research*, *49*, W285–W292.
29. Sharma, N., Naorem, L. D., Jain, S., & Raghava, G. P. S. (2022). ToxinPred2: An improved method for predicting toxicity of proteins. *Brief Bioinformatics*, https://doi.org/10.1093/bib/bbac174
30. Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*, 59–60.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Disela, R., Le Bussy, O., Geldhof, G., Pabst, M., & Ottens, M. (2023). Characterisation of the *E. coli* HMS174 and BLR host cell proteome to guide purification process development. *Biotechnology Journal*, 1–13. https://doi.org/10.1002/biot.202300068

18607314, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/biot.202300068 by Tu Delft, Wiley Online Library on [22/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License