Analyst

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d4an00333k

Characterization of recombinant human lactoferrin expressed in *Komagataella phaffii*†

Xiaoning Lu, *[a] Chad Cummings, [a] Udodili A. Osuala, [a] Neela H. Yennawar, [b] Kevin E. W. Namitz, [b] Brittney Hellner, [a] Pamela B. Besada-Lombana, [a] Ross D. Peterson [a] and Anthony J. Clark [a]

Received 1st March 2024,
Accepted 25th May 2024

DOI: 10.1039/d4an00333k

rsc.li/analyst

This work presents a thorough characterization of Helaina recombinant human lactoferrin (rhLF, Effera™) expressed in a yeast system at an industrial scale for the first time. Proteomic analysis confirmed that its amino acid sequence is identical to that of native human LF. N-linked glycans were detected at three known glycosylation sites, namely, Asparagines-156, -497, and -642 and they were predominantly oligomannose structures having five to nine mannoses. Helaina rhLF's protein secondary structure was nearly identical to that of human milk lactoferrin (hmLF), as revealed by microfluidic modulation spectroscopy. Results of small-angle X-ray scattering (SAXS) and analytical ultracentrifugation analyses confirmed that, like hmLF, Helaina rhLF displayed well-folded globular structures in solution. Reconstructed solvent envelopes of Helaina rhLF, obtained through the SAXS analysis, demonstrated a remarkable fit with the reported crystalline structure of iron-bound native hmLF. Differential scanning calorimetry investigations into the thermal stability of Helaina rhLF revealed two distinct denaturation temperatures at 68.7 ± 0.9 °C and 91.9 ± 0.5 °C, consistently mirroring denaturation temperatures observed for apo- and holo-hmLF. Overall, Helaina rhLF differed from hmLF in the N-glycans they possessed; nevertheless, the characterization results affirmed that Helaina rhLF was of high purity and exhibited globular structures closely akin to that of hmLF.

1. Introduction

Human lactoferrin (hLF) is an iron-binding glycoprotein found in high concentrations in human milk and to a lesser degree in tears, saliva, intestinal secretions, and neutrophils.[1] Human milk LF (hmLF) is a bioactive protein that nourishes newborns and supports optimal growth and development.[2,3] Lactoferrin (LF) also supports immune health at every life stage.[4,5] It is the second most abundant glycoprotein in human milk, constituting an estimated 15% of the total whey protein of human milk.[6] Human milk LF is a single polypeptide chain containing 691 amino acid residues with a molecular weight (MW) of about 80 kDa, and its secondary structure comprises α-helices, β-pleated sheets, and turns.[7] Its structure consists of two folded lobes connected by a three-turn α-helix, with each lobe having the capacity to bind one ferric ion (Fe^{3+}),[7,8] a characteristic that may assist iron absorption in the small intestine of infants fed human milk.[1,9]

N-glycosylation, a common yet complex post-translational modification, is another characteristic of hmLF with important structural and functional roles.[10,11] The N-glycans are linked to the asparagine (Asn) residue of hmLF, which, in turn, affects its structure and potential biological function.[11,12] In nature, N-glycan structures of LF vary across cell types (*e.g.*, human milk-derived *vs.* neutrophil-derived) and species (*e.g.*, human *vs.* bovine), from mother to mother, and throughout the lactation period.[13,14] For example, hmLF contains primarily complex N-glycans that are frequently sialylated and fucosylated and low abundance of oligomannose,[15] while bovine lactoferrin (bLF) has a different glycan distribution that includes abundant oligomannose, complex and hybrid glycans.[15] In summary, hmLF contains more complex N-glycans than bLF, which contains all three glycan types (Fig. 1A).[16,17] In addition, glycosylation in yeast generally results in hyper-mannosylated glycoforms,[18] a characteristic not found in mammalian systems (Fig. 1B).

Data from preclinical and clinical studies have shown the effects of supplementing the diet with LF on markers of iron status, highlighting the potential benefits of hmLF, bLF, and recombinant hLF (rhLF) as a food ingredient.[19–23] A preclinical study demonstrated that supplementation with rhLF from transgenic cows improved iron status markers (*e.g.*, red blood cells, hemoglobin, serum iron, serum ferritin).[19] Clinical

[a]*Helaina Inc., 345 Park Avenue South, 5th Floor, New York, NY 10010, USA. E-mail: xiaoning@myhelaina.com; Tel: +(1)814-852-9623*
[b]*X-ray Crystallography Core Facility, The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA*
† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4an00333k

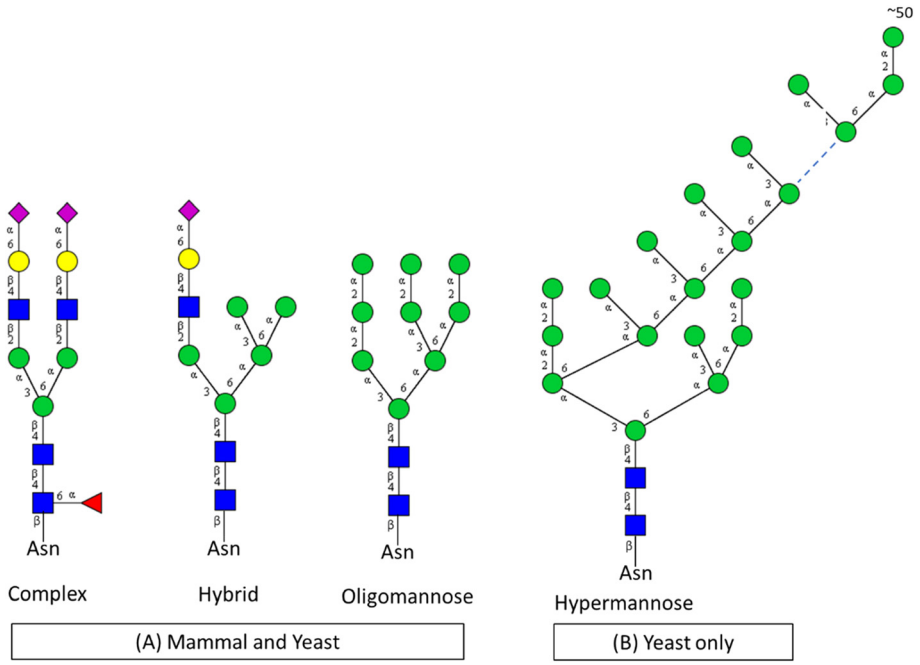

Fig. 1 The three types of N-glycans that are present in the glycoproteins of both mammals and yeast, namely complex, hybrid, and oligomannose structures (A). The N-glycans share a common core Man3GlcNAc2 which is linked to the Asparagine (Asn) residue of the glycoproteins having Asparagine (Asn)-X-Serine (Ser)/Threonine (Thr) sequons. Unlike mammals, the yeast tends to synthesize hypermannose N-glycans (B). The colored shapes represent monosaccharide residues: ■ *N*-acetylglucosamine (GlcNAc); ● mannose (Man); ○ galactose (Gal), ▲ fucose (Fuc); and ◆ *N*-acetylneuraminic acid (NeuAc).

studies have shown that extended dietary supplementation with LF increased serum iron, red blood cell number, and serum ferritin,[20,21] findings that have been confirmed in a recently published systematic review and meta-analysis of 11 intervention studies.[22]

In light of these declared benefits, there is a desire to use recombinant genetic technology to develop sustainable methods for the industrial scale production of rhLF. Indeed, a variety of recombinant systems have been used to express rhLF efficiently, including transgenic animals (*e.g.*, cows), plants (*e.g.*, rice), filamentous fungi (*e.g.*, *Aspergillus niger* var. *awamori*), and yeast (*e.g.*, *Saccharomyces cerevisiae*, *Komagataella phaffii* (*K. phaffii*, formerly classified as *Pichia pastoris*)).[24–31] For example, rhLF expressed from transgenic animal systems has a molecular weight, protein structure, and function similar to hmLF, yet their glycan structures differ.[16] Both hmLF and rhLF from transgenic cows contain sialylated and fucosylated glycans, whereas oligomannose glycans were reported with high abundance in the rhLF from transgenic cows and low abundance in hmLF.[11,16]

K. phaffii is a eukaryotic, methylotrophic yeast commonly used to express recombinant proteins efficiently, including animal-derived and human-derived recombinant LF.[32–35] Unlike most recombinant systems, *K. phaffii* can be modified to produce rhLF with glycan structures similar to hmLF. Also distinct from other host expression systems, *K. phaffii* can carry out high-density fermentation, which is a cost-effective way to generate high yields of recombinant proteins.[33] *K. phaffii* has been used as a production organism for numerous food ingredients that have received "no questions" letters from the U.S. Food and Drug Administration in the generally recognized as safe (GRAS) notification program.[36–38] Helaina has successfully used a proprietary technology to produce a modified strain of *K. phaffii* capable of expressing a form of rhLF substantively equivalent to hmLF at an industrial scale for the first time. Using multiple analytical methods, this publication comprehensively characterizes the identity, quality, glycosylation, structure, and iron saturation of Helaina rhLF.

2. Materials and methods

2.1 Materials

Samples of hmLF from human milk were purchased from MilliporeSigma (St Louis, MO, USA) and Abcam (Waltham, Boston, USA); bovine LF (bLF) was purchased from MilliporeSigma or The Lactoferrin Company (Byron Bay, Australia). Additionally, two hmLF samples were isolated and purified in-house from two different sources of human milk according to an established protocol.[39,40] Other reagents were purchased from MilliporeSigma or Thermo Fisher Scientific (Waltham, MA, USA) and used as is.

2.2 Methods

2.2.1 Production of Helaina rhLF. More than 10 fermentations of Helaina rhLF (Effera™) were produced from a

K. phaffii in an industrial scale (45 000 L) process, according to established procedure.[41–43] The fermentation process consisted of batch, fed-batch, and induction phases. At the end of fermentation, cells were separated from the protein-containing broth. Recombinant hLF was then isolated by microfiltration, ultrafiltration/diafiltration and ion exchange chromatography. The recovery of the purification is >60%. The resulting LF-rich solution was finally spray-dried to a powder. Samples are available upon request.

2.2.2 Total protein assay and simple western™. Total protein assay and Simple Western™ (automated western blotting) were performed on a JESS Simple Western™ instrument (ProteinSimple, San Jose, CA).[44,45] The proteins were separated by capillary electrophoresis over 30 minutes using the 12–230 kDa separation module. Immunostaining was completed with rabbit anti-hLF primary antibody (Novus Biologicals) and goat anti-rabbit secondary HRP conjugate (ProteinSimple, San Jose, Ca). Total protein staining was completed using biotinylated total protein dye (ProteinSimple) and streptavidin HRP (ProteinSimple). Total protein and simple western immunostaining were completed on the exact same samples in one experiment using RePlex functionality with chemiluminescence detection.

2.2.3 Reversed-phase HPLC determination of LF. Protein samples were analyzed using a 1290 UHPLC with a diode-array detector and an AdvanceBio RP-mAb C4 column (2.1 × 75 mm, 3.5 μm, Agilent, Santa Clara, CA). Mobile phases were (A) water with 0.1% trifluoroacetic acid and, (B) acetonitrile (ACN) with 0.09% trifluoroacetic acid. The gradient increased from 25% (B) to 35% (B) in 2 min, then to 45% (B) in 5 min and 55% (B) in 0.5 min, back to 25% (B) in 0.5 min, and was kept at 25% (B) for 2 min to re-equilibrate the column. The flow rate was 0.5 mL min^{-1}; the column temperature was set at 70 °C; and detection was set at 214 nm and 280 nm. The calibration standard was hmLF protein diluted in 50 mM Tris (pH 8.0). Peak area, concentration, peak retention time, peak area percentage, and calibration curve parameters were analyzed using ChemStation software.

2.2.4 Liquid chromatography and tandem mass spectrometry (LC-MS/MS)

2.2.4.1. Protein identification of rhLF with full protein sequence coverage. LF was solubilized in 1 mL Tris HCl buffer (pH 8.0). Samples were diluted 3 × 10 μg in 25 mM ammonium carbonate and incubated with 10 mM dithiothreitol at 60 °C, followed by 50 mM iodoacetamide at room temperature. Each aliquot of the samples was digested with trypsin, chymotrypsin, and elastase enzymes (Promega, Madison, WI, USA), separately, at 37 °C for 18 h, quenched with formic acid, and desalted using solid-phase extraction.

Protein digests were subjected to analysis by nano-flow LC-MS/MS on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled with a Waters™ M-class HPLC system (Milford, MS, USA). Peptides were loaded onto a trapping column and eluted over a 75 μm analytical column at 350 nL min^{-1}; both columns were packed with Luna C18 resin (Phenomenex, Torrance, CA, USA). A 30 min gradient was applied. The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed at resolutions of 60 000 and 15 000 full width half maximum, respectively. Advanced peak determination was enabled. The instrument was run using a 3-second cycle for MS and MS/MS.

The acquired MS data was processed using a local copy of the Byonic™ MS/MS search engine (Protein Metrics, Cupertino, CA, USA) with the following parameters: trypsin or none was set as the enzyme; a UniProt library containing sequencing information for *K. phaffii* and hLF was generated for protein sequencing; carbamidomethylation of cysteine (C) was specified as the fixed modification; and oxidation of methionine (M), acetylation of protein N-terminals, deamidation of asparagine (N) and glutamine (Q), phosphorylation of serine (S), threonine (T), and tyrosine (Y). Additional identification parameters included monoisotopic mass values with a peptide mass tolerance of 10 ppm, fragment mass tolerance of 20 ppm, and max missed cleavages of two. Byonic mzid files were parsed into the Scaffold software version 5.3.0 to validate, filter, and create a nonredundant list for each sample. Data was filtered using criteria that included a minimum protein value of 99%, a minimum peptide value of 50% (Prophet scores), and a minimum of two unique peptides per protein.

2.2.4.2 Site-specific glycosylation. Similar LC-MS/MS-based proteomics analyses were performed to determine site-specific glycosylation of rhLF proteins using previously established methods with minimal modification.[46–48] Briefly, an aliquot of 20 μg of rhLF proteins was reduced and alkylated and then digested with trypsin, followed by a cleanup by solid phase extraction using SOLA HRP 10 mg × 1 ml cartridges. Resulting protein digests were analyzed using an Orbitrap Fusion Lumos Tribrid mass spectrometer coupled with a Dionex UltiMate 3000 RSLCnano system (Thermo Fisher Scientific). All MS and MS/MS raw spectra from each sample were searched using Byonic™ MS/MS search engine against the *K. phaffii* NCBI database with added targeted protein hLF sequence. The peptide search parameters were as follows: two missed cleavages for full trypsin digestion with fixed carbamidomethyl modification of cysteine and variable modifications, including methionine oxidation and deamidation on asparagine/glutamine residues. Mass tolerance for peptides was set at 10 ppm, and fragments generated by HCD and EThcD at 0.05 Da and 0.6 Da, respectively. The maximum number of common and rare modifications was set at two. The glycan search was performed against a list of 309 mammalian N-linked glycans. Identified peptides were filtered for a maximum 1% false discovery rate.

The glycan occupancy rate of each site was calculated by dividing the total peak areas of the glycopeptides by the total peak areas of all peptides, including the identified glycopeptides and their respective non-glycosylated peptides as follows:

$$\text{Occupancy rate} = \frac{\text{total peak area of glycopeptides}}{\text{total peak area of all peptides}}$$

This provided an estimated occupancy rate and assumed the glycopeptides and their respective non-glycosylated peptides produced MS signals with similar efficiency.

2.2.5 Released N-glycan analysis by HPLC with fluorescence and mass spectrometry. N-glycans were released, labeled, and isolated from intact protein using the Agilent GlyX InstantPC kit according to the manufacturer's instructions (Agilent application note 5994-3482EN). Samples were chromatographically separated on an AdvanceBio glycan mapping column (300 Å, 2.1 × 150, 1.8 μm), detected by a fluorescence detector (excitation: 285 nm, emission: 345 nm), and further analyzed with an Agilent 6545XT AdvanceBio LC/Q-TOF mass spectrometer equipped with a high-resolution ion mobility (HRIM) device (MOBILion Systems, Inc., Chadds Ford, PA, USA). Alternatively, an Agilent 1290 Infinity UHPLC with a fluorescence detector was employed to record the glycan profiles. Mobile phases were 100 mM ammonium formate, pH 4.5 (A) and ACN (B). Glycan separation was carried out with a gradient from 80% to 60% (A) over 32 min at a constant elution rate of 0.5 mL min^{-1} and column temperature of 35 °C. Glycans were identified by matching their observed masses with the theoretical masses of 201 known InstantPC-labeled N-glycans with a mass tolerance of 10 ppm. The glycan identification was further verified by the HPLC retention time using IntantPC-labeled glycan standards whenever available. Agilent OpenLab ChemStation™ software produced a sequence report that included glycan identity, summary (retention time and peak area), and chromatograms.

2.2.6 Microfluidic modulation spectroscopy (MMS). The secondary structure (*i.e.*, folding of the protein backbone) of LF was measured using microfluidic modulation spectroscopy (MMS) with an AQS^3pro instrument from RedShiftBio (Boxborough, MA, USA). Dialysed samples in 50 mM Tris (pH 7.5) were loaded in triplicate into a 24-well plate in sample/buffer pairs and were run on the AQS^3pro MMS system that was equipped with sweep scan functionality and a flow cell with a 22.3 μm path length. An average of eight spectra were collected with a 5 psi backing pressure scan across the amide I band between 1588 and 1712 cm^{-1}. All samples were analyzed using the AQS^3pro delta Data Analysis software (version 2.5.98.4131). Lysozyme was used as the model protein. The samples were fit to the region between 1720 and 1680 cm^{-1}. The nominal displacement factor was set to 0.6, the second derivative smoothing window was set to 19 cm^{-1}, and the baselining setting (subtraction) was "Rubberband". A higher-order (secondary) structure was calculated using "Globular protein" pre-settings for the Gaussian curve deconvolution.

2.2.7 Iron saturation. The iron content of the rhLF and hmLF protein samples was measured according to the AOAC-2015.01 method. Iron saturation was calculated with the following formula:[49]

$$\text{Iron saturation} = 2^{\star}\ \frac{\text{mole(Fe)}}{\text{mole(rhLF)}}$$

Mole (Fe) and Mole (LF) were the moles of iron atoms and LF, respectively. They were calculated using the measured iron content and LF protein content with a MW of 56 for iron and 80 000 for LF.

2.2.8 Small-angle X-ray scattering (SAXS). Lactoferrin samples were prepared by dialysis in 50 mM Tris with 150 mM NaCl (pH 7.5) using 10 000 Da MW cut-off Slide-A-Lyzers™ (ThermoFisher). Lactoferrin protein samples were loaded onto a quartz capillary flow cell at 4 °C and precisely aligned with the X-ray beam under vacuum conditions (<1 × 10^{-3} Torr). Small-angle X-ray scattering (SAXS) data was gathered in triplicate at concentrations of 1.0 and 4.0 mg mL^{-1}. A Rigaku BioSAXS-2000 nano SAXS Kratky camera system (The Woodlands, TX, USA) was used to conduct the experiment. A Rigaku MM007 rotating anode generated X-rays with a wavelength of 1.5418 Å. The system was equipped with OptiSAXS confocal max-flux optics with a HyPix-3000 Hybrid Photon Counting detector. The useful q-space range, where q represented the scattering vector ($4\pi \sin\theta/\lambda$, with 2θ being the scattering angle), typically spanned from q_{min} = 0.008 $Å^{-1}$ to q_{max} = 0.6 $Å^{-1}$. The X-ray beam had an energy of 8.04 keV, with a Kratky block attenuation of 22% and a beam diameter of approximately 100 μm. A total of six 10 min images were collected and averaged. Obtained SAXS data underwent buffer subtraction, resulting in the final dataset we used for subsequent analysis. Data processing, including image integration, normalization, and background buffer data subtraction, was executed using Rigaku SAXSLAB software.

2.2.9 Analytical Ultracentrifugation (AUC). AUC analysis was carried out with an Optima multiwavelength instrument (Beckman Coulter, Indianapolis, IN) equipped with absorbance and interference optics. Samples were equilibrated to 20 °C. The instrument was operated under a full vacuum and the An-50 titanium rotor was spun at 30 000 rpm for 16.5 h; radial scans of the cells were recorded every two min at 280 nm. AUC data was analyzed with UltraScan III.[50,51] Reference scans were automatically selected to convert the data from raw radial intensity to pseudo-absorbance. The air-liquid meniscus was then manually selected for each sector. The data range was also manually selected (usually between 6.0 cm and 7.1 cm). The first five to 10 scans were excluded from analysis, as were scans at the end with little to no sedimentation signal. Data was fitted with an *S*-value range between 1 and 10, with a resolution of 100, a frictional ratio range of 1 and 4, and a resolution of 64. Time invariant noise was fitted during the first two-dimensional spectral analysis. When residuals were <0.003, data was re-fitted for time and radially invariant noise, and the meniscus was adjusted. When the correct meniscus was identified, an additional time and radial invariant noise fit was performed using an iterative fitting method with a maximum of 10 iterations. The resulting pseudo-three-dimensional plots were evaluated for final *S*-values, frictional ratios, and MW calculations.

2.2.10 Differential scanning calorimetry (DSC). Differential scanning calorimetry (DSC) was carried out using MicroCal VP-capillary DSC (Malvern Panalytical, Westborough, MA, USA). Data was collected over a temperature range of 25 °C to 105 °C with a scan speed of 90 °C h^{-1}. A buffer reference experiment was conducted for baseline subtraction. The cell was loaded with 400 μL of sample at a concentration of

2.0–5.0 mg mL^{-1} in the buffer 50 mM tris at pH 8.0. The running software was VPViewer2000. Raw data was converted into molar heat capacity using Microcal, LLC Cap DSC Version Origin70-L3.

3. Results and discussion

The objective of this study was to conduct a thorough characterization and analysis of Helaina rhLF protein produced at an industrial scale using a *K. phaffii* fermentation process. Various analytical techniques were employed to confirm identity, assess purity and impurities, measure size, and investigate the structure of Helaina rhLF protein, including glycosylation (Table 1). Additional comparisons to hmLF and bLF samples were made.

3.1 Protein identity

Helaina rhLF was designed according to the protein ID of hLF, specifically TRFL_HUMAN (UniProt P02788) of the UniProt database. Like mature hmLF, Helaina rhLF did not have the N-terminal signal peptide (*i.e.* amino acid sequence 1–19), as shown in the theoretical sequence (Fig. 2A). Helaina rhLF had a total of 691 amino acid residues with the N-terminal starting with the residue glycine (G)-19 and C-terminal ending with lysine (K)-710.

The nanoflow LC-MS/MS-based bottom-up proteomics approach was used to determine the amino acid sequence of Helaina rhLF. In this study, we used trypsin, chymotrypsin, and elastase to maximize the sequence coverage.[52,53] Trypsin specifically cleaved at the C-terminal side of lysine and arginine residues, while chymotrypsin and elastase cleaved non-specifically at the hydrophobic amino acid residues alanine, isoleucine, leucine, and valine. The combination of trypsin, chymotrypsin, and elastase produced complementary peptide sequences that covered the entire sequence of Helaina rhLF.

Table 1 Analytical techniques exploited to analyze and characterize Helaina rhLF as well as native hmLF

Analytical technique	Attributes of protein
Total Protein Assay and Simple Western	Molecular weight/identity
Reversed-phase high performance liquid chromatography (RP-HPLC)	Protein content and purity
HPLC-fluoresence detection	Released glycan profile
ICP-MS	Iron content and iron saturation (in conjunction with HPLC data)
Mass spectrometry (including LC-MS and high-resolution ion mobility)	Protein sequence Posttranslational modifications Glycosylation Host cell proteins
Microfluidic modulation spectroscopy (MMS)	Secondary structure
Analytical ultracentrifugation (AUC)	Tertiary structure and aggregation/degradation
Small angle X-ray scattering spectroscopy (SAXS)	Tertiary structure and aggregation/degradation
Differential scanning calorimetric (DSC)	Structure and thermostability

Proteomics analysis led to the identification of Helaina rhLF as TRFL_HUMAN with 100% sequence coverage (Fig. 2B). Protein identification was based on detecting 1522 unique peptides as well as over 2780 unique MS/MS spectra generated by multiple enzymatic digestions, in conjunction with the stringent criteria for peptide sequence identification as described in the Materials and Methods section (see ESI Table S1† for the details of the identified peptides). The data confirmed that the amino acid sequence of Helaina rhLF was identical to that of the hmLF. Additionally, total protein assay and Simple Western™ data showed that Helaina rhLF had a similar MW to that of hmLF and was positively detected with anti-hLF antibody (Fig. 3), further confirming that the size and identity of Helaina rhLF were identical to hmLF.

3.2 Protein purity

Helaina rhLF produced from the *K. phaffii* recombinant system underwent multiple isolation steps by microfiltration/diafiltration followed by cation exchange chromatography as the final purification step. The purity of Helaina rhLF as a percentage (%) of protein was assessed by multiple analytical techniques including total protein assay and Simple Western™, RP-HPLC with UV detection, and nanoflow LC-MS/MS-based proteomics approach.

Fig. 3 shows a representative image of total protein assay and Simple Western™, which is an equivalent to automated SDS-PAGE and western blotting, respectively.[44,45] The quantitative analysis using the peak area indicated the main lactoferrin peak at ~90 kDa accounted for 100% of peak area for total protein assay and 98% of Simple Western™ stain, respectively. The rest 2% peak area of the Simple Western™ stain was attributed to a band appearing at ~27 kDa, which was an internal standard included in the assay. RP-HPLC analysis provided an orthogonal measurement of Helaina rhLF quantification and purity. As shown in Fig. 4, Helaina rhLF was detected as the predominant peak along with a couple of minor peaks at both wavelengths of 280 nm (more specific to protein) and 214 nm (less specific to protein). It should be noted that the splitting peaks of the rhLF represented its glycoforms with different levels of glycosylation. The average purity of three batches of rhLF was 97.6% with a small relative standard deviation of 0.4%, based on peak area percentage calculations obtained at 214 and 280 nm. Further LC-MS/MS-based proteomics analysis of the Helaina rhLF revealed (identified) a total of 23 proteins, with human LF being the predominant protein along with 22 contaminating proteins (ESI Table S2†). No residual host cell proteins (HCP) had a relative abundance greater than 0.5%, with the top three identified HCP being phosphoglycerate kinase (UniProt i.d. C4QY07, 0.49% adjusted relative abundance), endoplasmic reticulum chaperone BiP (UniProt i.d. C4QZS3, 0.46% adjusted relative abundance), and MannosidaseTR (0.36% adjusted relative abundance). It should be noted that MannosidaseTR is not a residual protein from the yeast expression system but rather a protein introduced in the engineering process to help minimize the biosynthesis of hyper-mannosylation by yeast.

(A) Theoretical sequence

```
    1         11         21         31         41         51
    |         |          |          |          |          |
  1                   G RRRSVQWCAV SQPEATKCFQ WQRNMRKVRG PPVSCIKRDS    60
 61 PIQCIQAIAE NRADAVTLDG GFIYEAGLAP YKLRPVAAEV YGTERQPRTH YYAVAVVKKG   120
121 GSFQLNELQG LKSCHTGLRR TAGWNVPIGT LRPFLNWTGP PEPIEAAVAR FFSASCVPGA   180
181 DKGQFPNLCR LCAGTGENKC AFSSQEPYFS YSGAFKCLRD GAGDVAFIRE STVFEDLSDE   240
241 AERDEYELLC PDNTRKPVDK FKDCHLARVP SHAVVARSVN GKEDAIWNLL RQAQEKFGKD   300
301 KSPKFQLFGS PSGQKDLLFK DSAIGFSRVP PRIDSGLYLG SGYFTAIQNL RKSEEEVAAR   360
361 RARVVWCAVG EQELRKCNQW SGLSEGSVTC SSASTTEDCI ALVLKGEADA MSLDGGYVYT   420
421 AGKCGLVPVL AENYKSQQSS DPDPNCVDRP VEGYLAVAVV RRSDTSLTWN SVKGKKSCHT   480
481 AVDRTAGWNI PMGLLFNQTG SCKFDEYFSQ SCAPGSDPRS NLCALCIGDE QGENKCVPNS   540
541 NERYYGYTGA FRCLAENAGD VAFVKDVTVL QNTDGNNNEA WAKDLKLADF ALLCLDGKRK   600
601 PVTEARSCHL AMAPNHAVVS RMDKVERLKQ VLLHQQAKFG RNGSDCPDKF CLFQSETKNL   660
661 LFNDNTECLA RLHGKTTYEK YLGPQYVAGI TNLKKCSTSP LLEACEFLRK
```

(B) Detected sequence from Helaina rhLF

TRFL_HUMAN (100%), 76,165.4 Da
TRFL_HUMAN
1522 exclusive unique peptides, 2783 exclusive unique spectra, 5033 total spectra, 691/691 amino acids (100% coverage)

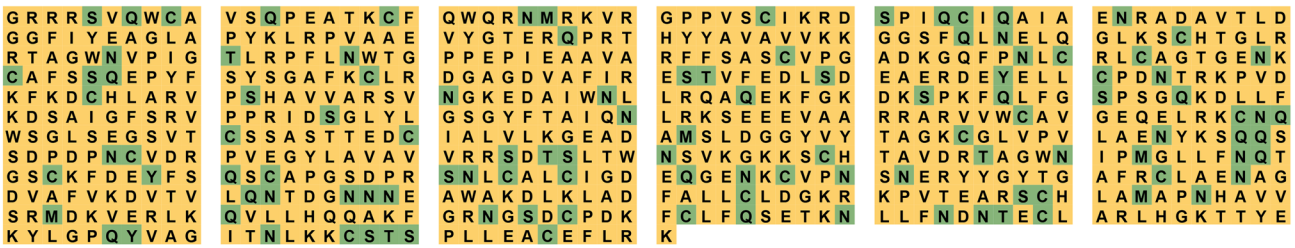

Fig. 2 (A) Protein sequence of native hmLF (Uniprot P02788) without the N-terminal signal peptide (amino acid sequence 1–19). The highlighted N indicates the three known N-glycosylation sites, namely, Asn-156, -497 and -642. (B) Detected sequence of native hmLF from Helaina rhLF as highlighted by the yellow and green color. The green color indicates posttranslational modifications.

3.3 Glycosylation

Two complementary analytical techniques; site-specific glycosylation analysis and released glycan profiling, were employed to gain a full understanding of the glycosylation of Helaina rhLF protein.

Site-specific glycosylation properties of LF were revealed by the LC-MS/MS-based proteomic analysis. N-glycans were detected at all three predicted asparagine (Asn) sites as described for the hmLF, Asn-156, Asn-497 and Asn-642.[11,54] In Helaina rhLF, all three asparagine sites share the same set of N-glycans- oligomannose N-glycans containing 5 to 12 mannose (hexose) residues (M5–M12) (ESI Table S3†). No complex or hybrid types of N-glycans were identified. The occupancy rates for the three sites, however, were different and none of the three sites were completely occupied by glycans. The occupancy rate of three batches of Helaina rhLF was 75% ± 7% for the site Asn-156; 80% ± 15% for Asn-497; and 18% ± 6% for Asn-642, respectively. These observations were similar to the results reported by Zlatina *et al.*, where there were approximately 90% occupancy rates for both Asn-156 and Asn-497, but only 9% for Asn-624, respectively.[12]

Fig. 5A shows a representative chromatogram of the released N-glycan profile of Helaina rhLF protein. Same as the results from the site-specific analysis, the N-glycans identified by mass spectrometry analysis were primarily oligomannoses M5–M14 (ESI Table S4†). Table 2 lists the relative abundance of the identified glycans, estimated based on their peak areas. The smaller oligomannose M5–M9 accounted for about 59 ± 2% of total identified N-glycans while the relatively larger oligomannose M10–M14 only took 41 ± 2% (Table 2). Interestingly the mass spectrometry data (ESI Table S4†) also suggested there were about 23% phosphorylated M5–M14. The phosphomannose glycans present on Helaina rhLF are the same oligomannose structures (M5–M14) with one phosphate group. Phosphomannose structures have been observed in

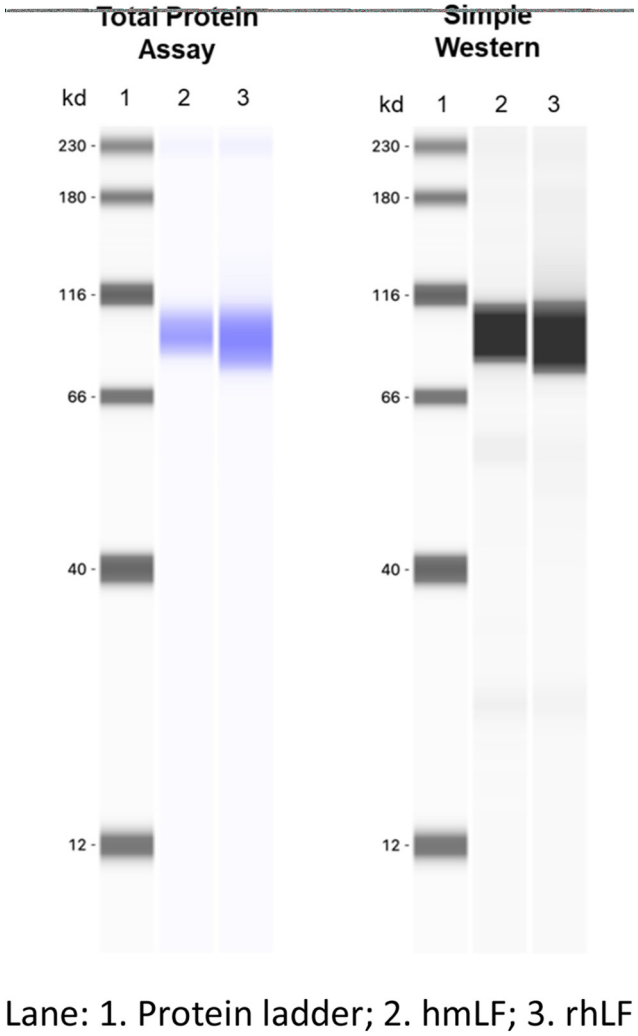

Fig. 3 Representative images of total protein assay and Simple Western™ of native hmLF (Lane-2) and Helaina rhLF (Lane-3) at the concentration of 100 μg mL^{-1}. Other conditions are described in Materials and Methods.

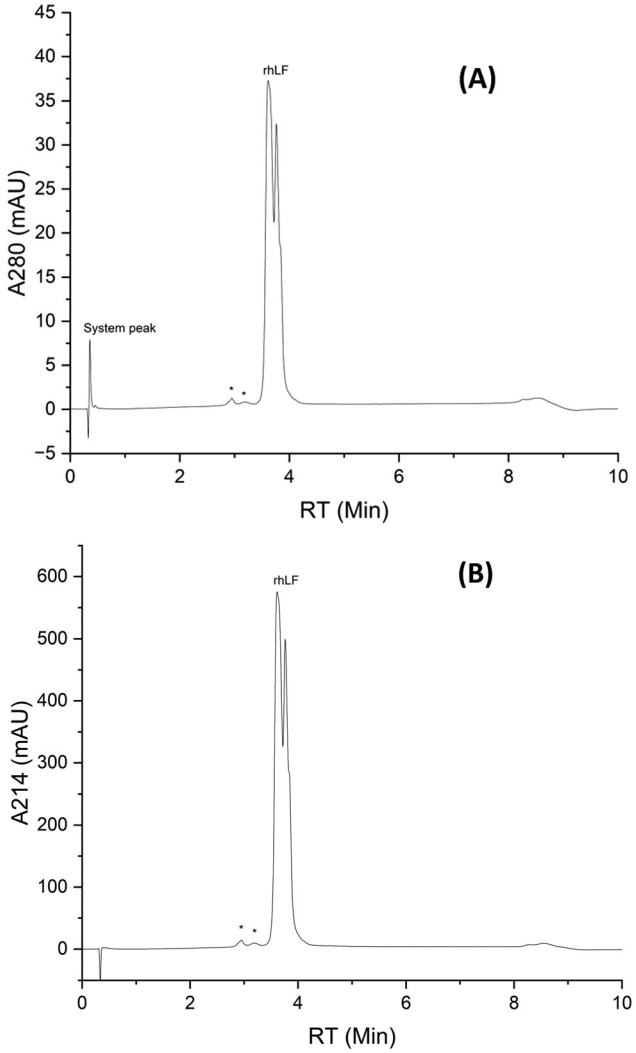

Fig. 4 Representative chromatograms of RP-HPLC analysis of Helaina rhLF with UV detection at 280 nm (A) and 214 nm (B). *Indicates impurity peaks observed from the rhLF samples. The chromatograms were generated from 1.0 μL injection of 1.0 mg mL^{-1} Helaina rhLF. Other conditions are described in Materials and Methods.

numerous recombinant glycoproteins produced in *K. phaffii*, in abundances ranging from 20% to 30%, although their presence depends on the protein being expressed.[55–58] This range of abundance coincides with that measured in Helaina rhLF. The results of the N-glycan profiles of Helaina rhLF were highly consistent across three different batches with an average relative standard deviation (RSD) of the relative abundance of 9% and a range of 2%–12% (Table 2 and Fig. S1†).

Fig. 5B shows that the smaller oligomannoses M5–M9 were detected in bovine milk LF at high abundance along with complex and hybid glycans. It should also be pointed out that M4–M9 have been reported in hmLF by different labs[16,59] In our study, we consistently detected M5–M6 in hmLF from different sources including those supplied by MilliporeSigma and Abcam as well as purified from breast milk by Helaina (Fig. 6 and S2†).

3.4 Protein structures

3.4.1 Secondary structure by MMS. The secondary structure of Helaina rhLF was assessed using MMS and then compared to secondary structure of hmLF and bLF. MMS is a novel infrared (IR) technology that has been gaining increasing interest for protein structural analysis and similarity comparison, due to its simplicity, sensitivity, and automation capabilities.[60,61] MMS monitors the vibrational response of the amide I band at 1600–1700 cm^{-1}, representing the stretching of the carbonyl group of the amide bonds. The amide I vibrational signals are sensitive to the H bonding environment and the folding of the protein backbone (the protein secondary structure).

Fig. 7A shows representative second derivative MMS spectra of Helaina rhLF, hmLF, and bLF. They all comprise the same types of secondary structure features, namely α-helix, β-sheet,

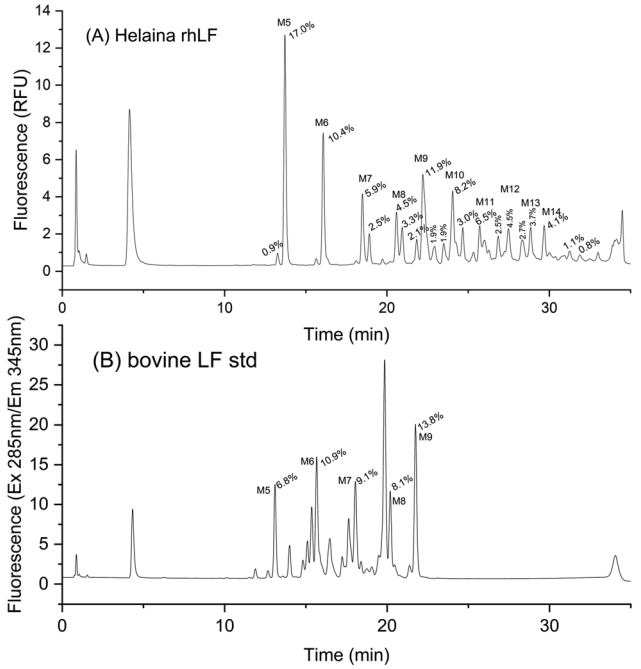

Fig. 5 Comparison of the released N-glycan profiles of Helaina rhLF (A) and bLF (B). Conditions for data acquisition and processing are detailed in the Material and Methods.

Table 2 The average relative abundance (%) and reproducibility of the N-glycans identified from three batches of Helaina rhLF

N-Glycan	Average abundance (%)	Standard deviation (%)	RSD
M5	15.6	2.3	15%
M6	11.6	1.6	14%
M7	9.0	0.8	9%
M8	7.3	0.5	6%
M9	15.8	0.1	1%
M10, 11, 12, 13 and M14	40.8	2.0	5%

RSD: relative standard deviation.

and turn, as evident by the observations of the distinct absorbance peaks at 1656, 1634, and 1686 cm^{-1}.[62,63] Helaina rhLF's MMS spectrum overlaps exactly with that of hmLF but not that of bLF. Fig. 7B further compares the compositions of the secondary structures of Helaina rhLF (averaged from three batches) with hmLF (averaged from three hmLF from three sources). Helaina rhLF comprises 38% ± 1% α-helices, 28% ± 1% β-sheets, and 23% ± 1% turns *versus* hmLF's 40% ± 1% α-helices, 29% ± 1% β-sheets, and 23% ± 1% turns. This data suggests Helaina rhLF secondary structure was nearly identical to that of hmLF. The MMS results of Helaina rhLF and hmLF closely match previously published data for native hmLF as measured. The MMS results of Helaina rhLF and hmLF closely match previously published data for native hLF as measured by circular dichroism (CD); where 37.4% α-helices, 28.2% β-sheets, and 15.8% turns were reported.[64] Additionally, the CD measurements of rhLF expressed in transgenic mice showed similar results with 40% α-helices, 20% β-sheets, and 18% turns.[65]

Similarity scores[66] are useful to quantify the degree of structural similarity between proteins. We computed similarity scores by normalizing the overlap of the second derivative MMS spectra of a LF protein (hmLF or Helaina rhLF) with the hmLF from MilliporeSigma, according to the approach proposed by Kendrick *et al.*[66] The mean similarity score was 98.2% ± 1.6% (range 96%–100%) for hmLF from three different sources, and 96.2%% ± 1.4% (range 95%–98%) for three Helaina rhLF batches, respectively. The hmLF from MilliporeSigma had a similarity score of 100% since it was used as the reference for the similarity score computation for all the LF. Nevertheless, the range of the similarities score of hmLF and Helaina rhLF overlapped and the difference in the means was only 2.0%. This provides further support that Helaina rhLF's MMS secondary structures were nearly identical to those of hmLF. The nearly identical secondary structures between Helaina rhLF and hmLF can be attributed to their identical amino acid sequences as determined by the proteomics analysis, since the secondary structure of a protein is predominantly determined by its primary structure.

3.4.2 Iron saturation. As an iron-binding protein, each LF molecule has two lobes, each capable of binding one iron atom. Iron saturation reflects the percentage of two lobes filled with iron atoms. The mean (standard deviation) iron saturation rate of three batches of Helaina rhLF was 54% (±4%) and ranged between 50% and 60%. These results indicate that about one-half to two-thirds of the lobes of Helaina rhLF contained iron. By way of comparison, hmLF is typically the apoform, meaning its saturation was under 20%.[67]

3.4.3 Tertiary structure by SAXS and AUC. We used SAXS to determine and compare the tertiary structure of Helaina rhLF to that of hmLF and bLF. The SAXS analysis reveals information about the size, shape, and conformational flexibility of proteins in solution.[68–71] We then utilized the reported crystallographic structures of hmLF to validate the accuracy of the molecular envelopes reconstructed from SAXS. Additionally, we employed AUC to further validate the structural information obtained by SAXS.

SAXS raw data (Fig. S3†) for Helaina rhLF, hmLF, and bLF was acquired at 1.0 mg mL^{-1} (A) and 4.0 mg mL^{-1} (B) concentrations. The SAXS data sets were collected for 60 min, with six 10 min images for each concentration. The overlays of these images were averaged and showed there was no X-ray radiation damage. Replicate datasets were also averaged. A comparison of the SAXS data between low and high concentrations showed near-identical signal profiles and no concentration-dependent multimerization. As expected, compared to the lower concentration, the higher concentration produced a greater signal to noise SAXS data.

Raw data (Fig. S3†) was reconstructed into Guinier plots (Fig. S4†), Pair distance distribution function $P(r)$ plots (Fig. S5†), and Kratky plots (Fig. S6†), molecular envelopes (Fig. 8) to gain the understanding of the characteristics of the

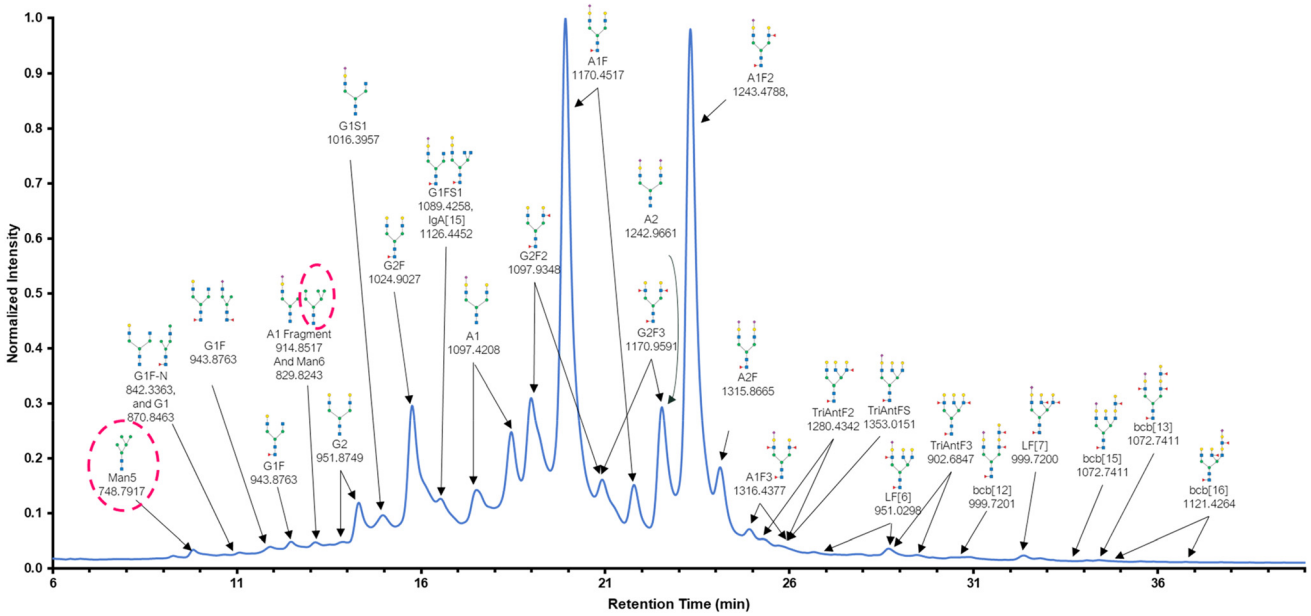

Fig. 6 Representative LC-fluorescence chromatogram of the released N-glycan profile of native hmLF. The peak assignments were supported by a LC-HRIM-MS trace with good alignment between the peaks. Each of the peaks is labeled with the *m/z* value(s) of the most abundant species observed and the best possibly matched N-glycan(s). The two red circles indicate the detection of oligomannoses M5 and M6, which are also present in Helaina rhLF. Other conditions are described in Materials and Methods.

size, shape, and conformation of the LF proteins. The results are summarized in Table 3.

The plots of Guinier fit (Fig. S4†) and pair distance distribution function $P(r)$ (Fig. S5†) display the overall size and multimerization/aggregation of the LF proteins in solution (50 mM tris with 150 mM sodium chloride, pH 7.5 for our analysis). The Guinier fit determined the radii of gyration (R_g) of LF proteins (Table 3). No aggregation was observed for all tested LF proteins, except for the hmLF-A (hmLF from Abcam) and bLF-T (bLF from The Lactoferrin Company) proteins. As can be seen, Helaina rhLF and the hmLF proteins, except hmLF-A, had R_g had values between 32 and 35 Å, which was in agreement with the size of a protein monomer from the X-ray crystal structure model of hmLF (PDB codes 2BJJ and 1LFH). Fig. S4† shows a "smiley Guinier" fit with a radius of gyration, R_g of 42.8 (±0.3) Å for the hmLF-A and 110.7 (±0.6) Å for bLF-T, indicating protein aggregation. Fig. 9 is a comparison of the R_g of Helaina rhLF and native hmLF proteins, showing similar average R_g with an overlapping distribution. The $P(r)$ plot (Fig. S5†) is a histogram of the distances between pairs of points within the LF proteins in solution. R_{max}, where the $P(r)$ curve intersects the X-axis, is the maximum diameter of the proteins. All Helaina rhLF and hmLF and bLF, except hmLF-A and bLF-T, appeared to be globular and peaked around 35 Å with a diameter of 128 Å in the range of the R_g values from the Guinier analysis. The hmLF-A and bLF-T proteins had broad humps following the main peak and corresponding to the large multimers with a maximum diameter over 250 Å, respectively. These results indicate that both Helaina rhLF and hmLF were present as monomers in aqueous solution and with similar overall size.

Kratky plots derived from SAXS data provided a further qualitative assessment of the flexibility and/or degree of unfolding in the tested LF protein samples (Fig. S6†). Proteins that are unfolded or highly flexible typically exhibit a plateau in the Kratky plot at high scattering vectors (s), while compact, globular proteins display a bell-shaped Gaussian peak. The Kratky plots we generated indicate that all tested human and bovine LF were well-folded globular proteins with no discernible flexibility or disorder.

Fig. 8 illustrates the solvent envelopes of Helaina rhLF and hmLF derived from SAXS data. These envelopes closely matched the reported X-ray crystal structures of both Fe^{+3} bound (holo) and iron-free (apo) hmLF monomers (2BJJ and 1LFH in the RCSB PDB database, respectively). The fits were rigorously validated through refinement using normal mode analysis in Sreflex, leveraging SAXS data. The Pymol-generated representation incorporated a transparent surface overlying the electron density from solution scattering (DENSS), the latter being an advanced algorithm for direct computation of *ab initio* electron density maps using solution scattering data.

The SAXS refinement, employing the SASREF method, successfully refined the N and C-lobes of LF individually, resulting in improved Chi-square values and better fits into the DENSS envelopes. As indicated by the SASREF Chi-square values (Table 3), all Helaina rhLF and hmLF had a better fit to the holo monomer than to the apo monomer, with the exception of hmLF-S (from MilliporeSigma). An explanation for this discrepancy may be that Helaina rhLF was not in an apo form but rather 50%-60% iron-saturated, while hmLF-S was an essentially apo form with approximately 8% iron saturation.

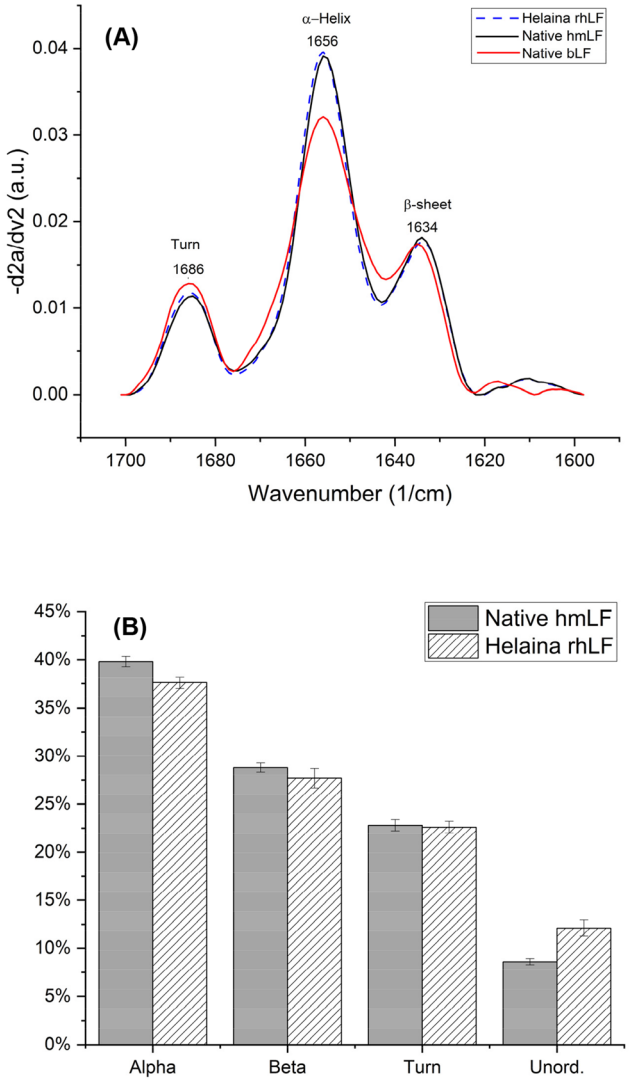

Fig. 7 Representative second derivative MMS spectra of Helaina rhLF, native hmLF and native bLF (A); and comparison of the compositions of the secondary structures of Helaina rhLF with native hmLF (B). The composition of the secondary structures of Helaina rhLF was the average from three production batches, while that of native hmLF was the average from three different sources of hmLF as described in the Material section.

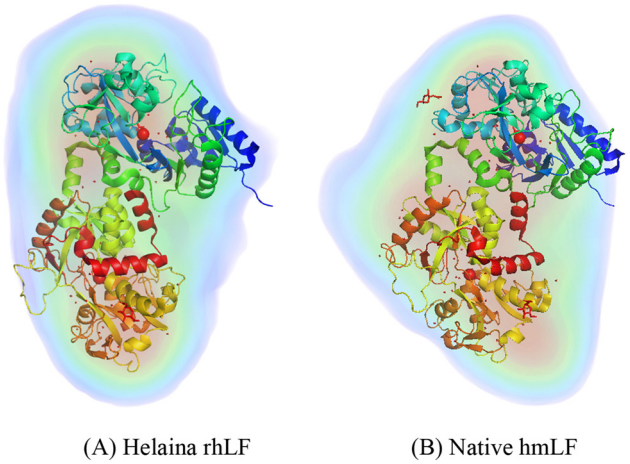

Fig. 8 Representative solvent envelopes of Helaina rhLF and native hmLF derived from the SAXS data and their fits with the Fe^{+3} bound human lactoferrin monomer (2bjj) model. Conditions for data acquisition and processing are detailed in the Material and Methods.

The AUC results (Fig. 10 and S7†) also demonstrated the consistency of Helaina's rhLF industrial-scale production process: Fig. S7A–C,† showing the three batches of rhLF in comparison to hmLF-S (from MilliporeSigma, Fig. S7D†), and hmLF-H (purified in house, Fig. S7F†). The hmLF-A (from Abcam) showed multiple species, suggesting more heterogeneity than the recombinant batches or the other native samples (Fig. S7E†). Each recombinant batch had highly similar frictional ratios of 1.25–1.30, suggesting a globular protein. This correlated with the crystal structure of hmLF and with the SAXS results. Furthermore, MW estimates from each batch were 76–80 kDa, and appear to be in agreement with the expected MW of native hLF (~ 80 kDa) and with the MW estimates obtained for hmLF-S (from MilliporeSigma) and hmLF-H (purified in house) (Fig. S7D and F†).

In summary, both Helaina rhLF and hmLF were confirmed as monomeric, well-folded, globular proteins. The slight variations in the shapes of their molecular envelopes can be attributed, in part, to the flexibility in the orientation of the N-lobe and C-lobe domains under various conditions, particularly in response to changes in iron saturation.

3.4.4 Thermostability by DSC. Protein unfolding under heat is primarily determined by the protein's structure, and the more tightly a protein is folded, the higher temperature that is required to unfold (*i.e.*, denature). We employed DSC to compare the thermal stability of Helaina rhLF and hmLF. Fig. 11A shows the hmLF as isolated, which comprises predominantly the apo form with a low (~8%) iron saturation, underwent denaturation (unfolding) at maximum peak temperature of approximately 67 °C. The holo-, iron-saturated (~99%) form of hmLF exhibited drastically higher thermal stability than the apo form with a much-elevated denaturation temperature of approximately 91 °C. Helaina rhLF had two denaturation temperatures, 68.7 ± 0.9 °C and 91.9 ± 0.5 °C (Table 4 and Fig. 11), due to Helaina rhLF's partial iron saturation (50–60%). In other words, Helaina rhLF was a mix of both apo- and holo-rhLF. Helaina rhLF having identical denaturation temperatures to that of apo and holo hmLF provided further evidence that they have highly similar secondary and tertiary structures.

3.5 Further and future work on safety and functions

The next steps, beyond the structural characterization described in this paper, include a food ingredient risk assessment. Unlike pharmaceuticals where both the risk and benefit are considered, new food ingredients are evaluated strictly on safety. The risk assessment will include immunogenicity potential, general safety and tolerability of the ingredient as

Table 3 Characteristics of LF proteins revealed by AUC and SAXS analyses

LF protein	f/f_0 (AUC)	Aggregated	Guinier R_g (Å)	$P(r)$ R_g (Å)	D_{max} (Å)	Porod volume (Å^3)	SASREF Chi-sq fit to holo-hmLF	SASREF Chi-sq fit to apo-hmLF
Helaina rhLF								
Mean	1.32	No	41.3	41.5	161.0	1.3×10^5	3.2	3.4
STDEV	0.04		13.37	13.59	83.71	3.6×10^4	1.79	2.33
RSD	2.7%		32.4%	32.7%	52.0%	26.7%	56.4%	67.6%
Native hmLF								
hmLF-S	1.31	No	35.4	35.5	122.5	1.2×10^5	3.1	2.9
hmLF-A	1.29	Yes	56.6	57.1	257	1.8×10^5	5	6
hmLF-H	1.36	No	31.9	32	103.4	1.1×10^5	1.42	1.44
Mean	1.32		41.3	41.5	161.0	1.3×10^5	3.2	3.4
STDEV	0.04		13.37	13.59	83.71	3.6×10^4	1.79	2.33
RSD	2.7%		32.4%	32.7%	52.0%	26.7%	56.4%	67.6%

The data of Helaina was calculated from the three batches of production runs. hmLF-A, -S, and -H are native hmLF from Abcam, MilliporeSigma, and Helaina, respectively. Conditions for data acquisition are detailed in the Material and Methods. STDEV: standard deviation. RSD: relative standard deviation.

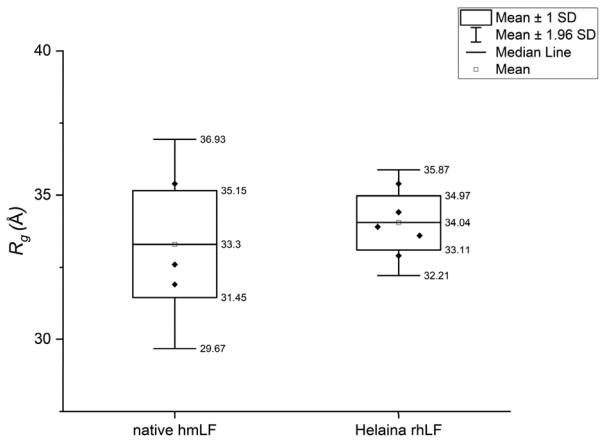

Fig. 9 Comparison of the radii of gyration (R_g) of the Helaina rhLF proteins with native hmLF proteins. Conditions for data acquisition and processing are detailed in the Material and Methods.

well as the and immunotoxicity potential of Helaina rhLF using cellular and animal preclinical models, respectively.[72] In addition, an allergenicity assessment using bioinformatics and pepsin digestion has be carried out,[73] and finally, a clinical study will evaluate the immunogenicity potential of Helaina rhLF along with iron homeostasis endpoints. These safety studies aim to meet the US FDA safety standard of reasonable certainty of no harm.[74]

Multiple functional assays have been carried out or planned. Firstly, this study showed that Helaina rhLF had affinity for binding to irons (section 3.4.2), an important function of LF. Helaina rhLF underwent denaturation at almost the same temperatures as did the apo- and holo-native hmLF as determined by DSC (section 3.4.4). These data indicate Helaina rhLF has a similar affinity for binding to iron compared to native hmLF. This study also demonstrated Helaina

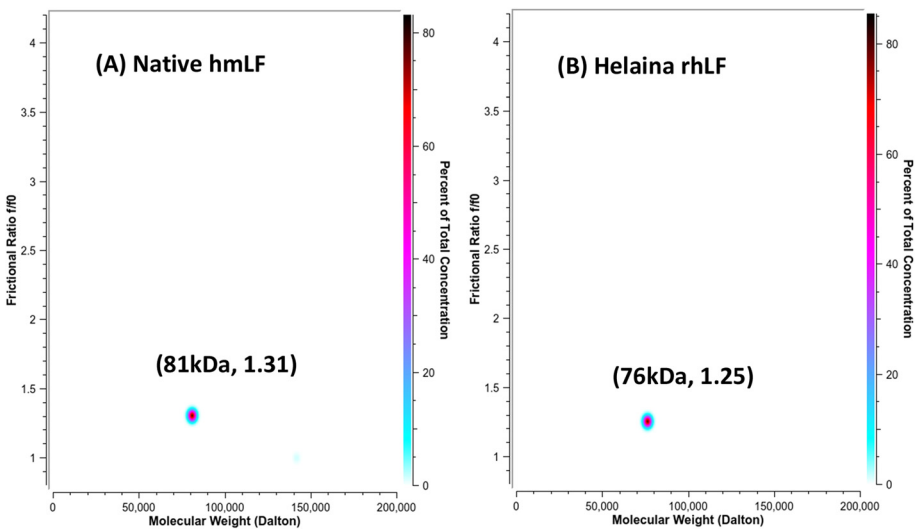

Fig. 10 AUC measurements of native hmLF (A) and Helain rhLF (B). The data presented a pseudo-three-dimensional distribution of the observed protein species, with the calculated molecular weight on the *x*-axis, the calculated frictional ratio on the *y*-axis, and the % concentration in the *Z*-plane, with the heat map on the right axis. Conditions for data acquisition and processing are detailed in the Material and methods.

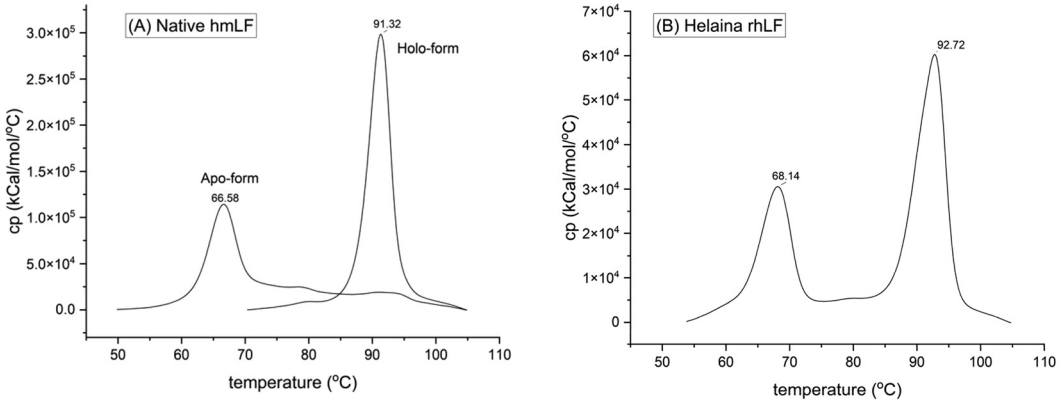

Fig. 11 Comparison of differential scanning calorimetry of native hmLF (apo- and holo-form) (A) and Helaina rhLF (B). Conditions for data acquisition and processing are detailed in the Material and Methods.

Table 4 Maximum peak temperature (T_{max}) of Helaina rhLF and native hLF

	T_{max1} (°C)	T_{max2} (°C)
Helaina rhLF	68.7 ± 0.9	91.9 ± 0.5
Native hmLF-Apo	66.4 ± 0.5	
Native hmLF-Holo		91.1 ± 0.5

rhLF was recognized and detected in the western blotting by the same antibody against native hmLF (section 3.1 and Fig. 3.). Additionally, the ability of Helaina rhLF in prohibiting the growth of bacteria will be assessed in the future *e.g.* by a minimum inhibitory concentration assay. Lastly, peripheral blood mononuclear cells exposed to Helaina rhLF or hmLF showed nearly identical cytokine production profiles, indicating similar immune-stimulating potential between the two proteins and low immunogenicity risk (manuscript in preparation).

4. Conclusion

In summary, our study presents a comprehensive analysis of industrial scale batches of Helaina rhLF (Effera™) expressed in the *K. phaffii* system. We showed that Helaina rhLF has a protein purity of 98% or higher, confirmed through RP-HPLC, SDS-PAGE, and LC-MS/MS-based proteomics. LC-MS/MS analysis verified the 100% identical amino acid sequence to hLF, with N-linked glycans detected at the same sites found on hmLF. The analysis has also identified that the N-glycans in Helaina rhLF are oligomannose structures (M5–M9), with M5 and M6 detected in hmLF samples, although complex N-glycans are the predominant ascribed glycoforms. MMS revealed that the secondary structure of Helaina rhLF closely mirrors those of hmLF, with an average 96% similarity that can be attributed to identical amino acid sequences. Additionally, data generated from the SAXS, AUC, and thermostability analyses confirmed that Helaina rhLF structure closely resembles hmLF, exhibiting well-folded globular structures and undergoing denaturation at temperatures consistent with apo- and holo-hmLF. Overall, these findings demonstrate that Helaina rhLF is of high purity and has a structure akin to hmLF.

Author contributions

X.L., C.C, A.O., B.H., P.B.B-L, R P., and A J C. are employees of Helaina Inc. N.H.Y. and K.E.W.N. are employees of the Pennsylvania State University.

Conflicts of interest

The authors declare no conflict of interests.

Acknowledgements

The authors thank Julia Fecko of the X-ray Crystallography core at the Penn State University, Huck Institutes of the Life Sciences, for her invaluable assistance in the DSC analyses. Additionally, support for small-angle X-ray scattering was provided by SIG S10 of the National Institutes of Health, USA, under award number # S10-OD028589 to N. H. Y. The Optima AUC utilized in this research was made possible by NIH grant S10 OD032215-01 to N. H. Y. The authors would also like to thank Drs Sheng Zhang and Qin Fu at the Proteomics and Metabolomics Facility of Cornell University for providing the mass spectrometry data and NIH SIG grant 1S10 OD017992-01 support for the Orbitrap Fusion mass spectrometer. The authors thank Dr Valerie Collins and Rich DiVirgilio at RedShiftBio for help in the MMS analyses and Julie Wushensky, Dr Greg Kilby, and Chuck Cloyd at MOBILion Systems, Inc for assistance in the glycan analyses. The authors want to specially thank Carolyn Alish for her assistance in writing, editing, and proofreading the manuscript.

References

1 B. Lönnerdal and S. Iyer, Lactoferrin: molecular structure and biological function, *Annu. Rev. Nutr.*, 1995, **15**(1), 93–110.

2 G. Schirmbeck, S. Sizonenko and E. Sanches, Neuroprotective role of lactoferrin during early brain development and injury through lifespan, *Nutrients*, 2022, **14**(14), 2923. Available from: **https://pubmed.ncbi.nlm.nih.gov/35889882/** [cited 2023 Apr 4].

3 B. Wang, Y. P. Timilsena, E. Blanch and B. Adhikari, Lactoferrin: structure, function, denaturation and digestion, *Crit. Rev. Food Sci. Nutr.*, 2019, **59**(4), 580–596.

4 T. J. Ochoa, E. Chea-Woo, M. Campos, I. Pecho, A. Prada, R. J. McMahon, *et al.*, Impact of lactoferrin supplementation on growth and prevalence of Giardia colonization in children, *Clin. Infect. Dis*, 2008, **46**(12), 1881–1883.

5 B. S. Berthon, L. M. Williams, E. J. Williams and L. G. Wood, Effect of lactoferrin supplementation on inflammation, immune function, and prevention of respiratory tract infections in humans: A systematic review and meta-analysis, *Adv. Nutr.*, 2022, **13**(5), 1799–1819.

6 S. M. Donovan, Human milk proteins: composition and physiological significance, in *Nestlé Nutrition Institute Workshop Series [Internet]*, ed. S. M. Donovan, J. B. German, B. Lönnerdal and A. Lucas, S. Karger AG, 2019, pp. 93–101. Available from: **https://www.karger.com/Article/FullText/490298** [cited 2022 Sep 20].

7 B. F. Anderson, H. M. Baker, E. J. Dodson, G. E. Norris, S. V. Rumball, J. M. Waters, *et al.*, Structure of human lactoferrin at 3.2-A resolution, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**(7), 1769–1773.

8 B. F. Andersen, H. M. Baker, G. E. Morris, S. V. Rumball and E. N. Baker, Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins, *Nature*, 1990, **344**(6268), 784–787.

9 B. Lönnerdal, M. K. Georgieff and O. Hernell, Developmental physiology of iron absorption, homeostasis, and metabolism in the healthy term infant, *J. Pediatr.*, 2015, **167**(4 Suppl), S8–14.

10 M. Barboza, J. Pinzon, S. Wickramasinghe, J. W. Froehlich, I. Moeller, J. T. Smilowitz, *et al.*, Glycosylation of human milk lactoferrin exhibits dynamic changes during early lactation enhancing its role in pathogenic bacteria-host interactions, *Mol. Cell. Proteomics*, 2012, **11**(6), M111.015248.

11 K. Zlatina and S. P. Galuska, The N-glycans of lactoferrin: more than just a sweet decoration, *Biochem. Cell Biol.*, 2021, **99**(1), 117–127.

12 B. Guan, Z. Zhang, Y. Chai, X. Amantai, X. Chen, X. Cao, *et al.*, N-glycosylation of milk proteins: a review spanning 2010–2022, *Trends Food Sci. Technol.*, 2022, **128**, 1–21.

13 J. W. Froehlich, E. D. Dodds, M. Barboza, E. L. McJimpsey, R. R. Seipert, J. Francis, *et al.*, Glycoprotein expression in human milk during lactation, *J. Agric. Food Chem.*, 2010, **58**(10), 6440–6448.

14 E. Goonatilleke, J. Huang, G. Xu, L. Wu, J. T. Smilowitz, J. B. German, *et al.*, Human milk proteins and their glycosylation exhibit quantitative dynamic variations during lactation, *J. Nutr.*, 2019, **149**(8), 1317–1325.

15 R. L. Valk-Weeber, T. Eshuis-de Ruiter, L. Dijkhuizen and S. S. van Leeuwen, Dynamic temporal variations in bovine lactoferrin glycan structures, *J. Agric. Food Chem.*, 2020, **68**(2), 549–560.

16 A. Le Parc, S. Karav, C. Rouquié, E. A. Maga, A. Bunyatratchata and D. Barile, Characterization of recombinant human lactoferrin N-glycans expressed in the milk of transgenic cows, *PLoS One*, 2017, **12**(2), e0171477.

17 P. Stanley, K. W. Moremen, N. E. Lewis, N. Taniguchi and M. Aebi. N-Glycans, in *Essentials of Glycobiology [Internet]*, ed. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart and M. Aebi, *et al.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 4th edn, 2022. Available from: **https://www.ncbi.nlm.nih.gov/books/NBK579964/** [cited 2023 Feb 9].

18 A. Pekarsky, L. Veiter, V. Rajamanickam, C. Herwig, C. Grünwald-Gruber, F. Altmann, *et al.*, Production of a recombinant peroxidase in different glyco-engineered Pichia pastoris strains: a morphological and physiological comparison, *Microb. Cell Fact.*, 2018, **17**(1), 183.

19 X. Wang, S. Liu, H. Xu and W. Yan, Effects of recombinant human lactoferrin on improving the iron status of IDA rats, *Weisheng Yanjiu*, 2012, **41**(1), 13–17, 22.

20 M. S. Lepanto, L. Rosa, A. Cutone, M. P. Conte, R. Paesano and P. Valenti, Efficacy of lactoferrin oral administration in the treatment of anemia and anemia of inflammation in pregnant and non-pregnant women: an interventional study, *Front. Immunol.*, 2018, **9**, 2123. Available from: **https://www.frontiersin.org/articles/10.3389/fimmu.2018.02123** [cited 2023 Nov 29].

21 R. Paesano, F. Torcia, F. Berlutti, E. Pacifici, V. Ebano, M. Moscarini, *et al.*, Oral administration of lactoferrin increases hemoglobin and total serum iron in pregnant women, *Biochem. Cell Biol.*, 2006, **84**(3), 377–380.

22 X. Zhao, X. Zhang, T. Xu, J. Luo, Y. Luo and P. An, Comparative effects between oral lactoferrin and ferrous sulfate supplementation on iron-deficiency anemia: a comprehensive review and meta-analysis of clinical trials, *Nutrients*, 2022, **14**(3), 543.

23 L. Rosa, A. Cutone, M. S. Lepanto, R. Paesano and P. Valenti, Lactoferrin: a natural glycoprotein involved in iron and inflammatory homeostasis, *Int. J. Mol. Sci.*, 2017, **18**(9), 1985.

24 P. H. C. van Berkel, M. M. Welling, M. Geerts, H. A. van Veen, B. Ravensbergen, M. Salaheddine, *et al.*, Large scale production of recombinant human lactoferrin in the milk of transgenic cows, *Nat. Biotechnol.*, 2002, **20**(5), 484–487.

25 M. Wang, Z. Sun, T. Yu, F. Ding, L. Li, X. Wang, *et al.*, Large-scale production of recombinant human lactoferrin from high-expression, marker-free transgenic cloned cows, *Sci. Rep.*, 2017, **7**(1), 10733.

26 Y. A. Suzuki, S. L. Kelleher, D. Yalda, L. Wu, J. Huang, N. Huang, *et al.*, Expression, characterization, and biologic activity of recombinant human lactoferrin in rice, *J. Pediatr. Gastroenterol. Nutr.*, 2003, **36**(2), 190–199.

27 S. Nandi, Y. A. Suzuki, J. Huang, D. Yalda, P. Pham, L. Wu, *et al.*, Expression of human lactoferrin in transgenic rice grains for the application in infant formula, *Plant Sci.*, 2002, **163**(4), 713–722.

28 P. P. Ward, J. Y. Lo, M. Duke, G. S. May, D. R. Headon and O. M. Conneely, Production of biologically active recombinant human lactoferrin in Aspergillus oryzae, *Nat. Biotechnol.*, 1992, **10**(7), 784–789.

29 P. P. Ward, C. S. Piddington, G. A. Cunningham, X. Zhou, R. D. Wyatt and O. M. Conneely, A system for production of commercial quantities of human lactoferrin: a broad spectrum natural antibiotic, *Nat. Biotechnol.*, 1995, **13**(5), 498–503.

30 B. K. Choi, J. K. Actor, S. Rios, M. d'Anjou, T. A. Stadheim, S. Warburton, *et al.*, Recombinant human lactoferrin expressed in glycoengineered Pichia pastoris: effect of terminal N-acetylneuraminic acid on in vitro secondary humoral immune response, *Glycoconjugate J.*, 2008, **25**(6), 581–593.

31 R. Jiang, X. Du and B. Lönnerdal, Comparison of bioactivities of talactoferrin and lactoferrins from human and bovine milk, *J. Pediatr. Gastroenterol. Nutr.*, 2014, **59**(5), 642–652.

32 L. Bernauer, A. Radkohl, L. G. K. Lehmayer and A. Emmerstorfer-Augustin, Komagataella phaffii as emerging model organism in fundamental research, *Front. Microbiol*, 2021, **11**, 607028. Available from: **https://www.frontiersin.org/articles/10.3389/fmicb.2020.607028** [cited 2022 Sep 1].

33 G. Ying, W. S. Hua, J. Wang, Z. X. Dong, C. J. Ming, Z. X. Guang, *et al.*, Producing human lactoferrin by high-density fermentation recombinant Pichia pastoris, *Zhonghua Shiyan He Linchuang Bingduxue Zazhi*, 2004, **18**(2), 181–185.

34 B. Iglesias-Figueroa, N. Valdiviezo-Godina, T. Siqueiros-Cendón, S. Sinagawa-García, S. Arévalo-Gallegos and Q. Rascón-Cruz, High-level expression of recombinant bovine lactoferrin in Pichia pastoris with antimicrobial activity, *Int. J. Mol. Sci.*, 2016, **17**(6), 902.

35 E. H. Wegner, *Biochemical conversions by yeast fermentation at high cell densities*, Bartlesville, OK, 4,414,329, 1983, pp. 1–16.

36 US FDA. GRN No. 967. Soluble egg-white protein produced by Komagataella phaffii strain GSD-1209 [Internet]. 2021 [cited 2023 Nov 29]. Available from: **https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=GRASNotices&id=967&sort=GRN_No&order=DESC&startrow=1&type=basic&search=clara**.

37 US FDA. GRN No. 1001. Myoglobin preparation from a strain of Pichia pastoris expressing the myoglobin gene from Bos taurus [Internet]. 2021 [cited 2023 Dec 20]. Available from: **https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=GRASNotices&id=1001&sort=GRN_No&order=DESC&startrow=1&type=basic&search=pichia**.

38 US FDA. GRN No. 1056. β-lactoglobulin produced by Komagataella phaffii strain "yRMK-66" [Internet]. 2023 [cited 2023 Nov 29]. Available from: **https://www.cfsanappsexternal.fda.gov/scripts/fdcc/index.cfm?set=GRASNotices&id=1056**.

39 M. Boesman-Finkelstein and R. A. Finkelstein, Sequential purification of lactoferrin, lysozyme and secretory immunoglobulin A from human milk, *FEBS Lett.*, 1982, **144**(1), 1–5.

40 Z. Yang, R. Jiang, Q. Chen, J. Wang, Y. Duan, X. Pang, *et al.*, Concentration of Lactoferrin in Human Milk and Its Variation during Lactation in Different Chinese Populations, *Nutrients*, 2018, **10**(9), 1235.

41 W. C. Liu, S. Inwood, T. Gong, A. Sharma, L. Y. Yu and P. Zhu, Fed-batch high-cell-density fermentation strategies for Pichia pastoris growth and production, *Crit. Rev. Biotechnol.*, 2019, **39**(2), 258–271.

42 K. Kobayashi, S. Kuwae, T. Ohya, T. Ohda, M. Ohyama and K. Tomomitsu, High level secretion of recombinant human serum albumin by fed-batch fermentation of the methylotrophic yeast, Pichia pastoris, based on optimal methanol feeding strategy, *J. Biosci. Bioeng.*, 2000, **90**(3), 280–288.

43 A. Kale and R. Patnaik, Recombinant protein recovery methods [Internet]. WO2023192957A1, 2023 [cited 2024 Apr 23]. Available from: **https://patents.google.com/patent/WO2023192957A1/en?q=(egg+white+protein+clara+foods+-pichia)&oq=egg+white+protein+clara+foods+pichia&page=2**.

44 U. Nguyen, N. Squaglia, A. Boge and P. A. Fung, The Simple WesternTM: a gel-free, blot-free, hands-free Western blotting reinvention, *Nat. Methods*, 2011, **8**(11), v–vi.

45 A. Sormunen, E. Koivulehto, K. Alitalo, K. Saksela, N. Laham-Karam and S. Ylä-Herttuala, Comparison of Automated and Traditional Western Blotting Methods, *Methods Protoc.*, 2023, **6**(2), 43.

46 Y. Yang, T. W. Thannhauser, L. Li and S. Zhang, Development of an integrated approach for evaluation of 2-D gel image analysis: Impact of multiple proteins in single spots on comparative proteomics in conventional 2-D gel/MALDI workflow, *Electrophoresis*, 2007, **28**(12), 2080–2094.

47 C. J. Thomas, T. P. Cleland, S. Zhang, C. M. Gundberg and D. Vashishth, Identification and characterization of glycation adducts on osteocalcin, *Anal. Biochem.*, 2017, **525**, 46–53.

48 M. Song, T. A. Sandoval, C. S. Chae, S. Chopra, C. Tan, M. R. Rutkowski, *et al.*, IRE1α–XBP1 controls T cell function in ovarian cancer by regulating mitochondrial activity, *Nature*, 2018, **562**(7727), 423–428.

49 G. Majka, K. Śpiewak, K. Kurpiewska, P. Heczko, G. Stochel, M. Strus and M. Brindell. A high-throughput method for the quantification of iron saturation in lactoferrin preparations. 2013.

50 B. Demeler, UltraScan – A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. In 2007. Available from: **https://api.semanticscholar.org/CorpusID:214705023**.

51 UltraScan – A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments, in

Analytical Ultracentrifugation [Internet], ed. D. Scott, S. E. Harding and A. Rowe. The Royal Society of Chemistry, 2005, pp. 210–230. Available from: **https://books.rsc.org/books/book/351/chapter/145927/UltraScan-A-Comprehensive-Data-Analysis-Software** [cited 2023 Dec 22].

52 D. Nardiello, C. Palermo, A. Natale, M. Quinto and D. Centonze, Strategies in protein sequencing and characterization: Multi-enzyme digestion coupled with alternate CID/ETD tandem mass spectrometry, *Anal. Chim. Acta*, 2015, **854**, 106–117.

53 D. Vincent, V. Ezernieks, S. Rochfort and G. Spangenberg, A multiple protease strategy to optimise the shotgun proteomics of mature medicinal cannabis buds, *Int. J. Mol. Sci.*, 2019, **20**(22), 5630.

54 S. Karav, J. German, C. Rouquié, A. Le Parc and D. Barile, Studying lactoferrin N-glycosylation, *Int. J. Mol. Sci.*, 2017, **18**(4), 870.

55 M. Miura, M. Hirose, T. Miwa, S. Kuwae and H. Ohi, Cloning and characterization in Pichia pastoris of PNO1 gene required for phosphomannosylation of N-linked oligosaccharides, *Gene*, 2004, **324**, 129–137.

56 L. S. Grinna and J. F. Tschopp, Size distribution and general structural features of N –linked oligosaccharides from the methylotrophic yeast, Pichia pastoris, *Yeast*, 1989, **5**(2), 107–115.

57 R. Montesino, R. García, O. Quintero and J. A. Cremata, Variation in N-Linked Oligosaccharide Structures on Heterologous Proteins Secreted by the Methylotrophic YeastPichia pastoris, *Protein Expression Purif.*, 1998, **14**(2), 197–207.

58 R. G. Miele, S. L. Nilsen, T. Brito, R. K. Bretthauer and F. J. Castellino, Glycosylation properties of the Pichia pastoris-expressed recombinant kringle 2 domain of tissue-type plasminogen activator, *Biotechnol. Appl. Biochem.*, 1997, **25**(2), 151–157.

59 J. Zhu, Y. H. Lin, K. A. Dingess, M. Mank, B. Stahl and A. J. R. Heck, Quantitative Longitudinal Inventory of the N -Glycoproteome of Human Milk from a Single Donor Reveals the Highly Variable Repertoire and Dynamic Site-Specific Changes, *J. Proteome Res.*, 2020, **19**(5), 1941–1952.

60 V. A. Ivancic, H. L. Lombardo, E. Ma, M. Wikström and D. Batabyal, Advancing secondary structure characterization of monoclonal antibodies using Microfluidic Modulation Spectroscopy, *Anal. Biochem.*, 2022, **646**, 114629.

61 L. L. Liu, L. Wang, J. Zonderman, J. C. Rouse and H. Y. Kim, Automated, high-throughput infrared spectroscopy for secondary structure analysis of protein biopharmaceuticals, *J. Pharm. Sci.*, 2020, **109**(10), 3223–3230.

62 H. Yang, S. Yang, J. Kong, A. Dong and S. Yu, Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy, *Nat. Protoc.*, 2015, **10**(3), 382–396.

63 A. Dong, Protein secondary structures in water from second-derivative amide I infrared spectra, *Biochemistry*, 1990, **29**(13), 3303.

64 J. Chamani, N. Tafrishi and M. Momen-Heravi, Characterization of the interaction between human lactoferrin and lomefloxacin at physiological condition: Multispectroscopic and modeling description, *J. Lumin.*, 2010, **130**(7), 1160–1168.

65 I. L. Goldman, S. G. Georgieva, G. YaG, A. N. Krasnov, A. V. Deykin, A. N. Popov, *et al.*, Production of human lactoferrin in animal milk1This article is part of a Special Issue entitled Lactoferrin and has undergone the Journal's usual peer review process, *Biochem. Cell Biol.*, 2012, **90**(3), 513–519.

66 B. S. Kendrick, A. Dong, S. D. Allison, M. C. Manning and J. F. Carpenter, Quantitation of the area of overlap between second-derivative amide I infrared spectra to determine the structural similarity of a protein in different states, *J. Pharm. Sci.*, 1996, **85**(2), 155–158.

67 J. M. Steijns and A. C. M. Van Hooijdonk, Occurrence, structure, biochemical properties and technological characteristics of lactoferrin, *Br. J. Nutr.*, 2000, **84**(S1), 11–17.

68 G. Edward, Analytical ultracentrifugation (AUC): an overview of the application of fluorescence and absorbance AUC to the study of biological macromolecules, *Curr. Protoc. Mol. Biol.*, 2020, **133**(1), e131.

69 S. Unzai, Analytical ultracentrifugation in structural biology, *Biophys. Rev.*, 2018, **10**, 229.

70 T. Grant, Small angle X-ray scattering as a complementary tool for high-throughput structural studies, *Biopolymers*, 2011, **95**(8), 517.

71 C. A. Brosey and J. A. Tainer, Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology, *Curr. Opin. Struct. Biol.*, 2019, **58**, 197–213.

72 R. Peterson, R. B. Crawford, L. K. Blevins, N. E. Kaminski, J. S. Sass, B. Ferraro, *et al.*, Dose Range-Finding Toxicity Study in Rats With Recombinant Human Lactoferrin Produced in Komagataella phaffii, *Int. J. Toxicol.*, 2024, 10915818241247013.

73 Y. Anaya, R. R. Martinez, R. E. Goodman, P. Johnson, S. Vajpeyi and X. Lu, *et al.*, Evaluation of the potential food allergy risks of recombinant human lactoferrin expressed in Komagataella phaffii [Internet]. bioRxiv, 2024, pp. 2024.01.18.576250. Available from: **https://www.biorxiv.org/content/10.1101/2024.01.18.576250v1** [cited 2024 Jan 23].

74 R. Morissette, J. Mihalov, S. J. Carlson and K. J. Kaneko, Trends in ingredients added to infant formula: FDA's experiences in the GRAS notification program, *Food Chem. Toxicol.*, 2023, **178**, 113876.