# Introduction of a C-terminal hexa-lysine tag increases thermal stability of the LacDiNac binding adhesin (LabA) exodomain from *Helicobacter pylori*

Vasiliki Paraskevopoulou[a,1], Verónica García Artiaga[a], Rachel Rowlinson[b,3], G. Sebastiaan Winkler[a], Paul Gellert[c], Snow Stolnik[a], Ross Overman[b,2], Franco H. Falcone[a,*]

[a] *School of Pharmacy, University of Nottingham, Nottingham, NG7 2RD, United Kingdom*
[b] *AstraZeneca, Alderley Park, SK10 4TG, United Kingdom*
[c] *AstraZeneca, Macclesfield, SK10 2NA, United Kingdom*

ABSTRACT

*Helicobacter pylori* is a pathogenic microorganism infecting approximately 50% of the global population, and establishes life-long colonization despite the hostile stomach environment. *H. pylori* employs a wide range of outer membrane proteins (adhesins) for epithelial attachment, which specifically bind to glycans or non-carbohydrate structures expressed on the gastric epithelium. A recently described adhesin from *H. pylori* is LabA, named after its ability to bind to a disaccharide present in gastric mucus (LacdiNAc-specific adhesin). Here, we describe the recombinant expression of LabA from *H. pylori* strains J99 and 26695 in *E. coli*. High yields of recombinant LabA were obtained using periplasmic expression. We found that the addition of a C-terminal hexalysine (6K) tag enhanced the thermal stability of LabA without affecting its secondary structure, using differential scanning fluorimetry and circular dichroism spectroscopy. In contrast to our previous report for another *H. pylori* adhesin (BabA), the 6K tag did not enhance recombinant protein yield or solubility. Both versions of LabA, with or without the 6K tag, were expressed and isolated from the periplasmic space of *Escherichia coli*, with a surprisingly high yield of at least 40 mg/L for each independent preparation, following a two-step purification protocol. The proteins were analyzed with mass spectrometry (MS). Unlike its reported effect on stability of BabA, the 6K tag did not appear to protect the N-term of recombinant LabA from partial periplasmic degradation.

## 1. Introduction

*Helicobacter pylori (H. pylori),* a Gram-negative microaerophilic bacterium discovered in 1982 by Marshall and Warren, infects approximately half the world's population and persistently colonizes the human gastric mucosa [1–3]. Although *H. pylori* mainly resides in the mucus layer covering the apical membrane of the epithelium lining the stomach, approximately 20% of the bacterial population colonizing the stomach reaches the apical membrane of the mucous surface cells [3].

Once *H. pylori* reaches the epithelium, specific interactions between bacterial proteins, called adhesins, and the epithelial cells take place. The bacterial adhesins recognize carbohydrate and non-carbohydrate structures present on the surface of the epithelial cells and bind them with high specificity. It is estimated that *H. pylori* has 60 outer-membrane proteins, of which 21 share extended similarity in the amino and carboxyl termini, with a signal peptide sequence in the N-terminus and a β-sheet transmembrane barrel structure in the C-terminus, and constitute the *Helicobacter* outer membrane proteins (Hop) family [4,5]. These have been recently suggested to form a novel class of Type 5-like secretion proteins [6].

The best characterized *H. pylori* adhesin is the blood group antigen-binding adhesin, BabA, which binds to the fucosylated blood group antigen Le[b] in healthy stomach, considered to be the main receptor involved in the initial adhesion of *H. pylori* [7–9]. Despite challenges related to poor solubility and extensive degradation during our initial attempts of protein expression, a recombinant form of the exodomain of

BabA (strain J99) was successfully expressed and purified in our group, after the insertion of a stabilizing hexalysine (6K) tag at the C-terminus of the protein [10]. The addition of the 6K tag not only strongly increased the yield achieved with periplasmic expression in *E. coli*, but also appeared to protect the recombinant protein from partial N-terminal degradation, an issue which can result in the lack of crystal formation for structural studies.

Also recently, a new *H. pylori* adhesin has been shown to bind to GalNAcβ1-4GlcNAc (LacdiNAc), a glycan structure which has been identified on gastric mucosal cells, and was therefore called LacdiNAc-binding adhesin, or LabA [11].

We were, therefore, interested to investigate the applicability of the expression protocol of BabA [10] for the successful recombinant expression of more bacterial adhesins. This would not only facilitate their biophysical and structural characterization, but also allow further understanding of their contribution to the adherence of *H. pylori,* and enable the rational design of (or screening for) attachment inhibitors.

Here, we report the expression and purification of the extracellular region of LabA from *H. pylori* strain J99 in order to assess the effect of the addition of a 6K tag in the C-terminus on the recombinant expression and the structural and biophysical properties of the protein.

## 2. Experimental procedures

### 2.1. Expression vector engineering

As described [10], *E. coli* XL10-Gold cells (Agilent Technologies, USA) were transformed with engineered pOPE101 vectors (Progen Biotechnik, Germany) containing the genes of interest and were used for protein expression.

The 1431 bp-long nucleotide sequence encoding the extracellular domain of the protein (LabA amino acids 21–496; GenBank accession no. AAD05605.1) was amplified from the bacterial genomic DNA isolated from *H. pylori* J99 strain, kindly donated by John Atherton, Nottingham. The first 20 amino acids in the N-terminus of the protein represent a conserved signal sequence, which is utilized by the Sec pathway for translocation of the protein and is cleaved off by a signal peptidase in *H. pylori,* and were excluded from the expression construct. Amino acids 497–690 also constitute a highly conserved sequence, which is predicted to form a β-barrel porin-like structure found in all adhesins of the Hop family [12].

The gene was amplified, purified, double-digested and ligated into a pOPE101 vector as described [10]. The primers used for gene amplification (Table 1) inserted restriction sites for PvuII and BamHI, as well as a sequence encoding a C-terminal hexalysine tag. Q5™ high-fidelity polymerase (New England Biolabs, USA) was used for the mutagenic deletion of the hexalysine tag using the primers in Table 2, according to the manufacturer's protocol.

The recombinant plasmids were used to transform *E. coli* cells and the bacterial colonies were screened with PCR for identification of the successfully engineered plasmids. The sequence of recombinant plasmids was confirmed by DNA sequencing (Source Bioscience).

### 2.2. Protein expression and purification

Protein expression purification and Immunoblotting were carried out as described [10], with slight modifications. For the induction of

**Table 2**
Primers used for the mutagenic deletion of the C-terminal nucleotide sequence encoding the hexalysine tag. [FOR] and [REV] stand for forward and reverse primers, respectively, del for deletion, and 6K for hexalysine tag.

| Name | Sequence (5′-3′) |
| --- | --- |
| LabA [FOR] (6K) del | [Phos]GGATCCGAACAAAAGCTGATCTCAG |
| LabA [REV] (6K) del | [Phos]GAGTTCTTGACTCCTAGATTGCAC |

protein expression, 0.2 mM IPTG was used instead of 0.1 mM IPTG. Also, the purification of the periplasmic extracts using Immobilized Metal Ion Chromatography (IMAC) was carried out using gravity flow, instead of using an ÄKTApurifier system. Following a 1-h incubation of the periplasmic extracts with Ni$^{2+}$ Sepharose 6 Fast Flow resin (GE Life Sciences, USA), the mixture was poured into a 2.5 cm diameter glass Econo-Column® chromatography column holder and the flow through was collected. Next, the column was washed with ten column volumes (CV) of buffer containing 20 mM Tris-HCl, pH 7.4 and 300 mM NaCl, followed by elution steps with 5CVs of 20, 40, 100 and 200 mM imidazole in washing buffer and 15 CV of 500 mM imidazole, sequentially. All elution fractions were analyzed by SDS-PAGE and stained with Coomassie using Instant*Blue*™ (Expedeon). The peak fractions were pooled and concentrated to 5 mL. The pooled sample was then further purified using Size Exclusion Chromatography (SEC), which was carried out as described in Ref. [10], using 25 mM Bicine, pH 8.4 with 150 mM NaCl. Samples were stored at −80 °C. The concentration of purified protein fractions was determined by measuring absorbance at 280 nm using an extinction coefficient of 50930 cm$^{-1}$M$^{-1}$, calculated based on the amino acid sequence of the proteins, using the following formula:

$$\varepsilon = (n1W * 5500) + (n2Y * 1490) + (n3C *125)$$

where n1W, n2Y and n3C are the number of tryptophan, tyrosine and cysteine residues in the protein of interest. Absorbance was measured using a Nanodrop 2000 instrument (ThermoScientific, USA).

### 2.3. Mass spectrometry analysis

The MWs of proteins in purified LabA$_{496}$ and LabA$_{496}$-6K protein samples were determined by Liquid chromatography (LC)-quadrupole time-of-flight (ToF) mass spectrometry (MS). Approximately 5 μg of purified protein sample was loaded onto an Agilent 1100 Series LC (Agilent Technologies, USA) which was coupled to a time-of-flight Q-ToF Premier mass spectrometer (Waters, USA), equipped with an electron spray ionizer for acquisition in a positive ionization mode. The software MassLynx (Waters, USA) was used to analyze the data. In addition, gel slices of fully purified LabA protein bands obtained by SDS-PAGE were analyzed with peptide mapping using trypsin and chymotrypsin for the fragmentation of the protein as described [10]. The bands were reduced with 10 mM dithiothreitol for 45 min at 65 °C and the resulting protein thiols were alkylated by incubation with 50 mM iodoacetamide at room temperature for 20 min in the dark. The proteins were digested with 250 ng of trypsin or chymotrypsin for 16 h at 37 °C. The protein fragments were analyzed with LC-ToF-MS. An UltiMate 3000 Nano LC system (Dionex, USA) coupled to a QSTAR Elite System (AB Sciex,USA) was used and data analyzed with the Mascot software (MatrixScience, USA). The antibody used for detection of the c-Myc Tag was biotinylated mouse anti c-Myc [clone 9E10] antibody

**Table 1**
Primers used for *labA$_{496}$* gene amplification from the bacterial genome. [FOR] and [REV] stand for forward and reverse primers, respectively and underlined are the restriction sites. 6K stands for the hexalysine tag.

| Name | Sequence (5′-3′) |
| --- | --- |
| LabA [FOR] 21- PvuII | CAGTAG<u>CAGCTG</u>GAAGACAACAACGGCTTTTTTGTG |
| LabA [REV] −496 (6K) BamHI | GCTGCT<u>GGATCC</u>CTTCTTCTTCTTCTTCTTGAGTTCTTGACTCCTAGATTGC |

(AbD Serotec, USA).

## 2.4. Circular dichroism spectroscopy

Circular dichroism spectra of the two constructs, LabA$_{496}$ and LabA$_{496}$-6K, were obtained with a J-810 Spectropolarimeter (Jasco, USA). Protein samples (0.3 μg/mL in 100 mM phosphate buffer pH 6.8) were analyzed at 25 °C in a quartz cell with a 0.1 cm path length, at a data pitch of 0.5 nm and scanning speed of 100 nm/min. Baseline correction was carried out using buffer only and the average of three independent scans was plotted. Analysis of the spectroscopic data was carried out with the CD Analysis & PlottIng Tool (CAPITO, Leibniz Institute on Aging – Fritz Lipmann Institute, Germany, [13]).

## 2.5. Differential scanning fluorimetry

A proprietary buffer screen designed by AstraZeneca scientists was used for buffer optimization purposes. The different buffer solutions covered a pH range from 4.8 to 9.1. The ionic strength of the buffer solutions varied from 0 to 300 mM NaCl. The buffer solutions contained no additives. A stock solution of 100 μM protein in 20 mM Tris-Cl (pH 7.4) and 300 mM NaCl mixed with SYPRO® Orange dye (100x concentration) was prepared and diluted with the buffer screen components at a ratio 1:5. Changes in fluorescence were measured across an increasing temperature gradient from 20 °C to 80 °C with a LightCycler 480 II (Roche, Switzerland) at a ramp rate of 0.01 °C/s. The data points from three independent experiments were fitted to a 6-parameter unfolding equation [14] using the Prism analysis package (GraphPad Prism 7 Software, USA).

## 2.6. Three-dimensional structure prediction

For the prediction of structural features of LabA J99, the amino acid sequence of the protein corresponding to the mature sequence of the extracellular region (GenBank accession no. AAD05605.1) was submitted to Phyre$^2$ [15].

## 3. Results

### 3.1. Protein expression

We have previously found that the addition of a 6K tag to the C-terminus of the protein strongly enhanced the total yield of recombinant BabA adhesin expressed in the periplasmic space of *E. coli* [10]. Additionally, the presence of the 6K tag in the construct appeared to protect the protein against partial N-terminal proteolytic cleavage in the periplasmic space. We therefore wanted to investigate whether addition of a 6K tag to LabA also resulted in similar enhancement of yield and protection against proteolytic cleavage in the periplasmic space.
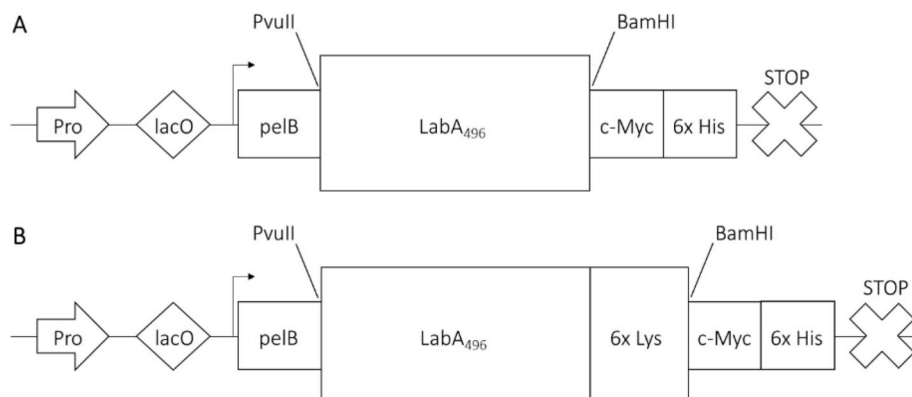
The recombinant expression of the two LabA constructs (LabA$_{496}$ and LabA$_{496}$-6K, both from strain J99), shown in Fig. 1, was carried out in parallel in order to determine the effect of the 6K tag on the expression yield. In the literature, polycationic amino acid tags, similar to the 6K tag, have been reported to improve expression and solubility of other proteins, as recently reviewed by us [16].

The proteins secreted in the periplasmic space of *E. coli* after induction with IPTG and 16 h-long incubation, were extracted following an osmotic shock procedure in the presence of protease inhibitors. The periplasmic proteins were separated by SDS-PAGE and visualized with Coomassie staining, as well as detected with chemiluminescence-based immuno-detection (Fig. 2A–B).

### 3.2. Protein purification

The recombinant LabA proteins were initially purified from the *E. coli* periplasmic extracts by immobilized metal ion affinity chromatography (IMAC) based on the inclusion of a C-terminal His-tag (Fig. 3A).

Protein of high purity and yield was obtained at this stage. Lanes 6-8 (top gel in Fig. 3A) and lanes 4-8 (bottom gel in Fig. 3A) were pooled and processed further using size exclusion chromatography (SEC). These fractions, corresponding to 100-500 mM imidazole, were chosen due to the large amounts of target protein and their apparent high purity, since impurities were eluting at 20 and 40 mM imidazole. As shown in Fig. 3B, LabA$_{496}$ and LabA$_{496}$-6K display very similar behavior on the gel filtration column, by giving almost identical chromatographic peaks. The first and lower peak of elution volume ($V_e$) ~45 mL may represent dimeric or aggregated protein. However, the protein predominantly eluted off the column as a monomer at $V_e$ ~60–70 mL. The purity of the final product was assessed by SDS-PAGE followed by Coomassie staining (Fig. 3C), which revealed high protein purity. In this first experiment, final purified yields of 40.5 and 67.2 mg/L were obtained for LabA$_{496}$ (pooled fractions 5-21) and LabA$_{496}$-6K (pooled fractions 4-16), respectively. In subsequent experiments, as mentioned earlier, LabA$_{496}$ was expressed at a higher yield than LabA$_{496}$-6K.

### 3.3. Mass spectrometry analysis

In order to determine the molecular weight (MW) of the purified recombinant LabA proteins and confirm their identity, LC-qTOF mass spectrometry (MS) was employed. For both purified proteins, obtained by pooling fractions 5-21 (LabA$_{496}$) and fractions 4-16 (LabA$_{496}$-6K) two separate peaks were obtained in the mass spectrum, which were 1,436 Da apart (Fig. 4A and B).

This was the outcome of at least three independent expression attempts. The lower MW peak, which represented the more abundant product, corresponds to a truncated version of the proteins, lacking the first nine amino acids of the expected LabA$_{496}$ amino acid sequence (shown in Fig. 5). The higher MW peak corresponds to the full-length proteins plus four additional amino acids in the N-terminus, found
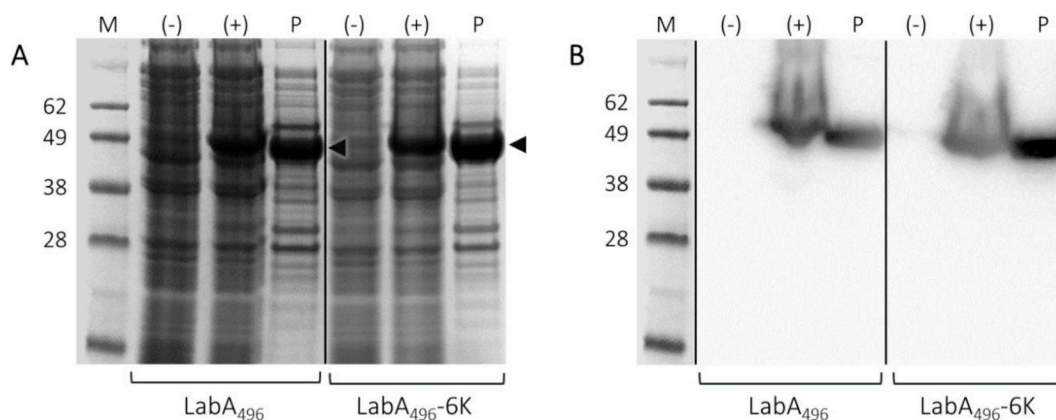


**Fig. 1.** Schematic diagram of (A) pOPE101-LabA$_{496}$ and (B) pOPE101-LabA$_{496}$-6K expression cassettes. Inserts were cloned into the pOPE101 vector using PvuII and BamHI restriction enzyme sites. The abbreviations and symbols represent the following: *Pro* - synthetic lac promoter (P/A1/04/03); *lacO* - lac operator; *arrow* - ATG transcription start codon; *pelB* - cleavable pelB leader sequence; *LabA$_{496}$* – LabA exodomain; *6x Lys* - hexa-lysine peptide tag; *c-Myc* - Myc 9E10 epitope tag; *6x His* - hexa-histidine peptide tag; *X/STOP* - TAA transcription stop codon. **2-Column Fitting Image.**

**Fig. 2.** Secretion of recombinant LabA$_{496}$ and LabA$_{496}$-6K into the periplasmic space. Coomassie staining (left) and c-Myc immunodetection (right) of periplasmic extracts from *E. coli* XL10 Gold cells containing pOPE101-LabA$_{496}$ and pOPE101-LabA$_{496}$-6K expression constructs. The (-) symbol represents uninduced total cell lysates, while (+) represents total cell lysates samples induced with 0.2 mM IPTG at 24 °C for 16 h. Equal amounts of sample were loaded. The LabA proteins are indicated by black arrowheads. The predicted molecular weight of LabA (21-496) is 51.1 kDa, without the tags. With tags, it is 53.342 (LabA$_{496}$) and 54.111 kDa (LabA$_{496}$-6K). **2-Column Fitting Image.** Fig. 2B shows how the addition of the 6K tag to the C-terminus of LabA$_{496}$ initially suggested higher levels of the c-Myc tagged recombinant protein in the periplasmic extracts. However, repeats of the expression did not confirm this finding. Thus, the presence of the 6K tag in the C-terminus of LabA does not appear to affect protein expression yield, and both LabA constructs were expressed at surprisingly high yields.

between the pelB leader sequence and the expressed gene (introduced during the cloning procedure). We assumed that the N-terminal truncation occurs during the cleavage of the pelB leader sequence, while periplasmic secretion takes place.

Mass spectrometry was also used in combination with Coomassie staining of SDS-PAGE gels containing purified recombinant LabA proteins, in order to assess the final purity and integrity of the products (Fig. 4C). The presence of single bands on the gels shows that both LabA$_{496}$ and LabA$_{496}$-6K are stable and, with the exception of a small N-terminal truncation in a larger fraction of the protein, do not undergo proteolytic degradation during expression. Due to the very small MW difference of the two peaks obtained in the mass spectra of LabA$_{496}$ and LabA$_{496}$-6K, the two different sizes of the proteins cannot be resolved on the SDS-PAGE gel.

In order to obtain additional confirmation regarding the identity of the proteins and ensure absence of potential unwanted mutations, the gel band of LabA$_{496}$-6K was fragmented with the use of trypsin and chymotrypsin and the identity of the produced peptides was matched to the expected protein sequence. Overall, 92.6% coverage was achieved and it was confirmed that there was no truncation in the C-terminus of the protein, which contains the appropriate tags for c-Myc immunodetection and IMAC purification (Fig. 5).

### 3.4. Circular dichroism spectroscopy

Circular dichroism spectra were obtained for both LabA$_{496}$ and LabA$_{496}$-6K in order to carry out a preliminary investigation of their secondary fold, but also to verify that the presence of the 6K tag in the C-terminus did not affect the folding of LabA$_{496}$. The spectra of the two proteins confirmed that the two proteins have identical secondary structure, i.e. the same percentage of α-helices and β-strands (Fig. 6). Overall, the curves obtained are consistent with a predominantly α-helical secondary protein structure, and are very similar to the spectra obtained for BabA [17].

### 3.5. Differential scanning fluorimetry

Differential scanning fluorimetry is useful for determination of the melting temperature (T$_m$) of a protein, which constitutes a measure of the protein's thermal stability, with higher T$_m$ values of a protein indicating a higher stability of the protein fold. The method was used in order to discover the effect of the 6K tag on the stability of LabA$_{496}$ and to assess the thermal stability of the proteins at 37 °C, where *H. pylori*

spends most of its life as a mesophilic microorganism. As shown in Fig. 7 (A and B), LabA$_{496}$-6K had consistently higher T$_m$ values than LabA$_{496}$ at the different pH conditions tested, acidic (pH 4.8), close to neutral (pH 7.5) and basic (pH 8.4). Thus, the addition of the 6K tag resulted in enhanced thermal stability of LabA$_{496}$, but without affecting the overall fold. The thermal stability of the protein was surprisingly sensitive to acidic conditions at 37 °C, perhaps in part explaining the known tropism of the bacterium towards the surface of the gastric epithelium [18], where the pH is close to neutral [19] (Fig. 7C).

### 3.6. Three-dimensional structural prediction

The amino acid sequence of LabA J99 submitted to Phyre[2] was automatically aligned to the amino acid sequence of BabA J99 and the structure of LabA predicted based on the known crystal structure of BabA (PDB ID: 4ZH7). Based on the BabA structure, which included the amino acids 27–528, the predicted LabA structure included the amino acids 28–478 (Fig. 8A–C). The comparison of the structures indicates the likelihood of a very similar fold, with three main regions, which we have previously named 'handle', 'head' and 'crown' [17], building on the names introduced for SabA, the first *H. pylori* adhesin crystal structure [20]. The largest difference can be seen in the crown region, which appears incomplete in LabA. The tags in LabA$_{496}$ and LabA$_{496}$-6K are located at the C-terminal end, i.e. at the bottom of the 'handle' region.

## 4. Discussion

The addition of a 6K-tag to the C-terminus of BabA offered the solution to the particularly challenging recombinant expression and purification of the protein. It had improved the solubility of BabA$_{547}$ and increased the levels of expressed proteins in the periplasmic space. Most importantly perhaps, it also appeared to have a protective role against degradation from the proteolytic enzymes of *E. coli* during periplasmic expression [10].

As both yields and the integrity of the expressed protein are of paramount importance in some downstream activities such as X-ray crystallography, it was our interest to investigate further whether the 6K tag had the same effect on the solubility and stability of more adhesins from *H. pylori* and whether it could be used as a generic approach for the periplasmic expression of poorly soluble and/or poorly stable proteins. Two versions of LabA were expressed, with and without the 6K tag. Both proteins were expressed with similarly high yields and
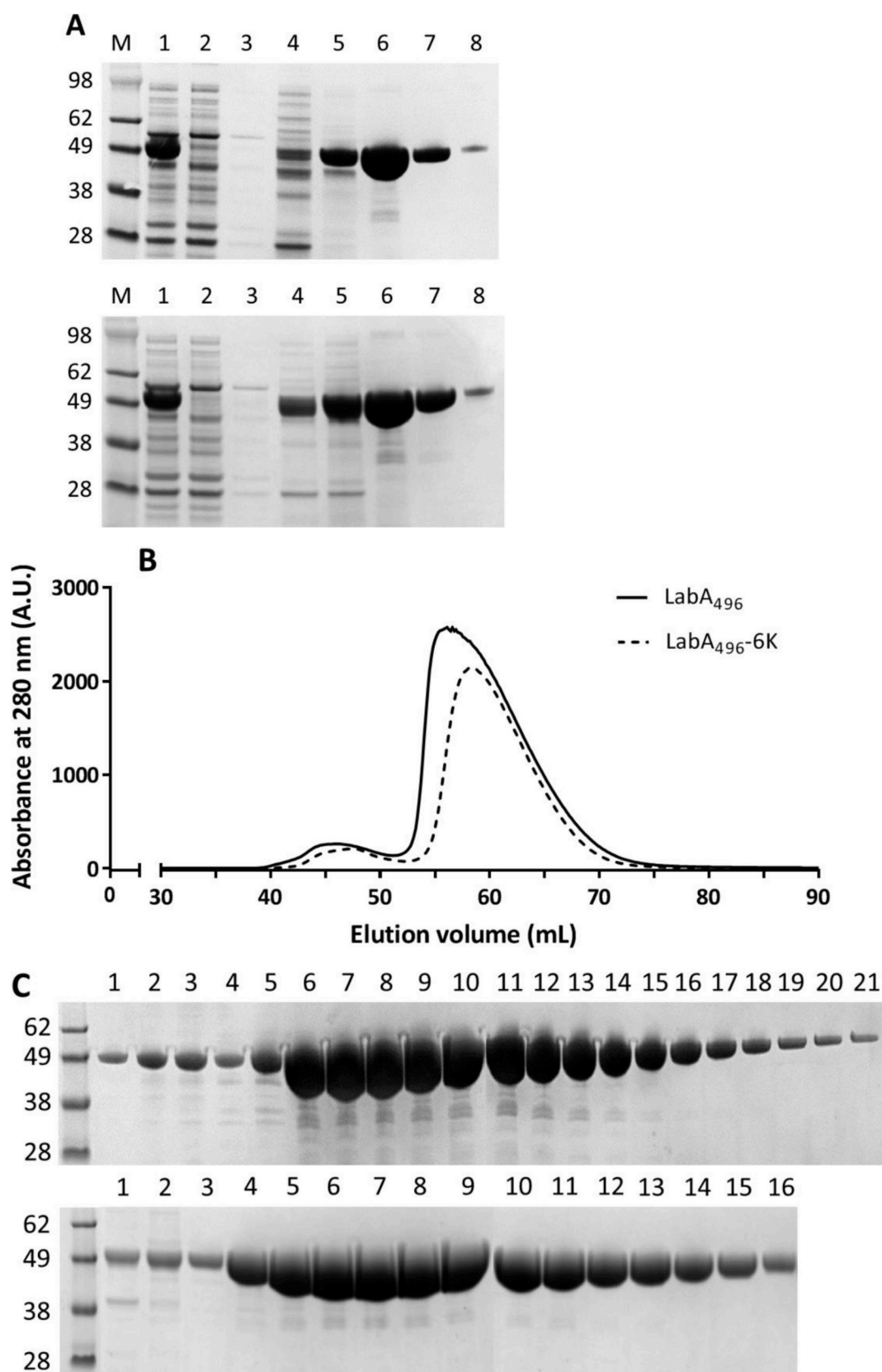
**Fig. 3.** (A) The different protein fractions from the IMAC purification carried out for the purification of LabA$_{496}$ (top) and LabA$_{496}$-6K (bottom), separated with SDS-PAGE and stained with Instant*Blue*™. The symbols stand for: M, molecular weight standards; 1, protein loaded on the nickel NTA column; 2, flow-through; 3, wash; 4–8, protein fractions eluted sequentially with 20, 40, 100, 200 and 500 mM imidazole. (B) Size exclusion chromatography traces of LabA$_{496}$ and LabA$_{496}$-6K proteins loaded onto a HiLoad 16/600 Superdex 75 column and (C) analysis of the protein-containing fractions with SDS-PAGE and Coomassie staining (top: LabA$_{496}$; bottom: LabA$_{496}$-6K). The numbers in the lanes correspond to the fractions of the eluted proteins corresponding to the main peaks in the chromatograms.Lanes 1-4 in the top gel and 1-3 in the bottom gel represent elution volumes of 40-52 mL, while the rest of the lanes represent elution volumes of 52-75 mL. **2-Column Fitting Image.**
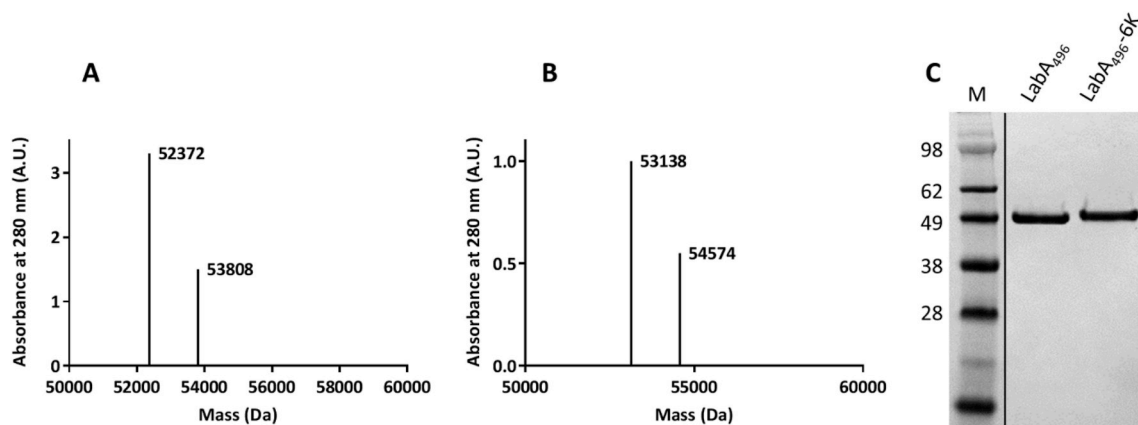
**Fig. 4.** Molecular mass profiles of purified (A) LabA$_{496}$ and (B) LabA$_{496}$-6K. IMAC and SEC purified LabA proteins were used for both SDS-PAGE and liquid chromatography – time-of-flight mass spectroscopy. Two peaks of higher and lower molecular weight than the expected (53,342 and 54,111 Da for LabA$_{496}$ and LabA$_{496}$-6K respectively) are shown. The lower molecular weight peak corresponds to the truncation of the first nine N-terminal amino acids (amino acid sequence: EDNGFFVSA) and the higher molecular weight peak corresponds to the addition of cloning-derived four amino acids in the N-terminus (additional amino acids: QVQL). (C) Coomassie-stained SDS-PAGE gel of purified LabA$_{496}$ and LabA$_{496}$-6K samples. **2-Column Fitting Image.**



**Fig. 5.** LabA$_{496}$-6K amino acid sequence matched with the peptides identified after fragmentation of the protein both with trypsin and chymotrypsin. The identity of the peptides was determined through peptide mass matching only, following liquid chromatography-quadrupole time of flight mass spectrometry. The peptides highlighted in red were identified after digestion with trypsin and the underlined peptides were identified after digestion with chymotrypsin. 92.6% overall coverage of the protein sequence was obtained.
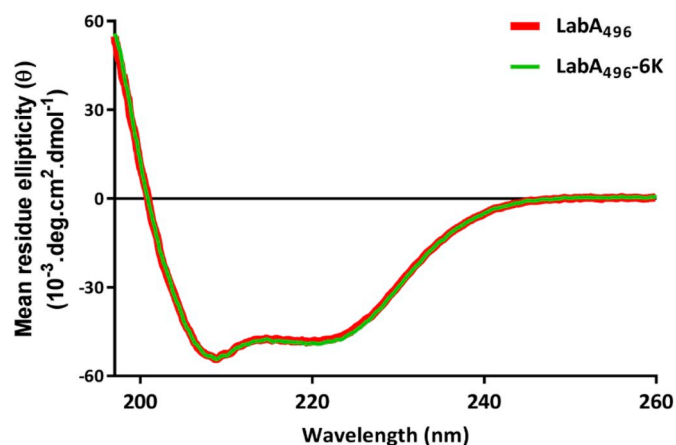


**Fig. 6.** Overlay of far-UV circular dichroism spectra of LabA$_{496}$ (red) and LabA$_{496}$-6K (green) averaged from three independent experiments.

remained soluble during the purification process. These results led us to the conclusion that, in the case of LabA and in contrast to BabA, the 6K tag did not significantly improve the solubility of the protein or enhanced the expression. Similarly, MS data indicated that LabA$_{496}$ is

partially degraded at the N-terminus. Again, the protection against N-terminal degradation in the periplasm conferred by the 6K tag to BabA was not seen for LabA.

The differential scanning fluorimetry experiments suggested that the thermal stability of LabA decreases with increasing acidity, with a T$_m$ of 34.0 °C at pH 4.8, which is below core body temperature, while the T$_m$ increases to 42.5 °C at neutral pH. The thermal stability of LabA, and its dependency on the pH value, may be physiologically significant. Bugaytsova et al. [21] described a sophisticated acid-sensitive regulatory mechanism for BabA, enabling *H. pylori* to detach from the epithelium and mucus being shed into the gastric lumen. Whether a similar mechanism also operates for LabA or other adhesins is currently unknown, but we suggest here that for LabA, a less sophisticated attachment/detachment mechanism may operate via pH-dependency of the thermal stability of the protein. A rapid unfolding of the protein upon a drop in pH may enable the bacterium to detach from the epithelium or mucus which have been shed into the acidic stomach lumen, and allow it to seek refugium using its characteristic adaptation mechanisms, which include increased flagellar movement [22] and the TlpB-dependent negative pH taxis [23] amongst other adaptations, recently reviewed by Ansari and Yamaoka [18]. The caveat of this hypothesis however is that the situation we described for the recombinant exodomain of LabA may not be reflected in the *in vivo* context of the protein, where the adhesin is embedded in the outer membrane lipopolysaccharide layer of *H. pylori*, which may affect its stability and folding.

Independently of its physiological significance, the determination of the T$_m$ values of the two LabA proteins revealed that LabA$_{496}$-6K had higher Tm values than those of LabA$_{496}$ at various pH conditions, so was more stable. This added thermal stability could be advantageous during characterization experiments, when they are carried out under conditions that mimic the natural habitat of the adhesin, i.e. at 37 °C. Such an increase in stability by the addition of an extra polycationic peptide tag in the C-terminus of the recombinant protein would not be desirable though, if this led to a change of properties. The circular dichroism spectra of LabA$_{496}$ and LabA$_{496}$-6K, however, being undistinguishable, revealed that the addition of the 6K tag did not have any measurable effect on the secondary structure of the protein and that the ratio of α-helices to β-sheets remains the same. Consequently, the 6K tag constitutes a beneficial addition to the expression construct, since it increases protein stability without altering the structural and potentially functional properties of this protein.

The functionality of the protein and its potential binding site can be revealed from the three-dimensional structure of the protein. In the
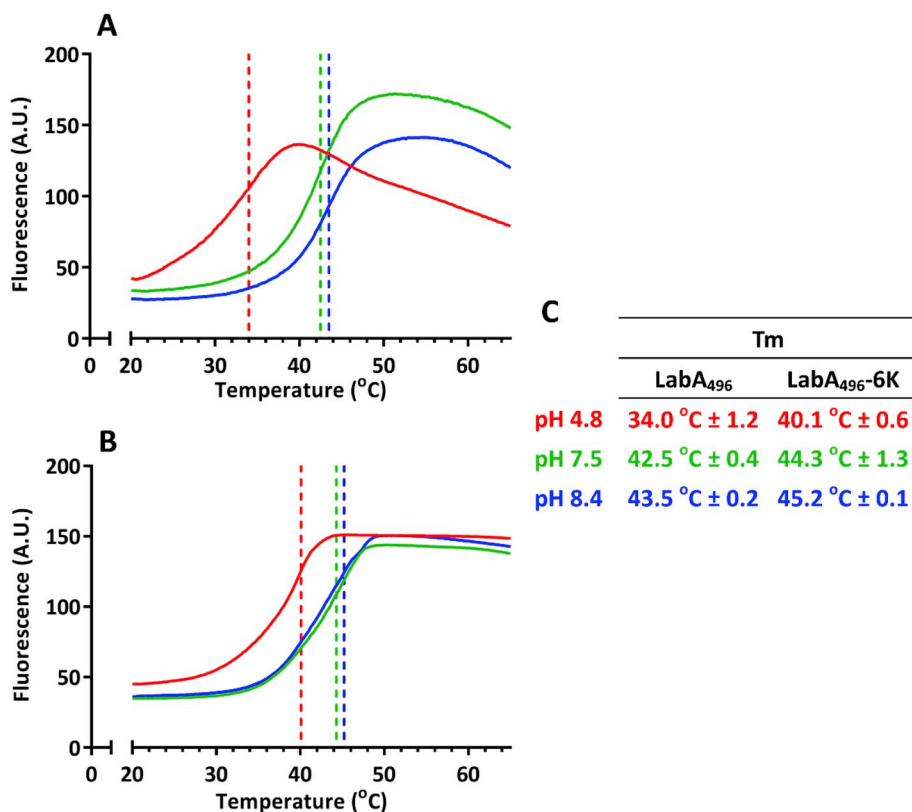
**Fig. 7.** Temperature-induced unfolding transition of (A) LabA$_{496}$ and (B) LabA$_{496}$-6K in three different buffer conditions: 50 mM Sodium Acetate pH 4.8 (red), 50 mM HEPES pH 7.5 (green) and 50 mM Bicine pH 8.4 (blue), in the presence of 100 mM NaCl. The reported midpoint temperature of each protein unfolding transition (T$_m$), intersected by vertical discontinuous lines, is the mean ± SEM from three independent experiments. Fluorescence was measured using 465 nm (exc) and 580 nm (em).

**C**

| | Tm | |
|---|---|---|
| | LabA$_{496}$ | LabA$_{496}$-6K |
| pH 4.8 | 34.0 °C ± 1.2 | 40.1 °C ± 0.6 |
| pH 7.5 | 42.5 °C ± 0.4 | 44.3 °C ± 1.3 |
| pH 8.4 | 43.5 °C ± 0.2 | 45.2 °C ± 0.1 |

current absence of a protein crystal structure and sufficient diffraction data, the three-dimensional structure of LabA was predicted based on the known structure of BabA. Although as seen in Fig. 8, the handle (blue) and head (purple) regions of the two proteins seem to share extended similarity, a pronounced structural difference is revealed in the crown (gold) region. It is known that the crown of BabA consists of four β-sheets, two of which are involved in the binding with the ligand Le$^b$. In contrast, the crown of LabA, which consists of approximately four times less amino acids, is predicted to have two shorter β-sheets, connected with a non-structured loop, and lacks the amino acids responsible for the binding of Le$^b$ in BabA. Based on our current knowledge, the binding site of LabA cannot be predicted and the X-ray crystallography data will be essential for the detailed description of binding. The predicted crown regions of LabA J99 and 26995 also showed clear differences, but in the absence of the knowledge of a ligand-binding site, these are difficult to interpret. This knowledge will need to await elucidation of co-crystal structures of LabA with its putative LacdiNAc ligand.

## 5. Conclusion

This work shows that, in contrast to the effect on BabA, the presence of a C-terminal 6K-tag in LabA does not improve the total yield or protect it against partial N-terminal degradation when the recombinant protein is expressed in the periplasmic space of *E. coli*. However, the addition of the 6K-tag increases the thermal stability of LabA without affecting its overall folding, which could represent an advantage for further studies (e.g. screening for binding inhibitors). Differential scanning fluorimetry shows that thermal stability is decreased at lower pH; this could represent a molecular mechanisms intended to enable *H. pylori* to detach from epithelium or mucus which have been shed into the gastric lumen, where the acidity threatens survival of this bacterium.

## Conflicts of interest

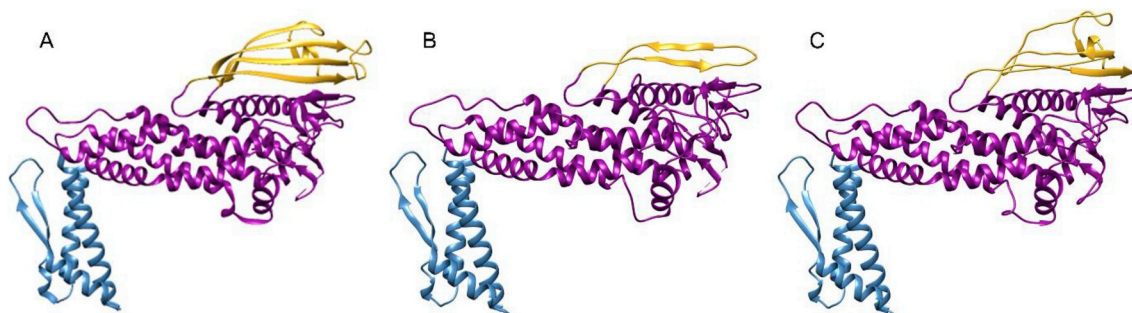The authors declare that they have no conflicts of interest with the contents of this article.



**Fig. 8.** Comparison of the N-terminal exodomains' three-dimensional structures of *H. pylori* BabA and LabA. (A) Crystal structure of BabA J99 (4ZH0) [17], consisting of the handle (blue), head (purple) and crown (gold) regions. (B) Predicted three-dimensional structure of LabA J99, based on the structure of BabA from the same strain. (C) Predicted three-dimensional structure of LabA 26695, based on the structure of BabA J99. The predicted structures were obtained with Phyre$^2$ [15].

## Acknowledgments

## References

[1] J.R. Warren, B. Marshall, Unidentified curved bacilli on gastric epithelium in active chronic gastritis, Lancet (London, England) 1 (1983) 1273–1275.

[2] L.M. Brown, *Helicobacter pylori*: epidemiology and routes of transmission, Epidemiol. Rev. 22 (2000) 283–297.

[3] S. Odenbreit, Adherence properties of *Helicobacter pylori*: impact on pathogenesis and adaptation to the host, Int. J. Med. Microbiol. 295 (2005) 317–324.

[4] N.R. Salama, M.L. Hartung, A. Müller, Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*, Nat. Rev. Microbiol. 11 (2013), https://doi.org/10.1038/nrmicro3016.

[5] M.E. Moore, T. Borén, J.V. Solnick, Life at the margins: modulation of attachment proteins in *Helicobacter pylori*, Gut Microb. 2 (2011) 42–46.

[6] K. Moonens, et al., , Structural insights into polymorphic ABO glycan binding by *Helicobacter pylori*, Cell Host Microbe 19 (2016) 55–66.

[7] T. Boren, et al., , Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens, Science 262 (1993) 1892–1895.

[8] D. Ilver, et al., , *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging, Science 279 (1998) 373–377.

[9] C. Prinz, et al., , Key importance of the *Helicobacter pylori* adherence factor blood group antigen binding adhesin during chronic gastric inflammation, Cancer Res. 61 (2001) 1903–1909.

[10] N. Hage, et al., , Improved expression and purification of the *Helicobacter pylori* adhesin BabA through the incorporation of a hexa-lysine tag, Protein Expr. Purif. 106 (2015) 25–30.

[11] Y. Rossez, et al., , The LacdiNAc-specific adhesin LabA mediates adhesion of *Helicobacter pylori* to human gastric mucosa, J. Infect. Dis. 210 (2014) 1286–1295.

[12] F. Coppens, et al., , Hop-family Helicobacter outer membrane adhesins form a novel class of Type 5-like secretion proteins with an interrupted β-barrel domain, Mol. Microbiol. 110 (2018) 33–46.

[13] C. Wiedemann, P. Bellstedt, M. Gorlach, CAPITO–a web server-based analysis and plotting tool for circular dichroism data, Bioinformatics 29 (2013) 1750–1757.

[14] V. Consalvi, et al., , Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile Thermotoga maritima, Protein Eng. 13 (2000) 501–507.

[15] L.A. Kelley, et al., , The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858.

[16] V. Paraskevopoulou, F. Falcone, Polyionic tags as enhancers of protein solubility in recombinant protein expression, Microorganisms 6 (2018) 47.

[17] N. Hage, et al., , Structural basis of Lewis (b) antigen binding by the *Helicobacter pylori* adhesin BabA, Sci. Adv. 1 (2015) e1500315–e1500315.

[18] S. Ansari, Y. Yamaoka, Survival of *Helicobacter pylori* in gastric acidic territory, Helicobacter 22 (2017) e12386.

[19] A. Allen, G. Flemström, Gastroduodenal mucus bicarbonate barrier: protection against acid and pepsin, Am. J. Physiol. Physiol. 288 (2015) C1–C19.

[20] S.S. Pang, et al., , The three-dimensional structure of the extracellular adhesion domain of the sialic acid-binding adhesin SabA from *Helicobacter pylori*, J. Biol. Chem. 289 (2014) 6332–6340.

[21] J.A. Bugaytsova, et al., , *Helicobacter pylori* adapts to chronic infection and gastric disease via pH-responsive baba-mediated adherence, Cell Host Microbe 21 (2017) 376–389.

[22] D.S. Merrell, et al., , pH-regulated gene expression of the gastric pathogen *Helicobacter pylori*, Infect. Immun. 71 (2003) 3529–3539.

[23] M.A. Croxen, et al., , The *Helicobacter pylori* chemotaxis receptor tlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa, J. Bacteriol. 188 (2006) 2656–2665.

## Abbreviations

*6K:* hexalysine
*MALT:* mucosa-associated lymphoid tissue lymphoma
*IPTG:* isopropyl β-D-1-thiogalactopyranoside
*SDS-PAGE:* sodium dodecyl sulphate- polyacrylamide gel electrophoresis
*ELISA:* enzyme-linked immunosorbent assay
*ITC:* isothermal titration calorimetry
*CD:* circular dichroism
*IMAC:* immobilized-metal ion affinity chromatography
*SEC:* size-exclusion chromatography
*CV:* column volume
*PDB:* protein data bank
*LB:* lysogeny broth
*EDTA:* ethylenediaminetetraacetic acid
*DTT:* Dithiothreitol
*PBS:* phosphate buffered saline
*HSA:* human serum albumin
*PCR:* polymerase chain reaction
*ESI-MS:* electrospray ionization mass spectrometry