Refinement of the Fusion Tag PagP For Efficient Inclusion Body-Targeted Expression of Recombinant Antimicrobial Peptides in *Escherichia Coli*

Xuefeng Li
South China Agricultural University
Baorong Zhang
South China Agricultural University
Quan Hu
South China Agricultural University
Changchao Chen
South China Agricultural University
Lu Liu
South China Agricultural University
Wenshi Luo
South China Agricultural University
Shengbin Wang (swwl@scau.edu.cn)
South China Agricultural University https://orcid.org/0000-0001-7402-2753

Research

Keywords: fusion tag, inclusion body, PagP, target protein recovery, antimicrobial peptides

Posted Date: October 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-961647/v1

Abstract

The methods developed for efficient insoluble protein production are less well explored. Our data demonstrated that PagP, an *E. coli* outer membrane protein with high β-sheet content, could function as an efficient fusion partner for inclusion body-targeted expression of antimicrobial peptide Magainin II, Metchnikowin and Andropin. The primary structure of a given polypeptide determines to a large extent its propensity to aggregate. The aggregation "hot spots" (HSs) in PagP was subsequently analyzed with the web-based software AGGRESCAN, leading to identification of the C-terminal region with high dense distribution of HSs. The absolute yields of recombinant antimicrobial peptide Metchnikowin and Andropin could be increased significantly when expressed in fusion with this version of PagP. Moreover, a Proline-rich region was found in the β-strands of PagP. Substitution for these prolines by residues with high β-sheet propensity and hydrophobicity significantly improved its ability to form aggregates, and greatly increased the yield of the recombinant passenger peptides. Fewer examples have been presented to separate the recombinant target proteins expressed in fusion inclusion bodies. Here, we reported an artificial linker peptide NHT with three motifs, by which separation and purification of the authentic recombinant antimicrobial peptides could be implemented.

Highlights

- PagP functions effectively for inclusion body-targeted expression of antimicrobial peptides.
- The C-terminal region of PagP harbours high density of aggregation "hot spots", and can also be used as an effective fusion tag, allowing the production of antimicrobial peptide Metchnikowin and Andropin to be increased significantly.
- Substitution for proline in PagP by amino acid residues with high β-sheet propensity and hydrophobicity helps it to form aggregates rapidly and stably, leading to the increased production of the potent antimicrobial peptide Magainin II.
- Artificial peptide NHT allows the recombinant antimicrobial peptides to be conveniently purified.

Introduction

Diverse conditions can lead to an alteration of protein homeostasis, resulting in protein aggregation in all living cells. Nevertheless, the commonest event is provided by the fast and overexpression of recombinant proteins in bacteria, leading to formation of the inclusion bodies (IBs). In the last decade, IBs have been investigated from a different perspective and their study gained considerable interest. Targeting recombinant proteins into IBs provides a new interesting approach for their production and IBs are also proposed as an almost pure source of recombinant proteins. Unordered sequences are highly exposed and therefore suitable for protein–protein interactions and posttranslational modifications [1]. However, the high degree of exposure that confers these sequences biological activities also renders them vulnerable to proteolysis in the setting of heterologous expression. Targeting their expression into IBs will provide an alternative strategy for protecting them from proteolytic degradation [2]. Although IBs

formation is so common in the heterologous protein expression system, especially under overexpression condition, a robust fusion partner is often necessary. Newly expressed proteins may exist in three states: properly folded, partially folded intermediates, and aggregated *in vivo*, which are in dynamic equilibrium and subjected to the action of chaperones and proteases [3]. Recombinant proteins that exist primarily as partially folded or misfolded intermediates can be rapidly degraded *in vivo*. Thus, their accumulation can be improved significantly by shifting the equilibrium towards the formation of IBs in view of the protection from proteolytic degradation. For a given protein, it is often difficult to assess which insoluble fusion partner is the most effective at targeting it into IBs. Nevertheless, some insoluble fusion partners have been demonstrated to work well with multiple target peptides or proteins, such as ketosteroid isomerase (KSI), EDDIE, PagP and so on [2].

When overexpressed in *E. coli*, fungal prion HET-s accumulated predominantly in IBs with high β-sheet content, and displayed very similar characteristics of amyloid. The kinetics of HET-s amyloid fibril formation demonstrated that the amyloid growth was nucleation-dependent [4]. An increasing amount of evidence has showed that IBs are highly ordered aggregates and bear a characteristic cross β structure similar to amyloid fibers. As occurs for amyloids, the formation of IBs is promoted by intermolecular interactions in a nucleation-dependent manner through hydrophobic protein patches [5]. Mutation of Pro-102 or Pro-105 in human prion (PrP) to leucine could greatly promote the formation of PrP aggregates, since proline residue normally disfavored the formation of β-sheet conformation [6]. IBs have been recognized as a valuable model to understand protein aggregation in eukaryotes, and for the search of specific inhibitors or disaggregation approaches [7].

As one of the important effectors of the innate immune system in multi-cellular organisms, antimicrobial peptides (AMPs) protect their hosts against a large variety of invading pathogens. These small cationic peptides have been recognized as ancient evolutionary molecules that have been effectively preserved in mammals [8]. More than 2000 AMPs have been reported in the Antimicrobial Peptide Database (http://aps.unmc.edu/AP/main.php). AMPs are generally composed of less than 50 amino acid residues with overall positive charges, imparted by the presence of multiple lysine and arginine residues. These peptides are usually highly amphipathic molecules with both hydrophobic and hydrophilic patches on their surface. The amphipathic character facilitates their interaction with membranes. It is generally accepted that AMPs interact by electrostatic forces with the negatively charged phospholipid head groups on the bacterial membrane and cause its disruption. Several mechanisms, such as “barrel-stave model”, “carpet model” have been proposed [9]. Besides their antimicrobial activities, some AMPs are also recognized for their immunomodulatory properties. A human cationic antimicrobial peptide cathelicidin LL-37 could modulate the macrophage functions, and be taken as a potent antisepsis agent with its ability to inhibit the macrophage stimulation by bacterial components [10]. β-defensin induced the expression of diverse chemokines and cytokines in keratinocytes [11, 12]. At high concentrations, LL-37 could also act directly as chemokines to recruit a variety of immune and epithelial cells, thus enhancing the clearance of infections [13]. The direct effect of AMPs on adaptive immunity has been proposed as they could act directly on B and T cells [14]. Adjuvant combination of the bovine antimicrobial peptide

indolicidin with CpG and polyphosphazene can stimulate a long-term and well-balanced immune response in cattle [15].

Recombinant expression of AMPs in *E. coli* faces two challenges. AMPs, rich in basic amino acid residues, are highly susceptible to the proteolytic degradation for their unordered structure, and toxic to the producing hosts. To overcome both obstacles, a commonly used strategy is to fuse them to fusion tags. As the solubility-enhancing fusion tags, thioredoxin and glutathione transferase (GST) has been used as carrier proteins for the fusion expression of AMPs [16]. However, antibacterial peptide CM4 (ABP-CM4), a small cationic peptide with broad-spectrum activities, could hardly be expressed in fusion with GST in light of the proteolytic degradation [17]. Targeting AMPs into IBs is believed to be more effective than their soluble fusion expression in masking their toxic effects and protecting them from proteolytic degradation. Some fusion tags have been demonstrated to be effective in targeting AMPs or other intrinsically disordered peptides into IBs, bringing about the satisfying expression level [2, 18]. Even so, for a given AMP, selecting a suitable fusion tag to confer it efficient expression in IBs remains empirical or needs to be screened. Meanwhile, the repertoire of IB-targeted fusion tags is also needed to be expanded.

Development of effective methods for recombinant production and purification of AMPs is of practical significance [19]. In this report, the aggregation “hot spots” (HSs) of the fusion tag PagP was analyzed. Based on the structural analysis, we demonstrated that the C-terminal region of PagP with high dense distribution of HSs could also function as an effective fusion tag to target AMPs into IBs. Meanwhile, our data showed that mutation of residue proline in or near the aggregation HSs to hydrophobic residues could significantly improve its potency as IBs-targeted fusion tag. Much fewer examples have been presented to separate recombinant target proteins expressed in fused IBs. Here, we provided an alternative approach to recover the authentic AMPs expressed as IBs in fusion with PagP.

Materials And Methods

Bacterial strains and plasmid constructions

E. coli DH5α was used as a host strain for the gene cloning and preparation of plasmids. *E. coli* BL21 (DE3) was used as a host strain for expression of the recombinant proteins. All *E. coli* strains are cultured in LB medium supplemented with antibiotics as needed.

Antimicrobial peptide magainin II (Mag II) [33], Metchnikowin (Metch) [35] and Andropin [34] were selected as interest peptides in this study. Mag II, Metch and Andropin gene were synthesized by overlap extension PCR according to their amino acid sequences, with a codon pattern adapted to the usage bias of *E. coli* [21, 43]. Fusion *Mag II* gene was first synthesized by overlap extension PCR with primer 1, primer 2, primer 3 and primer 4, then sequentially amplified with primer 8 and primer 4, primer 9 and primer 4, primer 10 and primer 4. The resulting amplicon encoded a fusion antimicrobial peptide Mag II with a sequence (ASRHWMAG) for the Ni(II)-dependent peptide bond hydrolysis [44], a His_6-tag (HHHHHH) and a recognition site (ENLYFQ) by the site-specific protease TEV [45] at its N-terminus. The resulting sequence

ASRHWMAGHHHHHHENLYFQ was denoted as NHT. Fusion *Metch* gene was also first synthesized by overlap extension PCR with primer 5, primer 6 and primer 7, and then sequentially amplified with primer 8 and primer 7, primer 9 and primer 7, primer 10 and primer 7. The resulting amplicon encoded a fusion peptide NHT-Metch. Similarly, fusion *Andropin* gene was first synthesized by overlap extension PCR with primer 11, primer 12, primer 13, primer 14 and primer 15, and then sequentially amplified with primer 8 and primer 15, primer 9 and primer 15, primer 10 and primer 15. The fusion *Andropin* gene was lastly constructed, encoding fusion peptide NHT-Andropin. Fusion antimicrobial peptides constructed were outlined in Fig. S1. The PagP gene (Gene ID: 946360) encoding its mature form was amplified using primer 16 and primer 17 from the chromosome of *E. coli* strain MG1655. The primers designed for gene amplification are listed in Supplementary data Table S1.

PagP is a Gram-negative bacterial outer membrane protein and extremely prone to accumulate in IBs when overexpressed in *E. coli* [2]. Base on the analysis of PagP sequence by web platform AGGRESCAN at http://bioinf.uab.es/aggrescan/ [27], seven aggregation “hot spots” (HSs) were identified, denoted as HS1, HS2, HS3, HS4, HS5, HS6 and HS7 successively. The C-terminal region of PagP (residue 101-161) is denoted as PagP-1, comprising four aggregation HSs (HS4, HS5, HS6 and HS7) (Fig. 3 and Fig. 4).

After being digested by *Eco*RI and *Hin*dIII, fusion Metch gene was inserted into expression vector pET30a and pET28b, generating vector pET30a-NHT-Metch and pET28b-NHT-Metch. The amplicon encoding mature PagP was digested with *Nde*I and *Bam*HI, and then inserted into pET30a-NHT-Metch, generating vector pET30a-PagP-NHT-Metch for intracellular expression of the fusion protein PagP-NHT-Metch in *E. coli*. Similarly, after being digested by *Eco*RI and *Hin*dIII, fusion Mag II gene and Andropin gene were inserted into expression vector pET30a respectively, generating vector pET30a-NHT-Mag II and pET30a-NHT-Andropin. The amplicon encoding mature PagP was digested with *Nde*I and *Bam*HI, and then inserted into pET30a-NHT-Mag II and pET30a-NHT-Andropin respectively, generating vector pET30a-PagP-NHT-Mag II and pET30a-PagP-NHT-Andropin for intracellular expression of the fusion protein PagP-NHT-Mag II and in PagP-NHT-Andropin in *E. coli*.

Using vector pET30a-PagP-NHT-Metch as template, fragment encoding fusion protein PagP-1-NHT-Metch was amplified with primer 18 and primer 7, and inserted into vector pET28b, generating expression vector pET28b-PagP-1-NHT-Metch. Similarly, corresponding fragments were amplified by using primer 19 and primer 7; primer 20, primer 21 and primer 7; primer 22 and primer 7; primer 23 and primer 7; primer 24 and primer 7, respectively. These fragments were inserted into vector pET28b respectively, generating vector pET28b-PagP-2-NHT-Metch, pET28b-PagP-3-NHT-Metch, pET28b-PagP-4-NHT-Metch, pET28b-PagP-5-NHT-Metch and pET28b-PagP-6-NHT-Metch. Furthermore, the fragment encoding NHT-Mag II was substituted for NHT-Metch gene, generating corresponding expression vector pET28b-PagP-1-NHT-Mag II, pET28b-PagP-2-NHT-Mag II and pET28b-PagP-3-NHT-Mag II. Similarly, the fragment encoding NHT-Andropin was substituted for NHT-Metch gene, generating corresponding expression vector pET28b-PagP-1-NHT-Andropin, pET28b-PagP-2-NHT-Andropin and pET28b-PagP-3-NHT-Andropin.

All constructs were checked by sequencing and primer sequences were listed in Supplementary data Table S1.

Site-specific mutation of PagP

Site-specific mutagenesis was performed to generate the mutants PagP on pET30a-PagP-NHT-Mag II background according to our previous report [46]. The resulting vectors are denoted as pET30a-P121L/P123L-PagP-NHT-Mag II, pET30a-P127I-PagP-NHT-Mag II, pET30a-P135L-PagP-NHT-Mag II, pET30a-P127I/P135L-PagP-NHT-Mag II, pET30a-P121L/P123L/P135L-PagP-NHT-Mag II, pET30a-P121L/P123L/P127I/P135L-PagP-NHT-Mag II, respectively. In the case of P121L/P123L-PagP, the residue Pro at position 121 and 123 was mutated to Leu with the primer 25 and primer 26. Similarly, mutant P127I-PagP was constructed with the primer 27 and primer 28, and P135L-PagP with primer 29 and primer 30. These mutants were also used as templates for further mutation. By performing the second round of mutation and the third round of mutation, the mutant P127I/P135L-PagP, P121L/P123L/P135L-PagP and P121L/P123L/P127I/P135L-PagP were constructed. All mutations were checked by sequencing.

Expression of fusion proteins

The *E. coli* cells (BL21 (DE3)) harboring the expression vectors were cultured overnight at 37°C in 6 mL of LB medium. The cultures were then diluted 100-fold and allowed to grow at 37°C until they reached the mid-log phase (with an optical density of around 0.6–0.8 at 600 nm). The expression of fusion proteins was induced by adding isopropyl β-D-1-thiogalactopy-ranoside (IPTG) to a final concentration of 0.3 mM. The resulting culture was further incubated for 12 h at 37°C. 100 mL of culture was centrifuged for 10 min at 6000 rpm to harvest bacteria at 4 °C. The *E. coli* cells were re-suspended in 10 mL lysis buffer (50 mM NaH_2PO_4-Na_2HPO_4, 0.2 M NaCl, 20 mM imidazole, pH 8.0), and then lysed by sonication on ice. After that, inclusion bodies were isolated by centrifugation for 15 min at 12000 rpm at 4°C, washed twice with washing buffer I (20 mM Tris-HCl, 50 mM NaCl, 0.1 % Triton X-100, 5 mM EDTA, pH 8.0), and then with washing buffer II (20 mM Tris-HCl, 50 mM NaCl, pH 8.0). IBs extracted or whole cell lysates were subjected to analysis by SDS-PAGE. The images of the gels were further analyzed using software (Image Lab^{TM}, BIO-RAD) to evaluate the expression levels of fusion proteins.

The specific hydrolysis of fusion proteins by Ni (II) ion

The specific hydrolysis of fusion proteins by Ni (II) ion was conducted under denatured condition according to the previous report [23]. The inclusion bodies isolated were dissolved in hydrolysis reaction buffer (20 mM Hepes, 6 M GuHCl, pH 8.2) with a concentration of about 200 μM. After that, fusion proteins were subjected to hydrolysis by addition of $NiSO_4$ to final concentration of 5 mM, and incubated at 60°C for 12 h, 24 h or 36 h to investigate the hydrolysis process. The mixture was lastly diluted 5-fold with addition of lysis buffer (50 mM NaH_2PO_4-Na_2HPO_4, 0.3 M NaCl, 20 mM imidazole, pH 8.0), and centrifuged for 15 min at 12000 rpm at 4°C to remove the precipitated fusion tag.

Purification of recombinant antimicrobial peptides

After being specifically hydrolyzed by Ni^{2+}, the fusion protein split into a tag and the NHT-AMP, which was further purified by Ni-chelating affinity chromatograph according to the protocol specified by the manufacturer (GE Healthcare Bio-Sciences). Firstly, the column was equilibrated with lysis buffer (50 mM NaH_2PO_4-Na_2HPO_4, 0.3 M NaCl, 20 mM imidazole, pH 8.0), and the protein sample was loaded. Then, the loaded column was washed three times with washing buffer III (50 mM NaH_2PO_4-Na_2HPO_4, 0.3 M NaCl, 40 mM imidazole, pH 8.0). After that, the recombinant NHT fusion (NHT-Metch, NHT-Mag or NHT-Andropin) was eluted with the elution buffer I (50 mM Na_2HPO_4-Na_2HPO_4, 200 mM imidazole, pH 7.0).

The fusion antimicrobial peptide (NHT-Metch, NHT-Mag or NHT-Andropin) contained in the elution buffer I was directly subjected to specific cleavage by addition of purified protease TEV at 25°C for 12 h according to the previous report [45]. The released AMP (Metch, Mag II or Andropin) was further purified using ion exchange chromatography (Macro-Prep® CM Resin, Bio-Rad). The column was first equilibrated with 2 column volumes (CV) of equilibration buffer (200mM imidazole, 50mM NaH_2PO_4-Na_2HPO_4, pH 7.0). 5 mL of the protein sample was loaded onto the column. Lastly, the recombinant AMP was eluted with elution buffer II (500 mM NaCl, 50mM Na_2HPO_4-Na_2HPO_4, pH 8.0).

Microbicidal activity assay of the recombinant antimicrobial peptides

The antimicrobial activities of the recombinant AMPs were analyzed by inhibition zone assay according to the protocol presented by previous report [47]. Briefly, gram-positive *Staphylococcus aureus* ATCC 25923 and gram-negative *E. coli* strain $K_{12}D_{31}$was grown overnight at 37°C in LB medium. 50 μL of these cultures were inoculated into 50 mL of fresh LB medium and incubated for additional 2-3 h at 37°C to OD_{600}=0.5, respectively. 200 μL of cell suspension was inoculated into 50 mL of pre-warmed (45°C) LB medium containing 0.8 % (w/v) agar and rapidly dispersed. The medium was then poured into the Petri dish (9.0 cm in diameter) to form a uniform layer with a depth of about 1.5 mm. Holes with a diameter of 2 mm were punched into the gelated medium. For microbicidal activity assay, recombinant AMPs were added into the holes punched, and the plate was incubated at 37°C for 12 h to investigate appearance of the inhibition zones.

Results

Targeting recombinant AMPs into IBs

Unfavorable patterns of codon usage affected the high-level expression of recombinant proteins [20]. All antimicrobial peptide Metch, Andropin and Mag II genes were designed with a codon pattern adapted to the codon usage bias of *E. coli* by the web platform at http://genomes.urv.es/OPTIMIZER [21]. Meanwhile, an artificial short peptide (NHT) with three functional motifs (motif I, motif II, motif III) was designed and appended at the N-terminus of AMPs (Fig. 1a). The motif I (ASRHWMAG) is to allow the fusion proteins to be hydrolyzed site-specifically by Ni^{2+} ion; the motif II (HHHHHH) to allow the fusion

AMPs (NHT-Metch, NHT-Andropin or NHT-Mag II) to be purified by Ni-chelating chromatography; the motif III (ENLYFQ) to allow the fusion AMPs to be cleaved site-specifically by protease TEV to release authentic antimicrobial peptide Metch, Andropin or Mag II.

Protein PagP is ready to accumulate in IBs when overexpressed in *E. coli* [2]. We fused NHT-Metch to its C-terminus, and investigated overexpression of the resulting fusion protein PagP-NHT-Metch in *E. coli*. As expected, fusion protein PagP-NHT-Metch accumulated predominantly in IBs (Fig. 1).

Site-specific hydrolysis of the fusion protein PagP-NHT-Metch catalyzed by Ni (II) ion

Insoluble fusions that existed in IBs were refractory to the site-specific cleavage by proteases, such as thrombin or TEV. An optimized amino acid sequence has been developed for Ni (II)-catalyzed cleavage [22], and it has been demonstrated that nearly complete nickel-catalyzed hydrolysis of the fusion proteins could be achieved under denaturing conditions [23]. As expected, full cleavage of the fusion protein PagP-NHT-Metch was achieved at 60°C after 24 h duration of hydrolysis, being confirmed by appearance of the fusion tag PagP (Fig. 2a). The resulting His_6-tagged passenger peptide NHT-Metch could be subsequently purified by Ni-affinity chromatography (Fig. 2b).

Refinement of the fusion tag PagP for more efficient production of recombinant peptides

As an integral outer membrane protein, *E. coli* PagP exhibits the β-barrel architecture with a hydrophobic exterior facing the membrane bilayer and a hydrophilic interior cavity [24]. The increasing accumulated data have suggested that the composition and sequence of a given protein define its propensity to aggregate to a large extent [25]. It has been demonstrated that some very short specific amino acid stretches can act as nucleators for formation of the amyloid fibril, usually known as aggregation "hot spots" (HSs). Many of these HSs have been characterized in the proteins governing the neurodegenerative and systemic amyloidogenic diseases [26]. Based on the aggregation-propensity scales for natural amino acids derived from *in vivo* experiments, an web-based software AGGRESCAN for prediction of the aggregation-prone segments in protein sequences was presented [27]. To identify the putative aggregation HSs of the fusion tag PagP, we analyzed its sequence with AGGRESCAN web tools. A total of seven HSs, termed as HS1, HS2, HS3, HS4, HS5, HS6 and HS7, were identified (Fig. 3). We also found that these HSs were distributed unevenly along its sequence, in which four HSs were present at its C-terminal region comprising residues 101-161, leading to the higher dense distribution of HSs at its C-terminal region. This finding implied that the C-terminal region of PagP have great potential to aggregate. We reconstruct a serial of fusion tags using the C-terminal region of PagP as the template. NHT-Metch was then fused with these constructed tags, respectively (Fig. 4). When targeted to express in IBs, PagP-1-NHT-Metch, PagP-2-NHT-Metch and PagP-3-NHT-Metch accumulated in considerable amounts (Fig. 5a and Fig. 5c). While, there were no obvious accumulation of fusion proteins for tag PagP-4, PagP-5 and PagP-6 (Fig. 5b), which possessed higher dense distribution of aggregation HSs owing to addition of HS3 or a combination of HS3 with HS2 and HS1. Our data showed that a simple addition of aggregation HSs to a given fusion tag might bring about the negative effect on its function to target passenger peptides

into IBs. Compared with the PagP-NHT-Metch, PagP-3-NHT-Metch accumulated in larger number of molecules, increasing absolute yield of the recombinant peptide up to about 40%.

We further fused NHT-Andropin and NHT-Mag II with these tags, respectively. All these fusion tags targeted NHT-Andropin into IBs in *E. coli* as well as they did for NHT-Metch, increasing the absolute yield of the recombinant Andropin up to nearly 50% (Fig. 6a and Fig. 6b). However, these fusion tags lost their efficacy when fused with passenger NHT-Mag II, making the accumulation of fusion proteins to be hardly detected by SDS-PAGE (Fig. 6c). Our data suggested that the effectiveness of a given fusion tag for IBs-targeted expression was affected significantly by the physicochemical properties of passenger proteins. Mag II is a potent antimicrobial peptide, and exerts its biological activity by destroying the integrity of cytoplasmic membrane [28]. The resulting fusion proteins might not form stable aggregates immediately, be still toxic to the host cells, and induce the multiple levels of cellular responses. Subsequently, these toxic intermediates were degraded by protease.

Recombinant IBs in *E. coli* possess the characteristic cross-β structure of amyloid fibers [29]. The single-point mutation at the central position of beta-amyloid peptide (beta 42) to change the hydrophobicity or to disfavor the formation of secondary structure (β-sheet) affected significantly its aggregation propensity and neurotoxicity [30, 31]. We analyzed the structure of PagP and highlighted its C-terminal region comprising the four β-sheet elements [24] (Fig. 3). This region is rich in residue proline (P-121, P-123, P-127, P-135), which are relatively less present in β-sheet. The mutants P127I-PagP-NHT-Mag II, P135L-PagP-NHT-Mag II and P121L/P123L-PagP-NHT-Mag II were constructed by substituting the hydrophobic residues (Ile or Leu) for proline. When induced to express in *E. coli*, these mutants accumulated in larger amounts in comparison with the counterpart PagP-NHT-Mag II. The yields of these recombinant mutants were increased significantly up to 44.3-60.5%, (Fig. 7). Compared with proline, residue isoleucine or leucine has higher propensity to form β-sheet [32]. Moreover, isoleucine and leucine have bulkier and hydrophobic side chains which favored the intermolecular or intramolecular interactions between side chains, facilitating the formation of aggregates. Our results demonstrated that the aggregation ability of a fusion tag could be further improved greatly by increasing its overall hydrophobicity or enhancing its propensity to form more stable β-sheet conformation.

Purification of the recombinant antimicrobial peptides and test of their biological activity

The fusion tag normally must be cleaved and removed due to its potential interference with the activities of passenger proteins. This is especially challenging in the case of IBs-targeted expression, since the IBs are usually solubilized in harsh denaturants or detergents, precluding the utilization of enzymatic cleavage (for example, with TEV protease or thrombin). With the help of artificial peptide NHT, the recombinant antimicrobial peptides in fusion with NHT could be conveniently purified by Ni-affinity chromatograph after Ni^{2+}-catalyzed specific cleavage. Lastly, authentic antimicrobial peptides were recovered with successive site-specific cleavage by protease TEV and ion exchange chromatography (Fig. 8). Our data showed that the artificial peptide NHT designed here provided an effective means for purification of the authentic peptides expressed in fusion IBs. Mag II and Andropin possessed the

biological activity against both Gram-positive and Gram-negative bacteria [33, 34], while peptide Metch only showed inhibitory activity towards Gram-positive bacteria [35]. We then tested their antimicrobial activity using classical inhibition zone assays. As expected, the appearance of clear inhibition zones showed that the recombinant peptides possessed their native antimicrobial activities against both Gram-positive *staphylococcus aureus* and Gram-negative *E. coli*, or only against Gram-negative bacteria *staphylococcus aureus* (Fig. 9).

Discussion

AMPs are generally rich in basic amino acid (lysine and arginine) residues, and consequently possess net positive charges at physiological pH. On the other hand, many linear AMPs have undefined structure in aqueous solution [36]. These properties make AMPs high susceptible to proteolytic degradation during heterogenous expression. Thus, recombinant AMPs are commonly fused to fusion tags to protect them from cellular proteases and mask their cellular toxicity to the expressing hosts [37]. Compared with solubility-enhancing tags, IB-targeted partners are believed to be more effective in masking the toxic activities of these peptides and protecting them from proteolytic degradation. Furthermore, insoluble expression facilitates the quick purification of these fusions. Even so, it remains a trial and error process for a given AMP to select a suitable fusion tag to effectively target it into IBs. Besides the potency of a fusion tag to form aggregates, its molecular size is another crucial factor affecting the final productivity of the passenger AMPs. The larger molecular size of a fusion tag implies the lower peptide-to-tag ratio, disfavoring the final yield of target peptides. Thus, screening and identification of effective shorter fusion tags are of practical significance [16].

Using s given fusion tag as template, fewer examples have been presented to illustrate how to refine its sequence to increase production of the passenger proteins. It has been proposed that the primary structure of a polypeptide intrinsically determine its propensity to aggregate [25]. Some short specific amino acid stretches paly crucial roles during aggregation as the initial nucleators (HSs) and can be effectively predicted [27]. We found that these HSs distributed unevenly along the PagP sequence, and the C-terminal region (residues 101-161) of PagP contained higher dense distribution of HSs. This finding promoted us to investigate whether this C-terminal region could alone retain high potency to direct passenger peptide into IBs. The promising absolute yields of the recombinant peptides demonstrated that these versions maintained high propensity to form aggregates (Fig. 5C and Fig. 6B). Nevertheless, we could not ignore the phenomena that these versions of PagP lost their efficiency in targeting potent antimicrobial peptide Mag II into IBs. Different approaches have revealed that IBs bare a characteristic cross-β structure. Moreover, a large number of IBs have been shown to have considerable biological activity, being characterized by a loose arrangement of protein molecules [38]. These versions of PagP might form less stable aggregates, making them vulnerable to proteolytic degradation induced by potent antimicrobial activity of peptide Mag II. Overall, we think that the strategy provided here is an effective attempt.

Bacterial IBs have been found to have some amyloid-like properties and contain similar structures [5]. Aβ 42 is an amyloidogenic peptide, in which a central hydrophobic stretch (residues 17-21) is predicted as a aggregation HS [27]. Reduction in hydrophobicity or β-sheet propensity in this stretch could impede its aggregation to different degrees [30]. Presence of the residue proline is normally regarded to decrease the overall protein aggregation propensity [39]. A hydrophobic region (residues 58-63) of Microcin E492 was identified. Compared with the wild type MccE492, mutants P57A and P59A exhibited the higher tendency to form amyloid aggregates *in vivo*, and aggregated significantly faster *in vitro* [40]. PagP is an out membrane protein rich in β-strands and ready to accumulate in IBs when expressed in *E. coli* cytoplasm [24, 41]. We found that the positions of aggregation HSs predicted by AGGRESCAN web tools were highly coincident with that of the β-strands (Fig. 3). Moreover, there are several proline residues (P121, P123, P127 and P135) inside or near the β6 strand. These findings prompted us to construct mutant P121L/P123L-PagP-NHT-Mag II, P127I-PagP-NHT-Mag II and P135L-PagP-NHT-Mag II, and investigated the effect of replacement of proline by hydrophobic residue leucine or isoleucine on their insoluble expression. The higher accumulation levels of these mutants further demonstrated that the aggregation capability of a given IBs-targeted fusion tag could be improved significantly by properly enhanced hydrophobicity. For newly synthesized recombinant proteins, the soluble properly folded forms, partially folded intermediates and insoluble aggregates co-exist in a dynamic equilibrium, and the partially folded form is very vulnerable to the proteolytic degradation [2]. The improvement in hydrophobicity or conformational stability of the HSs may allow the fusion protein to aggregate more rapidly, bringing about its enhanced resistance to the proteolytic degradation.

In many cases, fusion tag removal is necessary. For fusion proteins expressed in soluble form, a protease-sensitive linker sequence is often inserted between the fusion tag and the target protein. These fusion proteins are tractable to the site-specific cleavage by some proteases like thrombin or TEV protease [42]. However, this strategy may not be a feasible option when the fusion proteins exist in insoluble IBs. In this setting, fusion tag removal is biochemically challenged. Under denaturing conditions, utilization of cyanogen bromide in 70% formic acid for chemical cleavage may not only result in unwanted protein modifications, but also has much limitation due to the presence of residue methionine in the target peptide [2]. Fewer examples have been presented about how to separate target peptides from fusion proteins expressed in IBs. In this report, we designed a linker sequence NHT inserted between the fusion tag and AMPs (Fig. 1). This artificial peptide possessed three functional motifs, allowing the specific Ni (II)-catalyzed cleavage, Ni-chelating affinity chromatograph and protease TEV-mediated specific cleavage to be performed successfully. This approach is particularly applicable to production of the short unstructured peptide. We believe that the approach presented by us will promote extension of the IBs-targeted expression of short peptides.

Conclusions

Targeting recombinant proteins into inclusion bodies can make them refractory to proteolytic degradation effectively, especially for toxic or intrinsically disordered proteins, such as antimicrobial peptides. Based on the analysis of aggregation “hot spots” in fusion tag PagP, the version PagP-1 composed of the C-

terminal region of PagP with dense distribution of“hot spots”was identified, and has been demonstrated to effectively target recombinant antimicrobial peptide Metchnikowin and Andropin into inclusion bodies. Consequently, production of the recombinant antimicrobial peptides was increased significantly. Residue proline that is regarded to disfavour the formation of β-sheet was found to be distributed densely in the linking region between β-sheet element and the “hot spots”. Substitution for proline by residues with high β-sheet propensity and hydrophobicity in PagP could also greatly improve its ability to form aggregates, leading to the increased yield of the potent antimicrobial peptide Magainin II. The artificial peptide NHT designed here with three functional motifs allows the recombinant peptides to be conveniently purified. Refinement of a fusion tag for efficient inclusion body-targeted expression of the recombinant proteins is of significant practical importance. Few reports have been presented to improve the aggregation propensity of a fusion tag to increase production of a given passenger protein. Our results provide an alternative example to enhance the effectivity of a fusion tag,

Declarations

Acknowledgements

This work was supported by “Science and Technology Program of Guangzhou, China” (Grant No: 201510010072).

Author’s contributions

Xuefeng Li: Investigation, Visualization, Methodology, Formal analysis. **Baorong Zhang**: Investigation, Visualization, Methodology. **Quan Hu**: Methodology. **Changchao Chen**: Investigation, Visualization. **Wenshi Luo**: Methodology. **Lu Liu**: Investigation, Visualization. **Shengbin Wang**: Conceptualization, Supervision, Funding acquisition, Data curation.

Competing interests

The authors state that there are no competing interests to be reported.

Author details

[1] College of Life Sciences, South China Agricultural University, Wushan Road, Tianhe District, Guangzhou, People’s Republic of China.

References

1. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 2004;32:1037–49.
2. Hwang PM, Pan JS, Sykes BD. Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. FEBS Lett. 2014;588:247–52.

3. Villaverde A, Carrio MM. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol Lett. 2003;25:1385–95.
4. Sabate R, Espargaro A, Saupe SJ, Ventura S. Characterization of the amyloid bacterial inclusion bodies of the HET-s fungal prion. Microb Cell Fact. 2009;8:56.
5. Garcia-Fruitos E, Sabate R, de Groot NS, Villaverde A, Ventura S. Biological role of bacterial inclusion bodies: a model for amyloid aggregation. FEBS J. 2011;278:2419–27.
6. Kraus A, Anson KJ, Raymond LD, Martens C, Groveman BR, Dorward DW, Caughey B. Prion Protein Prolines 102 and 105 and the Surrounding Lysine Cluster Impede Amyloid Formation. J Biol Chem. 2015;290:21510–22.
7. Ramon A, Senorale-Pose M, Marin M. Inclusion bodies: not that bad. Front Microbiol. 2014;5:56.
8. Steinstraesser L, Kraneburg U, Jacobsen F, Al-Benna S. Host defense peptides and their antimicrobial-immunomodulatory duality. Immunobiology. 2011;216:322–33.
9. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol. 2005;3:238–50.
10. Scott MG, Davidson DJ, Gold MR, Bowdish D, Hancock RE. The human antimicrobial peptide LL-37 is a multifunctional modulator of innate immune responses. J Immunol. 2002;169:3883–91.
11. Niyonsaba F, Ushio H, Nagaoka I, Okumura K, Ogawa H. The human beta-defensins (-1, -2, -3, -4) and cathelicidin LL-37 induce IL-18 secretion through p38 and ERK MAPK activation in primary human keratinocytes. J Immunol. 2005;175:1776–84.
12. Niyonsaba F, Ushio H, Nakano N, Ng W, Sayama K, Hashimoto K, Nagaoka I, Okumura K, Ogawa H. Antimicrobial peptides human beta-defensins stimulate epidermal keratinocyte migration, proliferation and production of proinflammatory cytokines and chemokines. J Invest Dermatol. 2007;127:594–604.
13. Tjabringa GS, Ninaber DK, Drijfhout JW, Rabe KF, Hiemstra PS. Human cathelicidin LL-37 is a chemoattractant for eosinophils and neutrophils that acts via formyl-peptide receptors. Int Arch Allergy Immunol. 2006;140:103–12.
14. Kin NW, Chen Y, Stefanov EK, Gallo RL, Kearney JF. Cathelin-related antimicrobial peptide differentially regulates T- and B-cell function. Eur J Immunol. 2011;41:3006–16.
15. Kovacs-Nolan J, Mapletoft JW, Latimer L, Babiuk LA, Hurk S. CpG oligonucleotide, host defense peptide and polyphosphazene act synergistically, inducing long-lasting, balanced immune responses in cattle. Vaccine. 2009;27:2048–54.
16. Li Y. Carrier proteins for fusion expression of antimicrobial peptides in *Escherichia coli*. Biotechnol Appl Biochem. 2009;54:1–9.
17. Chen YQ, Zhang SQ, Li BC, Qiu W, Jiao B, Zhang J, Diao ZY. Expression of a cytotoxic cationic antibacterial peptide in *Escherichia coli* using two fusion partners. Protein Expr Purif. 2008;57:303–11.

18. Pane K, Durante L, Pizzo E, Varcamonti M, Zanfardino A, Sgambati V, Di Maro A, Carpentieri A, Izzo V, Di Donato A, et al. Rational Design of a Carrier Protein for the Production of Recombinant Toxic Peptides in *Escherichia coli*. PLoS One. 2016;11:e0146552.
19. Wibowo D, Zhao CX. Recent achievements and perspectives for large-scale recombinant production of antimicrobial peptides. Appl Microbiol Biotechnol. 2019;103:659–71.
20. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr Opin Biotechnol. 1995;6:494–500.
21. Puigbo P, Guzman E, Romeu A, Garcia-Vallve S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 2007;35:W126–31.
22. Krezel A, Kopera E, Protas AM, Poznanski J, Wyslouch-Cieszynska A, Bal W. Sequence-specific Ni(II)-dependent peptide bond hydrolysis for protein engineering. Combinatorial library determination of optimal sequences. J Am Chem Soc. 2010;132:3355–66.
23. Kopera E, Belczyk-Ciesielska A, Bal W. Application of Ni(II)-assisted peptide bond hydrolysis to non-enzymatic affinity tag removal. PLoS One. 2012;7:e36350.
24. Hwang PM, Choy WY, Lo EI, Chen L, Forman-Kay JD, Raetz CR, Prive GG, Bishop RE, Kay LE. Solution structure and dynamics of the outer membrane enzyme PagP by NMR. Proc Natl Acad Sci U S A. 2002;99:13560–5.
25. Nerelius C, Fitzen M, Johansson J. Amino acid sequence determinants and molecular chaperones in amyloid fibril formation. Biochem Biophys Res Commun. 2010;396:2–6.
26. Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem. 2006;75:333–66.
27. Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics. 2007;8:65.
28. Oren Z, Shai Y. Mode of action of linear amphipathic alpha-helical antimicrobial peptides. Biopolymers. 1998;47:451–63.
29. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R. Bacterial inclusion bodies contain amyloid-like structure. PLoS Biol. 2008;6:e195.
30. de Groot NS, Aviles FX, Vendrell J, Ventura S. Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. FEBS J. 2006;273:658–68.
31. Morimoto A, Irie K, Murakami K, Masuda Y, Ohigashi H, Nagao M, Fukuda H, Shimizu T, Shirasawa T. Analysis of the secondary structure of beta-amyloid (Abeta42) fibrils by systematic proline replacement. J Biol Chem. 2004;279:52781–8.
32. Fujiwara K, Toda H, Ikeguchi M. Dependence of alpha-helical and beta-sheet amino acid propensities on the overall protein fold type. BMC Struct Biol. 2012;12:18.

33. Zasloff M. Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. Proc Natl Acad Sci U S A. 1987;84:5449–53.
34. Samakovlis C, Kylsten P, Kimbrell DA, Engstrom A, Hultmark D. The andropin gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster*. EMBO J. 1991;10:163–9.
35. Levashina EA, Ohresser S, Bulet P, Reichhart JM, Hetru C, Hoffmann JA. Metchnikowin, a novel immune-inducible proline-rich peptide from *Drosophila* with antibacterial and antifungal properties. Eur J Biochem. 1995;233:694–700.
36. Nguyen LT, Haney EF, Vogel HJ. The expanding scope of antimicrobial peptide structures and their modes of action. Trends Biotechnol. 2011;29:464–72.
37. Li Y. Recombinant production of antimicrobial peptides in *Escherichia coli*: a review. Protein Expr Purif. 2011;80:260–7.
38. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. Microb Cell Fact. 2015;14:41.
39. Beerten J, Jonckheere W, Rudyak S, Xu J, Wilkinson H, De Smet F, Schymkowitz J, Rousseau F. Aggregation gatekeepers modulate protein homeostasis of aggregating sequences and affect bacterial fitness. Protein Eng Des Sel. 2012;25:357–66.
40. Aguilera P, Marcoleta A, Lobos-Ruiz P, Arranz R, Valpuesta JM, Monasterio O, Lagos R. Identification of Key Amino Acid Residues Modulating Intracellular and In vitro Microcin E492 Amyloid Formation. Front Microbiol. 2016;7:35.
41. Hwang PM, Pan JS, Sykes BD. A PagP fusion protein system for the expression of intrinsically disordered proteins in Escherichia coli. Protein Expr Purif. 2012;85:148–51.
42. Waugh DS. An overview of enzymatic reagents for the removal of affinity tags. Protein Expr Purif. 2011;80:283–93.

Figures

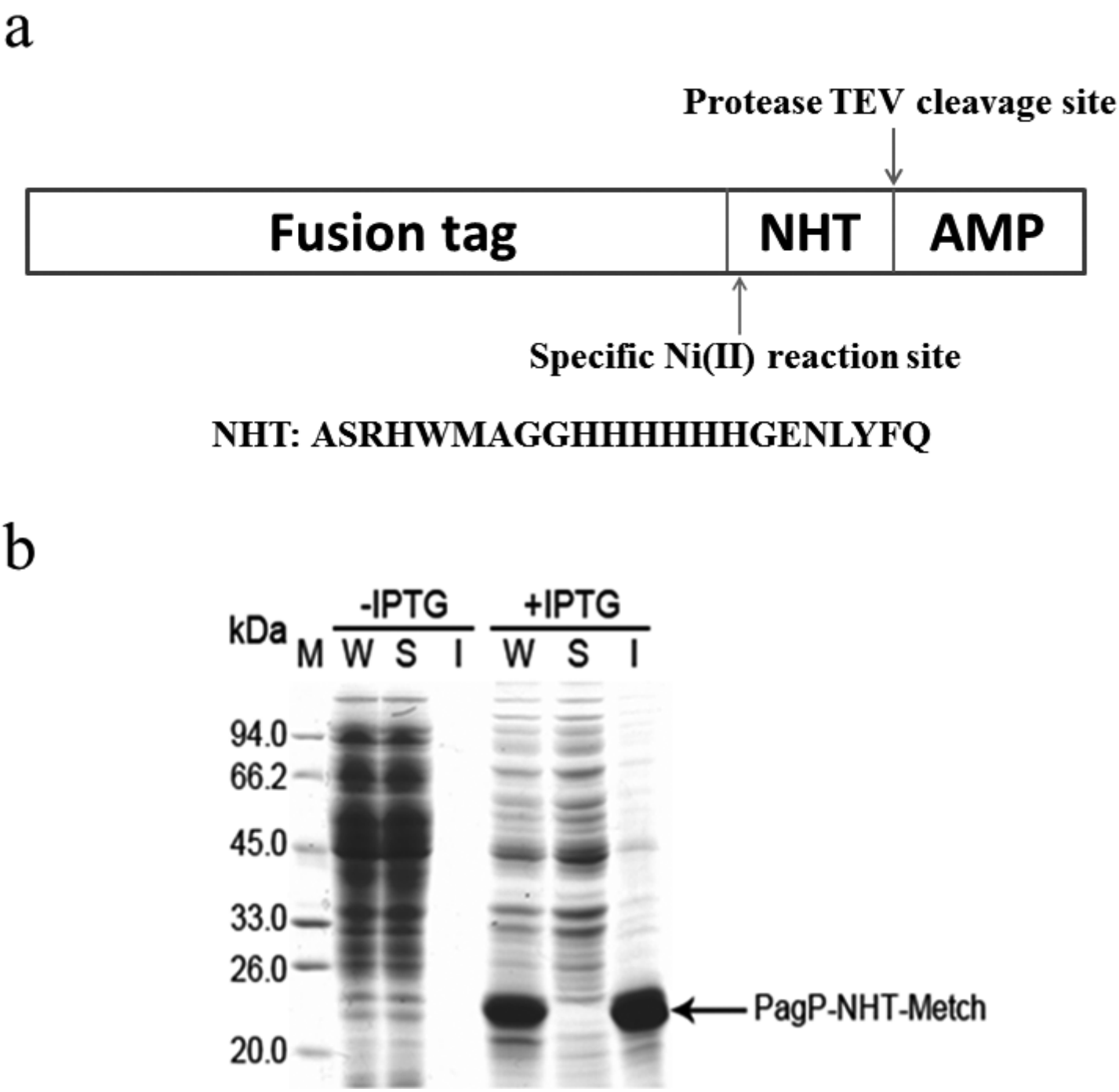

Figure 1

Schematic representation of the artificial peptide NHT (Panel a) and SDS-PAGE analysis of the expression of recombinant fusion protein PagP-NHT-Metch (Panel b). W: whole cell lysate; S: soluble supernatant after sonication; I: inclusion bodies.

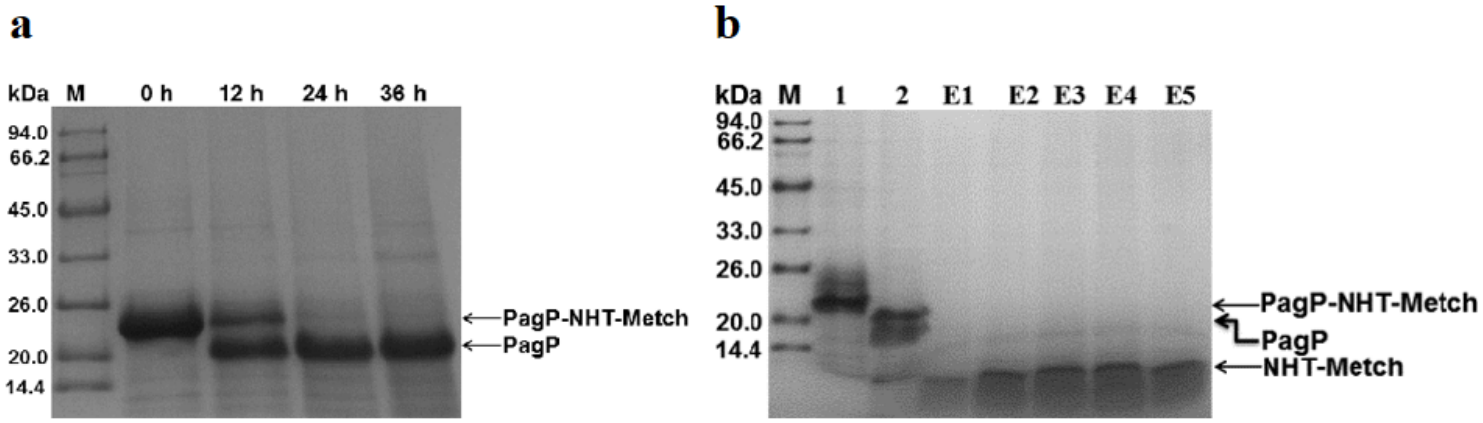

Figure 2

Ni+2-catalyzed cleavage of the fusion protein PagP-NHT-Metch and purification of the fusion peptide NHT-Metch. Panel a: Tris-SDS-PAGE analysis. Under denaturing condition, the fusion protein PagP-NHT-Metch was subjected to Ni+2-catalyzed cleavage at 60°C for 12 h, 24 h and 36 h, respectively. Panel b: Tricine-SDS PAGE analysis. Purification of the fusion peptide NHT-Metch by Ni-chelating affinity chromatograph. Lane 1: the fusion protein PagP-NHT-Metch; lane 2: products after Ni+2-catalyzed cleavage; lane E1, lane E2, lane E3, lane E4 and lane E5: the fusion peptide NHT-Metch eluted by elution buffer.

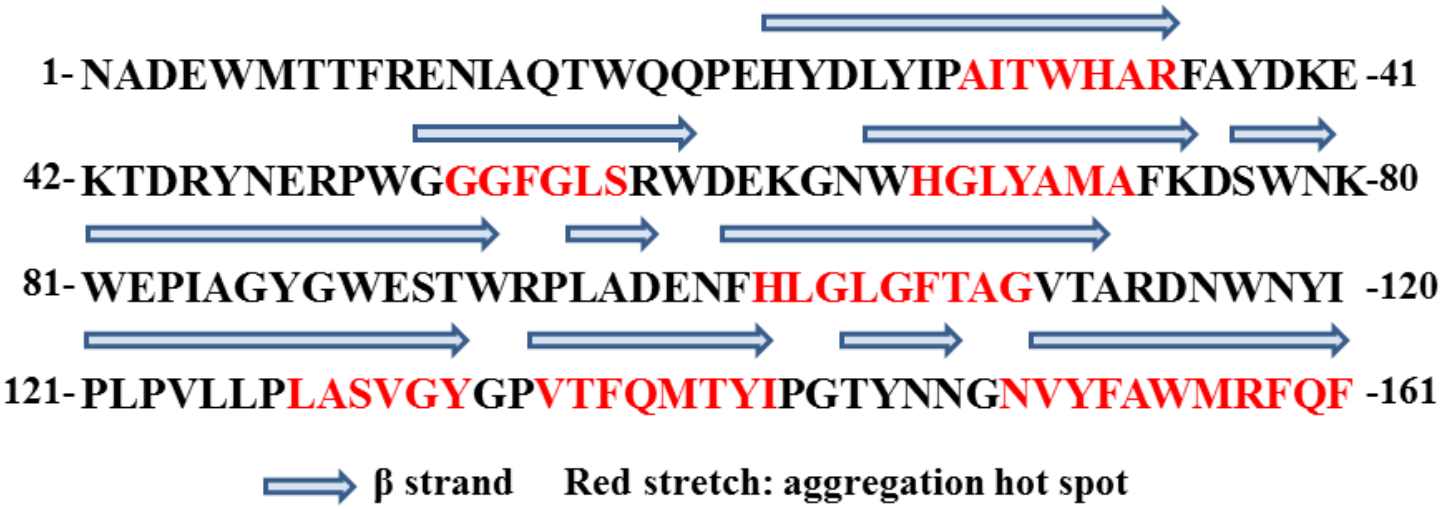

Figure 3

Schematic representation of the β-sheet elements in PagP reported previously [24], and the aggregation “hot sports” (HSs) predicted by web-based software AGGRESCAN [27].

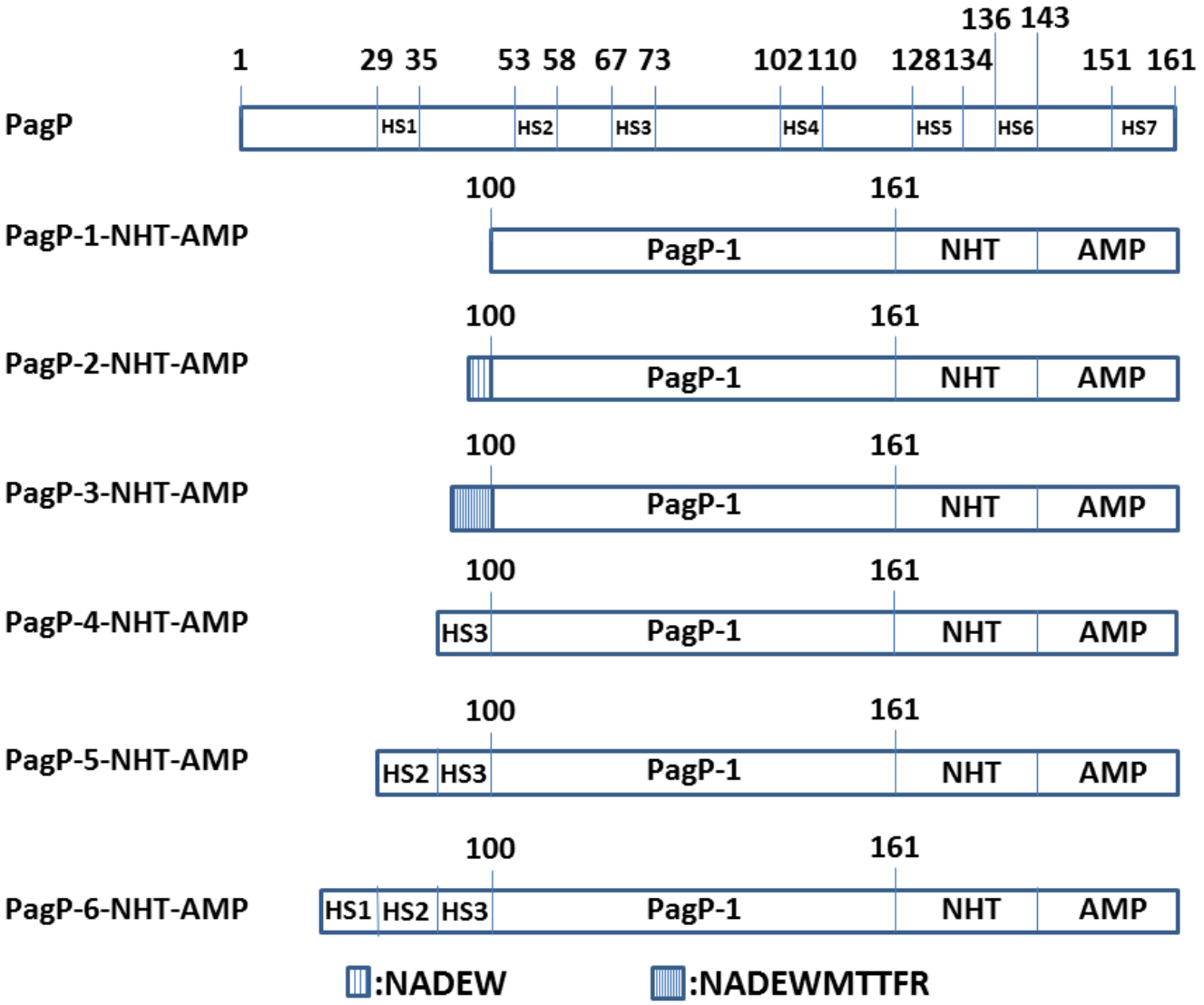

Figure 4

Schematic representation of construction of the fusion proteins.

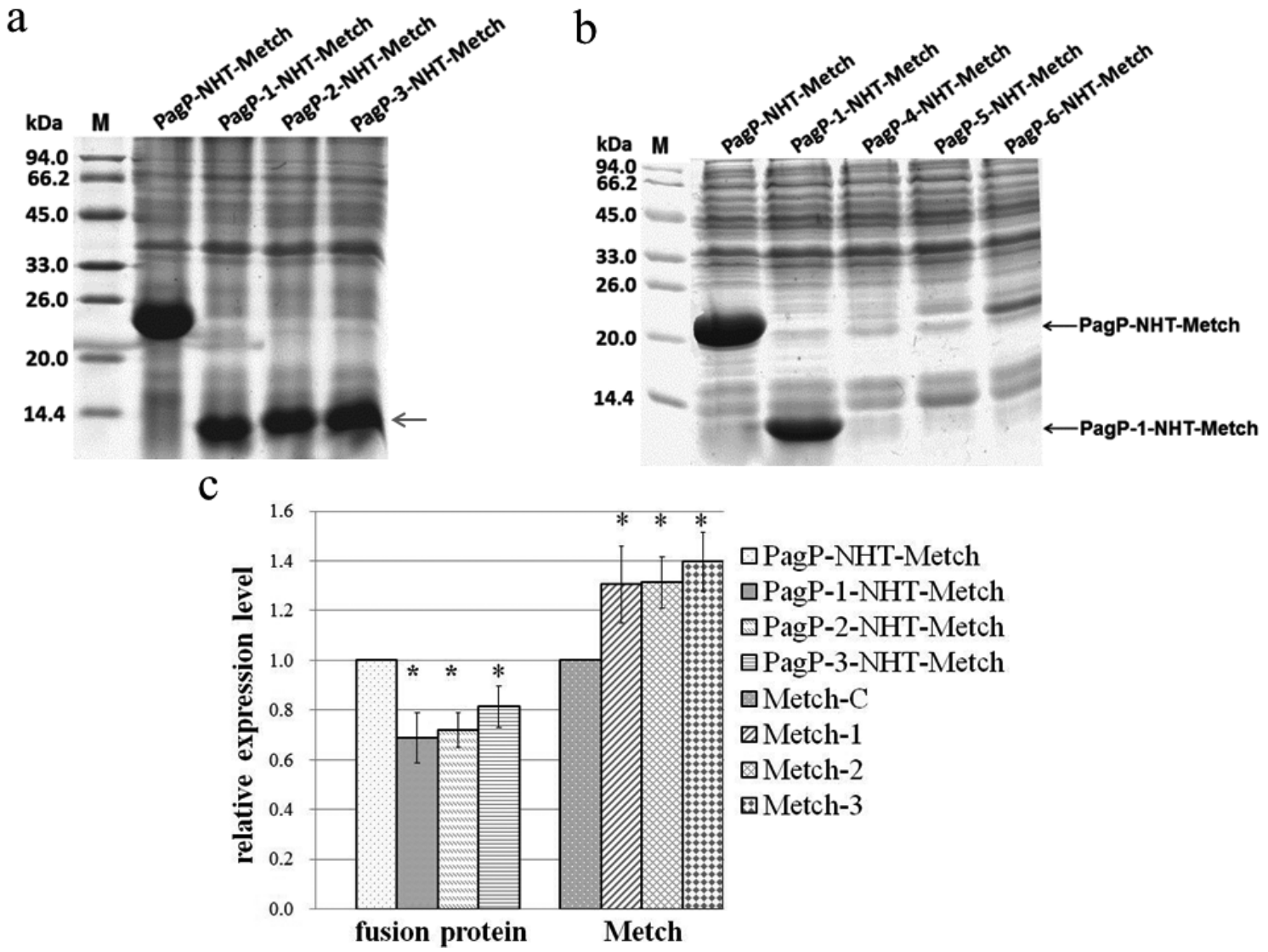

Figure 5

Refinement of the fusion tag PagP for efficient expression of antimicrobial peptide Metch. The expressions of six fusion proteins were analyzed by SDS-PAGE (Panel a and Panel b). Panel c, the densitometric value of the band of PagP-NHT-Metch was set to be 1. The yield of recombinant Metch was calculated according to the expression levels of fusion proteins and their corresponding molecular weight. Each bar represented the mean ± standard error, three replicates, *: $P < 0.05$. Metch-C: the yield of recombinant peptide Metch expressed by the fusion protein PagP-NHT-Metch, it was set to be 1; Metch-1: the yield of recombinant peptide Metch expressed by the fusion protein PagP-1-NHT-Metch; Metch-2: the yield of recombinant peptide Metch expressed by the fusion protein PagP-2-NHT-Metch; Metch-3: the yield of recombinant peptide Metch expressed by the fusion protein PagP-3-NHT-Metch. The molecular weight of PagP and PagP-1 is 18.98 kDa and 6.87 kDa, respectively.

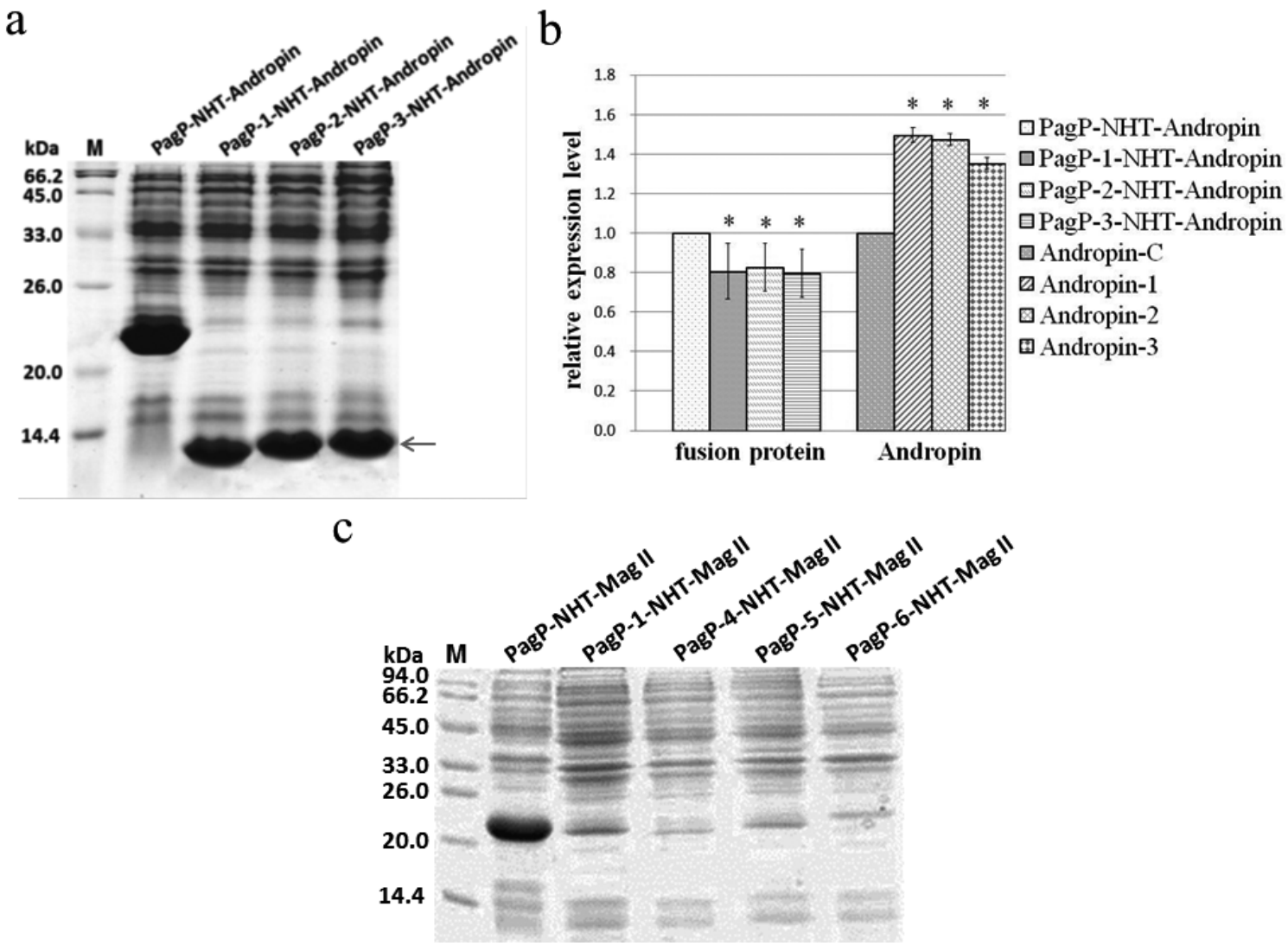

Figure 6

The expression of antimicrobial peptide Mag II and Andropin in fusion with the refined fusion tags. Panel a, the antimicrobial peptide Andropin was fused to PagP-1, PagP-2 and PagP-3 respectively, and their expression was analyzed by SDS-PAGE. Panel b, the densitometric value of the band of PagP-NHT-Andropin was set to be 1. The yield of recombinant Andropin was calculated according to the expression levels of fusion proteins and their corresponding molecular weight. Each bar represented the mean ± standard error, three replicates, *: $P < 0.05$. Andropin-C: the yield of recombinant Andropin expressed by the fusion protein PagP-NHT-Andropin, it was set to be 1; Andropin-1: the yield of recombinant Andropin expressed by the fusion protein PagP-1-NHT-Andropin; Andropin-2: the yield of recombinant Andropin expressed by the fusion protein PagP-2-NHT-Andropin; Andropin-3: the yield of recombinant Andropin expressed by the fusion protein PagP-3-NHT-Andropin. Panel c, antimicrobial peptide Mag II was fused to PagP, PagP-1, PagP-4, PagP-5 and PagP-6, respectively, and their expression was analyzed by SDS-PAGE.

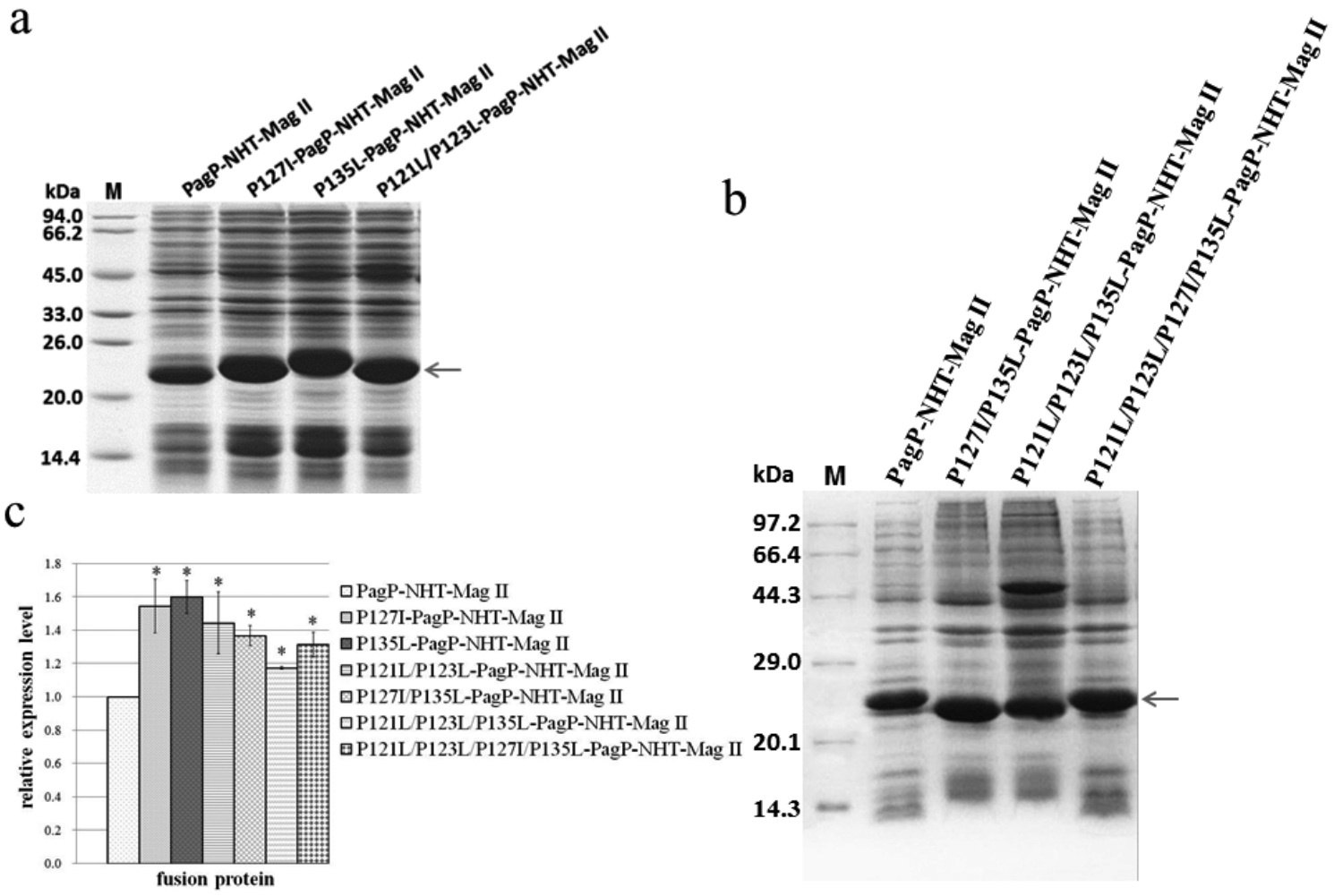

Figure 7

Mutation of the fusion tag PagP for efficient expression of the antimicrobial peptide Mag II. The mutated PagPs were fused to NHT-Mag II, respectively, and their expression was analyzed by SDS-PAGE (Panel a and Panel b). Panel c, the densitometric value of the band of PagP-NHT-Mag II was set to be 1. Each bar represented the mean ± standard error, three replicates, *: $P < 0.05$.

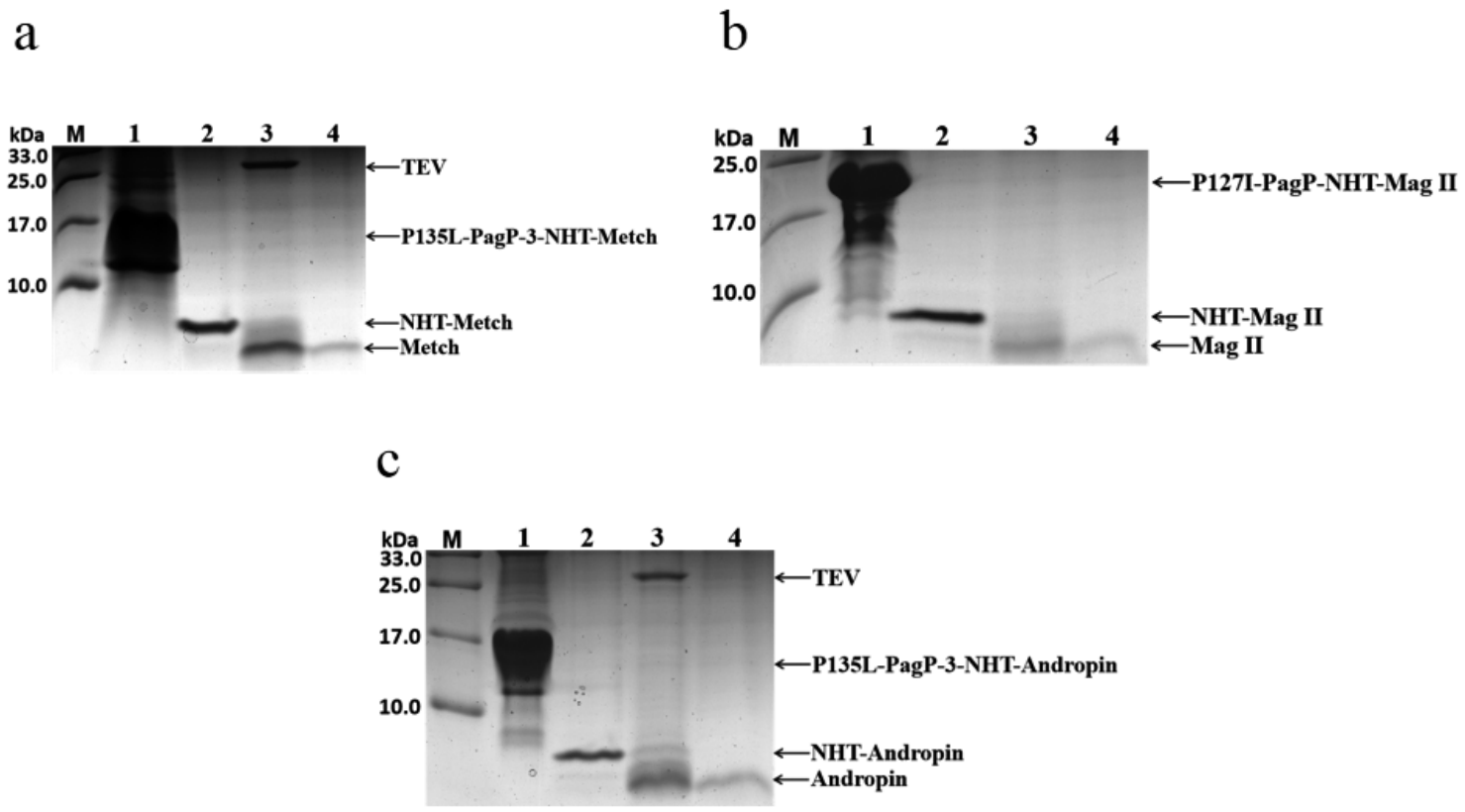

Figure 8

Purification of the recombinant antimicrobial peptides. Panel a, Lane 1: the fusion protein PagP-3-NHT-Metch; lane 2: the purified NHT-Metch; lane 3: the products of NHT-Metch targeted to site-specific cleavage by protease TEV; lane 4: the purified antimicrobial peptide Metch. Panel b, Lane 1: the fusion protein P135L-PagP-NHT-Mag II; lane 2: the purified NHT-Mag II; lane 3: the products of NHT-Mag II targeted to the site-specific cleavage by protease TEV; lane 4: the purified antimicrobial peptide Mag II. Panel c, Lane 1: the fusion protein PagP-2-NHT-Andropin; lane 2: the purified NHT-Andropin; lane 3: the products of NHT-Andropin targeted to the site-specific cleavage by protease TEV; lane 4: the purified antimicrobial peptide Andropin.

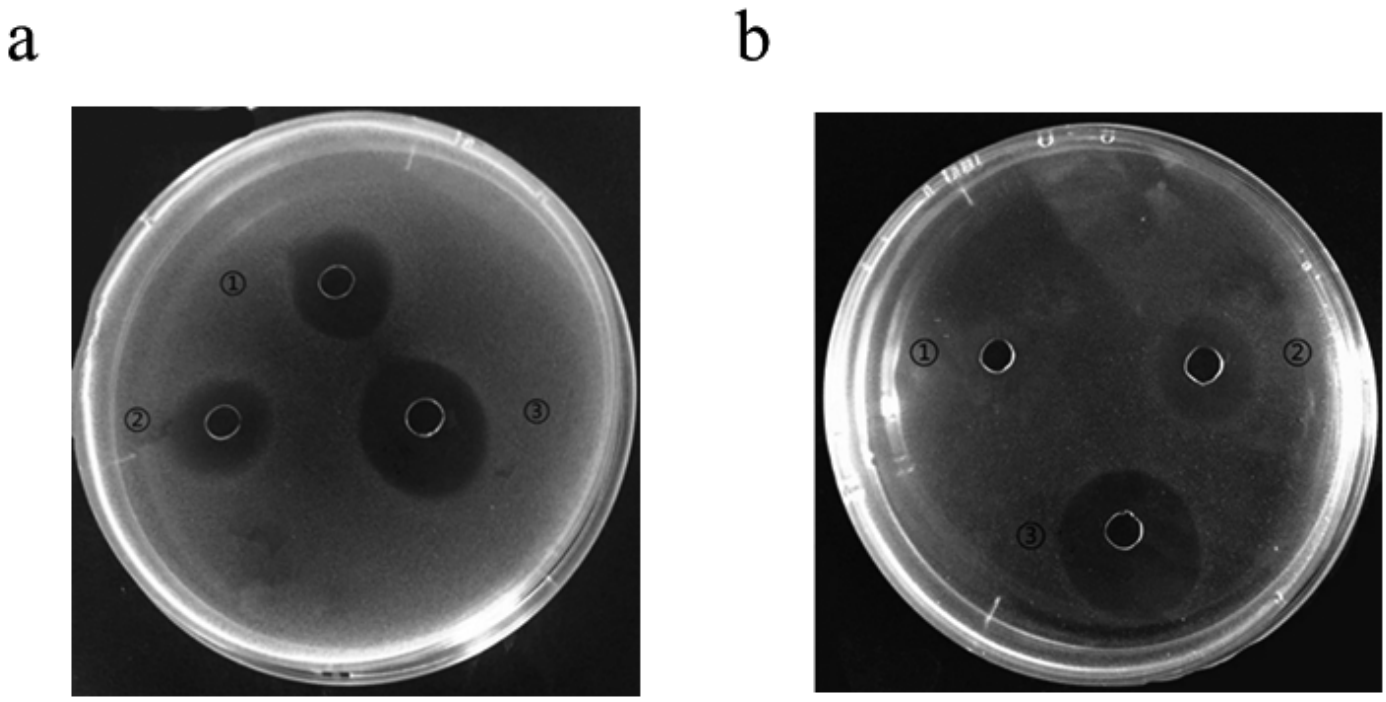

Figure 9

The antibacterial activity assay of recombinant antimicrobial peptides against Staphylococcus aureus ATCC 25923 (Panel a) and E. coli MG1655 (Panel b) by appearance of the inhibition zones. □: 30 μl purified recombinant Metch; □: 30 μl purified recombinant Andropin; □: 30 μl purified recombinant Mag II.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx