Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Efficient recombinant production of a short silica-forming peptide in Escherichia coli using an intrinsically disordered fusion tag

Gyun Taek Lim
Gyeongsang National University
Byung Hoon Jo
jobh@gnu.ac.kr

Gyeongsang National University

Research Article

Keywords: recombinant peptide production, intrinsically disordered peptide, NEXT tag, Escherichia coli, proteolytic cleavage, R5 peptide, biomimetic silica

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11746962/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Recombinant production of short peptides in microbial cell factories often relies on fusion tags that support intracellular accumulation and efficient peptide release. Large tags, however, account for much of the recombinant product, potentially wasting cellular resources. We evaluated the small, intrinsically disordered NEXT tag against the maltose-binding protein (MBP), a large globular tag, for the production of the silica-forming R5 peptide bearing a C-terminal hexahistidine tag (R5-His_6).

Results

Direct expression yielded no detectable R5-His_6 in *Escherichia coli* BL21(DE3) or C41(DE3), whereas both fusion proteins accumulated in both hosts. The NEXT fusion was detected exclusively in the soluble fraction, while the MBP fusion was partly insoluble. Soluble R5-His_6-equivalent expression yields were 46% and 37% higher with NEXT than with MBP in BL21(DE3) and C41(DE3), respectively, reaching 23.9 mg L^{-1} in C41(DE3). During cleavage of R5-His_6 from the fusions in cell lysates with separately expressed tobacco etch virus protease, the NEXT fusion reached near-complete cleavage within 60 min, whereas a fraction of the MBP fusion remained uncleaved. Affinity purification following cleavage of the NEXT fusion produced in C41(DE3) yielded approximately 13 mg of R5-His_6 per liter of culture. The recovered peptide readily induced biomimetic silica precipitation.

Conclusions

Despite its considerably smaller size, NEXT combined higher soluble peptide-equivalent expression yields with faster protease-mediated release than MBP. These findings establish NEXT as an effective fusion tag for the production of recombinant R5-His_6 and support its further evaluation for other short functional peptides.

Background

Short functional peptides are increasingly employed in therapeutics, biosensing, and biomaterials engineering [1–3]. Chemical peptide synthesis is widely used to produce short peptides and offers flexibility in introducing chemical modifications. However, it often requires large amounts of reagents and solvents, increasing production costs and waste generation [4, 5]. Incomplete reactions can also generate peptide impurities that complicate purification [4]. Recombinant expression in microbial hosts such as *Escherichia coli* offers an alternative route for peptide production [6–8]. Nevertheless, short peptides lacking a stable fold may be susceptible to intracellular proteolysis, limiting their accumulation

[9]. Depending on their sequence and biological activity, some peptides may also be toxic to the host cell [6, 7].

Fusion expression is widely used to improve the accumulation of short peptides by protecting them from intracellular degradation and, in some cases, reducing their adverse effects on the host [6, 7, 9, 10]. Inclusion-body-forming tags such as ketosteroid isomerase (KSI; 13.5 kDa) and PagP (19.0 kDa) promote insoluble accumulation and facilitate initial recovery by centrifugation, but generally require solubilization before peptide release [10, 11]. Alternatively, solubility-enhancing tags such as maltose-binding protein (MBP; 40.4 kDa), thioredoxin (Trx; 11.8 kDa), and small ubiquitin-like modifier (SUMO; 11.3 kDa) promote soluble accumulation, enabling purification and proteolytic release without the solubilization step [6, 12, 13]. In either approach, however, when the tag is much larger than the target peptide, biosynthesis of the fusion protein consumes cellular resources primarily to produce the fusion partner, which is subsequently removed during downstream processing. Smaller tags may therefore improve peptide productivity, provided that they still support efficient expression and protect the target peptide from intracellular degradation [9, 14]. Beyond expression, downstream recovery requires efficient, site-specific cleavage to release the intact target peptide [8]. Accessibility of the cleavage site is therefore another important consideration in fusion-tag design [15].

Intrinsically disordered proteins lack a stable three-dimensional structure and instead adopt a dynamic ensemble of different conformations [16]. When used as fusion tags, they can adopt expanded conformations, creating a larger excluded volume around the point of attachment than compact globular tags of comparable molecular mass [17–19]. These steric effects may suppress intermolecular association and aggregation, allowing a small disordered tag to promote soluble expression of fusion proteins. The NEXT tag (5.5 kDa) is a 53-residue intrinsically disordered tag previously developed to enhance soluble recombinant protein production in *E. coli* [18]. It also improves the colloidal stability of purified proteins by suppressing precipitation [18]. Despite its much smaller size, NEXT has been shown to outperform MBP in enhancing recombinant protein expression and solubility [18]. Whether these advantages extend to short peptides, whose production requires both sufficient intracellular accumulation and efficient release from the fusion tag, remains to be examined.

Here, we investigated the NEXT tag for the recombinant production of the silica-forming R5 peptide and compared its performance with that of MBP. R5 is a 19-residue Lys-rich cationic peptide (SSKKSGSYSGSKGSKRRIL) derived from a diatom silaffin protein sequence [20]. As one of the most widely studied silica-forming peptides, R5 efficiently promotes silica precipitation under mild aqueous conditions [3, 20–22]. This activity has been exploited for biomolecule encapsulation and immobilization in applications including enzyme immobilization, drug delivery, and biosensing [3, 22–28]. MBP was selected as a benchmark tag based on its prior use in the production of recombinant R5 [12]. A C-terminal hexahistidine tag (His_6) was appended to R5, generating R5-His_6, to facilitate affinity purification following its proteolytic release from the fusion proteins. Direct and fusion expression of R5-His_6 were examined in *E. coli* BL21(DE3) and C41(DE3). We compared soluble R5-His_6-equivalent expression yields, in-lysate cleavage, and affinity purification profiles for the two fusions and quantified purified peptide

recovery from the NEXT fusion. The purified R5-His_6 was then assessed for molecular integrity and silica-forming activity.

Methods

Plasmid construction

Escherichia coli TOP10 (Thermo Fisher Scientific, USA) was used for gene cloning and plasmid propagation. Four expression plasmids encoding R5-His_6, MBP-R5-His_6, NEXT-R5-His_6, and MBP-TEVp were constructed using pET-22b(+) (Novagen, USA). The R5 coding sequence was amplified from a previously constructed bCA-R5 plasmid [29]. For fusion expression, NEXT- and MBP-containing fragments were amplified from the pET-NEXT-GFPuv and pET-MBP-GFPuv plasmids [18], respectively, and joined to the R5 fragment by overlap extension PCR. The R5 insert for direct expression and the two assembled fusion inserts were cloned into pGEM-T Easy (Promega, USA), verified by sequencing, and subcloned into pET-22b(+) using the NdeI and XhoI sites.

In both R5 fusions, the tag was followed by a $(G_4S)_2$ linker and a TEVp cleavage junction (ENLYFQ↓S), with the first residue of R5 serving as the P1' serine. All three R5 constructs encoded the vector-derived C-terminal extension sequence LEHHHHHH. The direct expression construct also encoded an initiator methionine immediately before R5.

The TEVp coding sequence was synthesized by GenScript (USA). In the MBP-TEVp construct, MBP and TEVp were joined by the same linker and cleavage junction as used in the R5 fusions. The TEVp coding sequence encoded the S219V substitution [30] and a C-terminal Arg_5 tail [31]. Self-cleavage at the junction released TEVp with an SMGES sequence at its N terminus. A stop codon was inserted upstream of the XhoI site to prevent addition of the vector-encoded His_6 tag. The nucleotide and amino acid sequences of all constructs are provided in **Supplementary Table S1**.

Recombinant protein expression

The R5-His_6 expression constructs were introduced into *E. coli* BL21(DE3) (Novagen, USA) and C41(DE3) (Sigma-Aldrich, USA), whereas the MBP-TEVp plasmid was introduced into BL21(DE3) only. Seed cultures were grown for 10 h and used to inoculate 200 mL of Luria–Bertani (LB) medium supplemented with 50 μg mL^{-1} ampicillin at 1% (v/v). Cultures were grown in 500-mL Erlenmeyer flasks (DURAN; DWK Life Sciences, Germany) at 37°C with shaking at 180 rpm. Expression was induced at an OD_{600} of 0.6–0.8 by adding isopropyl β-D-thiogalactopyranoside (IPTG; Duchefa Biochemie, The Netherlands) to a final concentration of 1 mM. After induction, cultures for direct and fusion expression of R5-His_6 were maintained at 37°C, whereas cultures for TEVp production were shifted to 30°C.

To monitor R5 fusion expression over time, 10-mL samples were taken from both strains immediately before induction (0 h) and at 4, 8, and 12 h after induction. A separate set of C41(DE3) cultures was sampled at 12, 18, and 24 h after induction. For subsequent production of the R5 fusions, BL21(DE3) and C41(DE3) cultures were harvested 8 and 12 h after induction, respectively. Cultures expressing R5-His_6 directly (both strains) or producing TEVp (BL21(DE3) only) were harvested 12 h after induction.

Cell lysis and fractionation

Cells were harvested by centrifugation at 4,000 × g for 10 min at 4°C. Cell pellets from 10-mL time-course samples and 200-mL preparative cultures were resuspended in 1.5 and 30 mL of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, and 10 mM imidazole, pH 8.0), respectively. Cells were disrupted on ice by sonication using an ultrasonic dismembrator (Sonics and Materials, USA). The lysates were centrifuged at 10,000 × g for 10 min at 4°C, and the supernatants were collected as the soluble fractions. For analysis of the insoluble fractions, the pellets were resuspended in lysis buffer to a volume equal to that of the corresponding supernatants.

SDS-PAGE and gel analysis

SDS-PAGE and gel analysis
Samples were mixed with 4× sample buffer (Bio-Rad, USA) at a sample-to-buffer ratio of 3:1 (v/v) and heated at 100°C for 10 min. Samples were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 15% Tris–glycine gels. Gels were stained with 0.1% Coomassie Brilliant Blue R-250 (Bio-Rad, USA) or, for His-tag detection, with InVision His-Tag In-Gel Stain (Thermo Fisher Scientific) according to the manufacturer's instructions. Band intensities were quantified by densitometry using ImageJ for Coomassie-stained gels and a GelDoc Go imaging system (Bio-Rad, USA) for InVision-stained gels.

For time-course analysis, soluble and insoluble fractions were loaded on an equivalent culture-volume basis, and total accumulation was estimated by summing the intensities of the soluble and insoluble fusion bands. Soluble MBP-R5-His_6 was also quantified separately. Each series was normalized within each experiment to its own 8-h value for BL21(DE3) or 12-h value for C41(DE3).

To determine soluble fusion yields, cleared lysates were analyzed alongside known amounts of the corresponding purified fusion protein, with a separate linear calibration curve for each fusion in each experiment. Yields were calculated from the calibration curves, accounting for sample dilution, lysate volume, and culture volumes. R5-His_6-equivalent yields were obtained by multiplying the soluble fusion yields by the ratio of the molecular mass of released R5-His_6 to that of the intact fusion protein.

The original uncropped Coomassie-stained and in-gel His-tag-stained gel images corresponding to the SDS-PAGE panels in Figs. 2–5 are provided in Supplementary Figs. S1–S4, respectively.

In-lysate cleavage

A stock solution of protease inhibitor cocktail (Sigma-Aldrich, USA; cat. no. P2714) was prepared by dissolving the contents of one vial in 10 mL of water. To limit degradation of released R5-His_6 during cleavage, 300 µL of the stock was added per 30 mL of lysis buffer for cells expressing either R5 fusion before sonication. Control lysates were prepared without the cocktail. The lysis buffer for TEVp-producing cells was supplemented with 1 mM dithiothreitol (DTT; Sigma-Aldrich, USA).

For cleavage, cleared *E. coli* C41(DE3) lysate containing either MBP-R5-His_6 or NEXT-R5-His_6 was mixed with cleared *E. coli* BL21(DE3) lysate containing His-tag-free TEVp at a 9:1 (v/v) ratio. The mixtures were incubated at 30°C without shaking. For time-course SDS-PAGE analysis, aliquots were collected at 5, 10, 20, 30, 40, 60, and 90 min. Cleavage was stopped in each aliquot by immediately adding 4× sample buffer and heating at 100°C. For the 0-min reference, an equal volume of lysis buffer was used in place of the TEVp lysate. Apparent cleavage C(t) (%) at time t (min) was calculated as $100(1 - I_t/I_0)$, where I_t and I_0 are the intensities of the uncleaved fusion band at time t and 0 min, respectively. For each fusion, the mean apparent cleavage values at each time point were fitted in SigmaPlot (Systat Software, USA) to a single-exponential model with the plateau fixed at 100%:

$$C\,(\mathrm{t}) = 100(1 - \mathrm{e}^{-k_{\mathrm{obs}}\mathrm{t}})$$

where k_{obs} is the observed cleavage rate constant (min^{-1}). For preparative recovery, the NEXT fusion was cleaved for 60 min and the MBP fusion for either 60 or 90 min.

Ni-NTA affinity purification

His_6-tagged proteins and peptides were purified by affinity chromatography using nickel-nitrilotriacetic acid (Ni-NTA) agarose beads (QIAGEN, USA). Cleared lysates were applied to the resin to assess R5-His_6 recovery from direct expression cultures and to purify intact MBP-R5-His_6 and NEXT-R5-His_6 for use as standards. Released R5-His_6 was purified from the lysate mixtures after TEVp cleavage. Samples were incubated with the resin on ice for 30 min. The resin was washed with wash buffer (50 mM sodium phosphate, 300 mM NaCl, and 30 mM imidazole, pH 8.0). Bound proteins or peptides were eluted with elution buffer (50 mM sodium phosphate, 300 mM NaCl, and 250 mM imidazole, pH 8.0). After elution, MBP-R5-His_6 and NEXT-R5-His_6 were dialyzed at 4°C against deionized water using Spectra/Por tubing with a molecular weight cutoff (MWCO) of 6–8 kDa (Thermo Fisher Scientific). Released R5-His_6 was dialyzed against deionized water using Spectra/Por tubing with a 1 kDa MWCO (Thermo Fisher Scientific).

Protein concentration determination

For quantification, aliquots of dialyzed MBP-R5-His_6 were denatured at a final guanidine hydrochloride (GuHCl) concentration of 6 M, and absorbance was measured at 280 nm. The concentration was calculated using the molar extinction coefficient obtained with ExPASy ProtParam [32]. Dialyzed NEXT-R5-His_6 and released R5-His_6 were quantified at 205 nm using molar extinction coefficients calculated from their respective amino acid sequences [33]. Absorbance was measured with a UV-1800 spectrophotometer (Shimadzu, Japan) against the corresponding solvent blank, and concentrations were corrected for sample dilution.

To determine the yield of released R5-His_6, equal volumes of the Ni-NTA eluate and the corresponding water-dialyzed sample were analyzed on the same SDS-PAGE gel. The peptide concentration in the eluate was estimated by multiplying the concentration of the dialyzed sample by the ratio of the R5-His_6 band intensities in the eluate and dialyzed sample. The estimated mass concentration was multiplied by the eluate volume, and the recovered peptide mass was expressed in milligrams per liter of culture.

Mass spectrometry

Purified R5-His_6 was analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) on an autoflex maX MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Germany). Mass spectra were acquired over an m/z range of 700–4,000 with a 355 nm smartbeam-II laser operated at a repetition rate of 2,000 Hz. The mass scale was calibrated using the Peptide Calibration Standard (Bruker Daltonics), and the spectra were analyzed with flexAnalysis (Bruker Daltonics).

Silica precipitation

The silica precursor was prepared by hydrolyzing 1 M tetramethyl orthosilicate (TMOS; Sigma-Aldrich, USA) in 1 mM HCl for 15 min at room temperature. Purified R5-His_6 was dialyzed against 20 mM sodium phosphate buffer (pH 7.5). To initiate silica precipitation, 100 µL of hydrolyzed TMOS was added to 900 µL of the dialyzed peptide solution. A reaction with a blank buffer served as a negative control. After 30 min at room temperature, the precipitate was collected by centrifugation at 10,000 × g for 10 min at 4°C, washed three times with deionized water, and dried at room temperature for further analyses.

SEM and EDX analyses

The dried silica particles were coated with Pt and imaged using a field-emission scanning electron microscope (SEM; JEOL, Japan) equipped with an energy-dispersive X-ray spectroscopy (EDX) system (Oxford Instruments, UK). Elemental weight percentages (wt%) were determined from the summed EDX spectrum of the mapped area using AZtecLive.

Use of large language models

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to assist with language editing and manuscript refinement. All AI-assisted outputs were critically reviewed, verified, and edited by the authors, who take full responsibility for the final content of the manuscript.

Results

Design of an expression system for recombinant R5 peptide production

Three constructs were generated to compare direct expression of R5-His_6 with fusion expression using MBP or NEXT as an N-terminal fusion tag (Fig. 1a). The two tags differ in size and structural organization: MBP is a large (40.4 kDa) globular protein, whereas NEXT is a small (5.5 kDa) intrinsically disordered tag [18]. In both fusion constructs, the tag was followed by a flexible $(G_4S)_2$ linker and a TEVp cleavage site. The TEVp recognition sequence ENLYFQ was placed immediately upstream of R5, positioning its authentic N-terminal serine at the P1' position, where serine is a preferred residue for TEVp [34]. This arrangement was designed to release R5-His_6 without additional N-terminal residues. In contrast, the direct-expression construct encoded an initiator methionine preceding the R5 sequence. All three constructs were placed under the control of a T7*lac* promoter and introduced into *E. coli* BL21(DE3) and C41(DE3). *E. coli* BL21(DE3) was used as the standard T7-based expression host, whereas *E. coli* C41(DE3), a BL21(DE3)-derived strain with attenuated T7 RNA polymerase expression [35], was selected to evaluate whether moderating expression-associated stress could promote more sustained accumulation of the R5-His_6 fusion proteins.

For peptide release, a separate construct encoding TEVp with the autolysis-resistant S219V substitution and a C-terminal Arg_5 tail was introduced into *E. coli* BL21(DE3) (Fig. 1b). The protease was expressed as an N-terminal MBP fusion to support soluble production, following an established MBP-based expression strategy [36]. A TEVp cleavage site between MBP and the protease was incorporated to enable TEVp release via self-processing. The construct omitted the His_6 tag to permit selective capture of R5-His_6 after cleavage. The overall workflow involved mixing the fusion-containing lysate with the His-tag-free TEVp-containing lysate to enable direct peptide release, followed by Ni-NTA purification of R5-His_6 from the lysate mixture (Fig. 1c). This approach avoided the separate purification of the fusion protein and protease prior to cleavage.

Direct expression of R5-His_6

Direct expression of R5-His_6 was examined in *E. coli* BL21(DE3) and C41(DE3). No distinct band corresponding to R5-His_6 was detected in either the soluble or insoluble fraction, or in the Ni-NTA eluate, from either strain (Fig. 2). These results indicate that direct expression did not yield detectable accumulation or recovery of R5-His_6 under the conditions tested. Intracellular degradation may have

contributed to its low accumulation. We therefore examined whether fusion with MBP or NEXT could improve its accumulation and recovery.

Time-course analysis and quantification of soluble R5-His_6 fusion production

Expression of MBP-R5-His_6 and NEXT-R5-His_6 was evaluated in *E. coli* BL21(DE3) and C41(DE3). Samples collected at 0, 4, 8, and 12 h after IPTG induction were lysed and fractionated into soluble and insoluble fractions, which were analyzed by SDS-PAGE with Coomassie staining and in-gel His-tag staining (Fig. 3a, b). An additional time course at 12, 18, and 24 h was examined in separate C41(DE3) cultures (Fig. 3b). In contrast to direct expression of R5-His_6, both fusion constructs produced prominent bands corresponding to the fusion proteins in both strains. NEXT-R5-His_6 was detected exclusively in the soluble fraction, whereas MBP-R5-His_6 also accumulated in the insoluble fraction. These results indicate that NEXT was more effective than MBP in promoting soluble expression of the R5 fusion in both hosts, despite its much smaller size.

Time-dependent changes in relative fusion-protein accumulation were assessed by in-gel His-tag staining. Total accumulation was calculated by summing the soluble and insoluble target-band intensities, and soluble MBP-R5-His_6 accumulation was also analyzed separately. In BL21(DE3), both MBP and NEXT fusions showed similar overall temporal profiles of relative total accumulation, with the highest mean values at 8 h followed by a decline at 12 h (Fig. 3c). At 12 h, total accumulation decreased to 82.8% and 90.9% of the respective 8 h values for MBP-R5-His_6 and NEXT-R5-His_6. The decline was more pronounced in the soluble MBP fraction, which retained 68.7% of its 8 h value, compared with 90.9% for soluble NEXT-R5-His_6. Although the decrease in soluble MBP-R5-His_6 was accompanied by increased insoluble accumulation, the latter accounted for only part of the soluble loss. The substantial net loss of total fusion protein suggests that proteolysis also contributed to this decline during prolonged expression. In contrast, accumulation of both fusions in C41(DE3) increased up to 12 h and remained near the respective 12 h levels through 24 h without a noticeable decrease (Fig. 3d), indicating more sustained fusion-protein accumulation in C41(DE3) than in BL21(DE3). Based on these profiles, cultures were harvested at 8 h for BL21(DE3) and 12 h for C41(DE3) in subsequent experiments.

Next, soluble fusion-protein expression yields were quantified at the selected harvest times of 8 h for BL21(DE3) and 12 h for C41(DE3). Cleared lysates were analyzed by in-gel His-tag staining alongside known amounts of the corresponding purified fusion proteins, with separate calibration curves generated for each fusion in each independent experiment (Fig. 4a). The calibration curves shown exhibited a linear relationship between protein loading amount and fluorescence intensity ($R^2 > 0.99$). To compare production on a target-peptide basis, soluble fusion-protein yields were converted to R5-His_6-equivalent expression yields using the molecular mass ratio of R5-His_6 to each fusion protein. The NEXT fusion yielded 18.9 ± 2.3 and 23.9 ± 4.4 mg L^{-1} of R5-His_6 equivalents in BL21(DE3) and C41(DE3), respectively,

compared with 12.9 ± 1.6 and 17.4 ± 1.9 mg L^{-1} for the MBP fusion (Fig. 4b). The yields obtained with NEXT were therefore 46% and 37% higher than those obtained with MBP in BL21(DE3) and C41(DE3), respectively. Both fusions also showed higher yields in C41(DE3), with the highest yield obtained using NEXT in this host.

Lysate-based cleavage and recovery of R5-His_6

Both fusions gave higher soluble R5-His_6-equivalent expression yields in C41(DE3), so this strain was used to produce the fusion proteins for subsequent experiments. We then examined the release and recovery of R5-His_6 from the MBP and NEXT fusions. Fusion-containing lysates were mixed with a separate lysate containing His-tag-free TEVp at a 9:1 (v/v) ratio. Without a protease inhibitor cocktail, released R5-His_6 bands progressively weakened during incubation. The addition of the cocktail slowed this decline, indicating that proteolysis contributed to the loss of the released peptide (Fig. 5a). The cocktail was therefore included in subsequent cleavage experiments.

Cleavage was quantified from the decrease in uncleaved fusion-band intensity. NEXT-R5-His_6 showed faster cleavage than MBP-R5-His_6, with fitted k_{obs} values of 0.0552 and 0.0405 min^{-1}, respectively (Fig. 5b). Cleavage of the NEXT fusion was nearly complete by 60 min, whereas approximately 8% of the MBP fusion remained uncleaved. Thus, the NEXT fusion reached near-complete cleavage within a shorter incubation period.

R5-His_6 was recovered from both fusions by Ni-NTA purification following 60 min of cleavage, but residual uncleaved MBP-R5-His_6 co-purified with the released peptide (Fig. 5c). Although extending the MBP cleavage reaction to 90 min increased cleavage, no distinct R5-His_6 band was detected in the final eluate. This may reflect further degradation of the released peptide during prolonged incubation and subsequent purification. In contrast, the NEXT-derived eluate contained predominantly R5-His_6, with no distinct uncleaved fusion band (Fig. 5d). These results identified 60 min as a suitable cleavage duration for the NEXT fusion under our experimental conditions. The NEXT fusion yielded approximately 13 mg of R5-His_6 per liter of culture at the Ni-NTA elution stage, before dialysis (Fig. 5d). This yield corresponded to approximately half of the soluble R5-His_6-equivalent expression yield in C41(DE3) (Fig. 4b).

Characterization and silica-forming activity of purified R5-His_6

MALDI-TOF mass spectrometry revealed a dominant isotopic envelope with a first resolved peak at *m/z* 3078.096 (calculated monoisotopic *m/z* for $[M + H]^+$, 3077.600), supporting the recovery of full-length R5-His_6 (Fig. 6a). Its silica-forming activity was subsequently examined using acid-hydrolyzed TMOS as the silica precursor. Incubation with purified R5-His_6 produced a visible white precipitate within a few minutes, whereas no visible precipitate formed in the peptide-free control (Fig. 6b). SEM analysis

revealed predominantly spherical particles a few hundred nanometers in diameter, some of which formed aggregates (Fig. 6c). This morphology was consistent with that previously reported for R5-mediated silica formation [37]. EDX detected silicon and oxygen in the analyzed area (Fig. 6d). Together, these results demonstrate that the purified R5-His_6 retained silica-forming activity following recombinant production and recovery.

Discussion

The present study demonstrates that a small intrinsically disordered fusion tag can support efficient short-peptide production. Despite its considerably smaller size, the NEXT tag promoted soluble accumulation of the R5 fusion and yielded more R5-His_6 equivalents than MBP in both hosts. Its expanded conformations may contribute to this performance through steric exclusion, which has previously been proposed to explain the aggregation-suppressing effect of the NEXT tag [18]. The N-terminal fusion itself may also reduce the susceptibility of R5 to intracellular degradation. Together, the results indicate that a large globular fusion partner is not required to support R5 accumulation. By combining a lower biosynthetic demand for tag synthesis with effective soluble accumulation, the NEXT tag offers a promising alternative to larger fusion partners.

In addition to supporting soluble production, the NEXT fusion underwent faster TEVp-mediated cleavage than the MBP fusion. Although both constructs contained the same flexible linker and TEVp recognition sequence, differences in the surrounding structural environment may have contributed to their distinct cleavage rates. The conformational flexibility of NEXT and the absence of a bulky globular domain adjacent to the linker may favor TEVp access to the recognition sequence. This interpretation is consistent with previous evidence that improving cleavage-site accessibility can enhance TEVp-mediated tag removal [38].

The benefits of NEXT are expected to extend beyond the in-lysate workflow used here. Its ability to support high-level accumulation of soluble fusion protein and efficient peptide release should also benefit conventional peptide production processes, in which the fusion protein is first purified, followed by cleavage and a second purification step to isolate the released peptide. In this study, retaining His_6 on R5 enabled direct Ni-NTA purification after in-lysate cleavage. Near-complete cleavage was needed to minimize co-purification of uncleaved fusion, while faster cleavage shortened the incubation required to reach this endpoint, thereby reducing the exposure of released peptide to proteolytic degradation.

The NEXT-based process yielded approximately 13 mg L^{-1} of purified R5-His_6 in LB medium, comparable to the approximately 10 mg L^{-1} reported for MBP-based recombinant R5 production in Terrific Broth (TB) [12]. The purified yield was about half of the soluble peptide-equivalent expression yield. The decline in the released R5-His_6 band during incubation suggests that proteolysis contributed to this gap, and losses during affinity purification may also have played a role. Further improvements could focus on identifying inhibitor conditions that suppress residual proteolysis in the mixed lysates without compromising TEVp activity, thereby preserving the released peptide until cleavage is complete.

From a functional standpoint, the His_6 tag can mediate binding to silica [39], and histidine-rich peptides such as Kpt can promote silica formation [40]. These findings suggest that the retained His_6 tag may contribute to the interactions of R5-His_6 with silica or its precursors. R5-His_6 readily induced silica precipitation in this study, supporting the use of R5-His_6 for bioinspired silica synthesis.

Conclusions

The NEXT tag provides an effective alternative to large globular fusion tags for recombinant short-peptide production by enabling high-level soluble accumulation and efficient protease-mediated release. Its successful application to R5-His_6 production supports the use of small intrinsically disordered tags in peptide production workflows and motivates testing these tags with other short functional peptides.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Supplementary Information

Additional file 1: Table S1.

Nucleotide and amino acid sequences of the recombinant proteins used in this study.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (RS-2025-25442869 and RS-2024-00355356).

Author Contribution

BHJ conceived and supervised the study. GTL and BHJ designed experiments. GTL performed the experiments. GTL and BHJ analyzed the experimental results. GTL and BHJ wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Data Availability

All data supporting the findings of this study will be made available on request.

References

1. Wilding JPH, Batterham RL, Calanna S, Davies M, Van Gaal LF, Lingvay I, et al. Once-weekly semaglutide in adults with overweight or obesity. N Engl J Med. 2021;384:989–1002.
2. Xue B, Bashir Z, Guo Y, Yu W, Sun W, Li Y, et al. Strong, tough, rapid-recovery, and fatigue-resistant hydrogels made of picot peptide fibres. Nat Commun. 2023;14:2583.
3. Choi O, Kim BC, An JH, Min K, Kim YH, Um Y, et al. A biosensor based on the self-entrapment of glucose oxidase within biomimetic silica nanoparticles induced by a fusion enzyme. Enzyme Microb Technol. 2011;49:441–5.
4. Collins JM, Singh SK, White TA, Cesta DJ, Simpson CL, Tubb LJ, et al. Total wash elimination for solid phase peptide synthesis. Nat Commun. 2023;14:8168.
5. Mottola S, Del Bene A, Mazzarella V, Cutolo R, Boccino I, Merlino F, et al. Sustainable ultrasound-assisted solid-phase peptide synthesis (SUS-SPPS): Less waste, more efficiency. Ultrason Sonochem. 2025;114:107257.
6. Zhang Y, Wang Y, Lu J, Huang Z, Hua H, Li Y, et al. High-yield and cost-effective biosynthesis process for producing antimicrobial peptide AA139. Protein Expr Purif. 2024;219:106475.
7. Roca-Pinilla R, Lisowski L, Arís A, Garcia-Fruitós E. The future of recombinant host defense peptides. Microb Cell Fact. 2022;21:267.
8. Osiro KO, Duque HM, Sampaio de Oliveira KB, Melo NTM, Lima LF, Paes HC, et al. Cleaving the way for heterologous peptide production: An overview of cleavage strategies. Methods. 2025;234:36–44.
9. Lamer T, van Belkum MJ, Wijewardane A, Chiorean S, Martin-Visscher LA, Vederas JC. SPI sandwich: Combined SUMO-Peptide-Intein expression system and isolation procedure for improved stability and yield of peptides. Protein Sci. 2022;31:e4316.
10. Rodríguez V, Asenjo JA, Andrews BA. Design and implementation of a high yield production system for recombinant expression of peptides. Microb Cell Fact. 2014;13:65.

11. Li X, Zhang B, Hu Q, Chen C, Huang J, Liu L, et al. Refinement of the fusion tag PagP for effective formation of inclusion bodies in *Escherichia coli*. Microbiol Spectr. 2023;11:e0380322.
12. Wallace AK, Chanut N, Voigt CA. Silica nanostructures produced using diatom peptides with designed post-translational modifications. Adv Funct Mater. 2020;30:2000849.
13. Xie YG, Luan C, Zhang HW, Han FF, Feng J, Choi YJ, et al. Effects of thioredoxin: SUMO and intein on soluble fusion expression of an antimicrobial peptide OG2 in *Escherichia coli*. Protein Pept Lett. 2013;20:54–60.
14. Köppl C, Lingg N, Fischer A, Kröß C, Loibl J, Buchinger W, et al. Fusion tag design influences soluble recombinant protein production in *Escherichia coli*. Int J Mol Sci. 2022;23:7678.
15. Wang HZ, Chu ZZ, Chen CC, Cao AC, Tong X, Ouyang CB, et al. Recombinant passenger proteins can be conveniently purified by one-step affinity chromatography. PLoS ONE. 2015;10:e0143598.
16. Holehouse AS, Kragelund BB. The molecular basis for cellular function of intrinsically disordered protein regions. Nat Rev Mol Cell Biol. 2024;25:187–211.
17. Santner AA, Croy CH, Vasanwala FH, Uversky VN, Van YYJ, Dunker AK. Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. Biochemistry. 2012;51:7250–62.
18. Jo BH. An intrinsically disordered peptide tag that confers an unusual solubility to aggregation-prone proteins. Appl Environ Microbiol. 2022;88:e0009722.
19. Tang NC, Su JC, Shmidov Y, Kelly G, Deshpande S, Sirohi P, et al. Synthetic intrinsically disordered protein fusion tags that enhance protein solubility. Nat Commun. 2024;15:3727.
20. Kröger N, Deutzmann R, Sumper M. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. Science. 1999;286:1129–32.
21. Brandis D, Mollica G, Kurzbach D. R5 peptides constitute condensed phases with liquid-like properties in biomimetic silica capsules. J Phys Chem Lett. 2025;16:4326–35.
22. Luckarift HR, Spain JC, Naik RR, Stone MO. Enzyme immobilization in a biomimetic silica support. Nat Biotechnol. 2004;22:211–3.
23. Betancor L, Luckarift HR. Bioinspired enzyme encapsulation for biocatalysis. Trends Biotechnol. 2008;26:566–72.
24. Lim GT, Jo BH. Exploiting silica-binding and silica-forming proteins as versatile tools for one-step enzyme immobilization on siliceous materials. Int J Mol Sci. 2025;26:1304.
25. Reichinger D, Reithofer M, Hohagen M, Drinic M, Tobias J, Wiedermann U, et al. A biomimetic, silaffin R5-based antigen delivery platform. Pharmaceutics. 2023;15:121.
26. Ki MR, Kim JK, Kim SH, Nguyen TKM, Kim KH, Pack SP. Compartment-restricted and rate-controlled dual drug delivery system using a biosilica-enveloped ferritin cage. J Ind Eng Chem. 2020;81:367–74.
27. Lee AS, Kim KR, Kim JH, Jo BH, Song YH, Seo JH, et al. Bioinspired synthesis of micelle-templated ultrathin silica-layered mesoporous nanoparticles with enhanced mass transfer and stability for

biocatalysis. Chem Eng J. 2023;473:145108.
28. Kim KR, Lee AS, Heo HR, Park SY, Kim CS. Bioinspired synthesis of virus-like particle-templated thin silica-layered nanocages with enhanced biocompatibility and cellular uptake as drug delivery carriers. Colloids Surf B: Biointerfaces. 2025;247:114418.
29. Lim GT, Jo BH. Cation-assisted stabilization of carbonic anhydrase one-step *in situ* loaded in diatom-inspired silica nanospheres for potential applications in CO_2 capture and utilization. Sep Purif Technol. 2024;333:125926.
30. Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, et al. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng. 2001;14:993–1000.
31. Blommel PG, Fox BG. A combined approach to improving large-scale production of tobacco etch virus protease. Protein Expr Purif. 2007;55:53–68.
32. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The proteomics protocols handbook. Totowa, NJ: Humana; 2005. pp. 571–607.
33. Anthis NJ, Clore GM. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. Protein Sci. 2013;22:851–8.
34. Kapust RB, Tözsér J, Copeland TD, Waugh DS. The P1' specificity of tobacco etch virus protease. Biochem Biophys Res Commun. 2002;294:949–55.
35. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning *Escherichia coli* for membrane protein overexpression. Proc Natl Acad Sci U S A. 2008;105:14371–6.
36. Tropea JE, Cherry S, Waugh DS. Expression and purification of soluble His_6-tagged TEV protease. Methods Mol Biol. 2009;498:297–307.
37. Senior L, Crump MP, Williams C, Booth PJ, Mann S, Perriman AW, et al. Structure and function of the silicifying peptide R5. J Mater Chem B. 2015;3:2607–14.
38. Chen Y, Li Q, Yang J, Xie H. Promoting tag removal of a MBP-fused integral membrane protein by TEV protease. Appl Biochem Biotechnol. 2017;181:939–47.
39. Freitas AI, Domingues L, Aguiar TQ. Bare silica as an alternative matrix for affinity purification/immobilization of His-tagged proteins. Sep Purif Technol. 2022;286:120448.
40. Nguyen TKM, Ki MR, Son RG, Kim KH, Hong J, Pack SP. A dual-functional peptide, Kpt from *Ruegeria pomeroyi* DSS-3 for protein purification and silica precipitation. Biochem Eng J. 2020;163:107726.

Figures

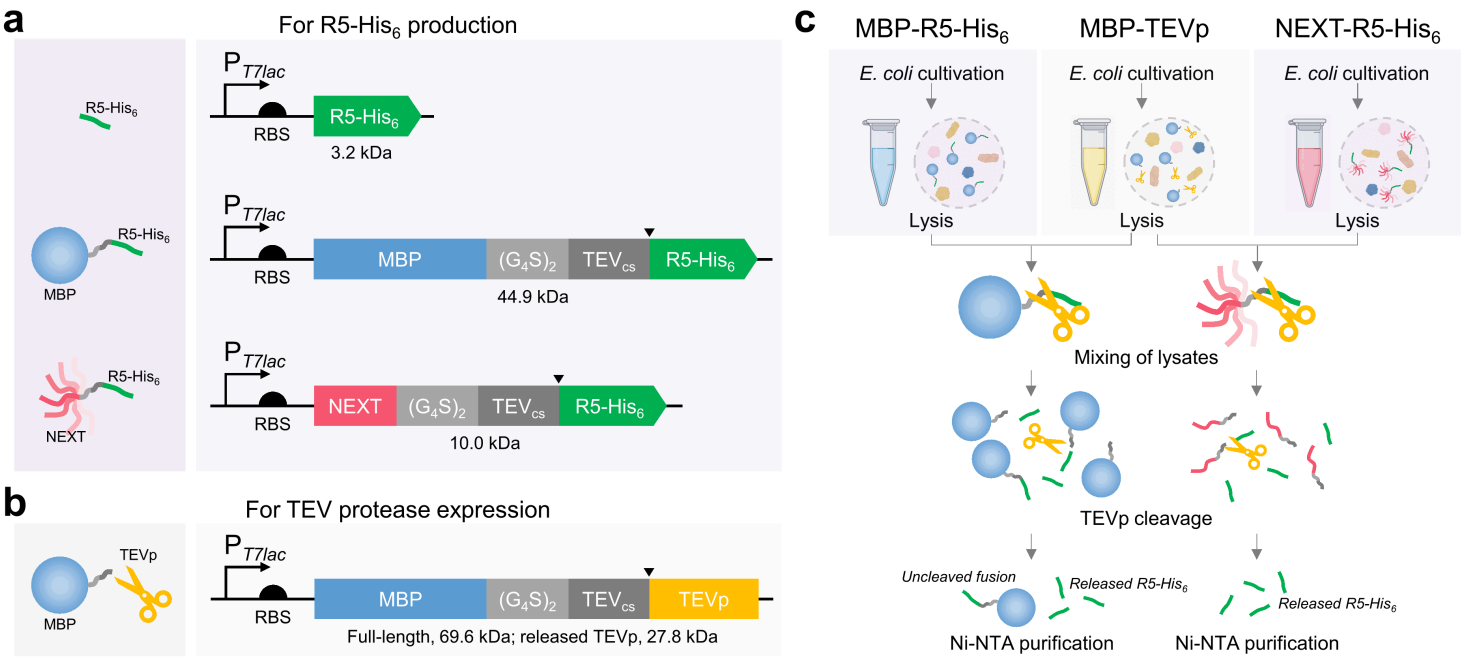


Figure 1

Construct design and lysate-based workflow for recombinant production of R5-His_6. (a, b) Schematic representations of the expressed proteins (left) and their corresponding genetic constructs (right) for R5-His_6 production **(a)** and TEVp expression **(b)**. Values indicate the calculated molecular masses of the full-length proteins; the mass of released TEVp is also shown in **(b)**. MBP is depicted as a compact sphere representing its globular structure, whereas NEXT is depicted as multiple red traces of varying intensity illustrating its intrinsically disordered nature and conformational dynamics. Arrowheads indicate TEVp cleavage sites in the encoded proteins. **(c)** Schematic overview of lysate-based R5-His_6 production and purification. Lysates containing either the MBP or NEXT fusion produced in *E. coli* C41(DE3) were mixed with a separate lysate containing His-tag-free TEVp produced in *E. coli* BL21(DE3). Following cleavage, released R5-His_6 was purified by Ni-NTA affinity chromatography. Cleavage time was limited to minimize degradation of the released peptide. Abbreviations: Ni-NTA, nickel-nitrilotriacetic acid; TEVp, tobacco etch virus protease; TEV_{cs}, TEVp cleavage site; RBS, ribosome-binding site; His_6, hexahistidine tag; $(G_4S)_2$, two repeats of the Gly-Gly-Gly-Gly-Ser linker.

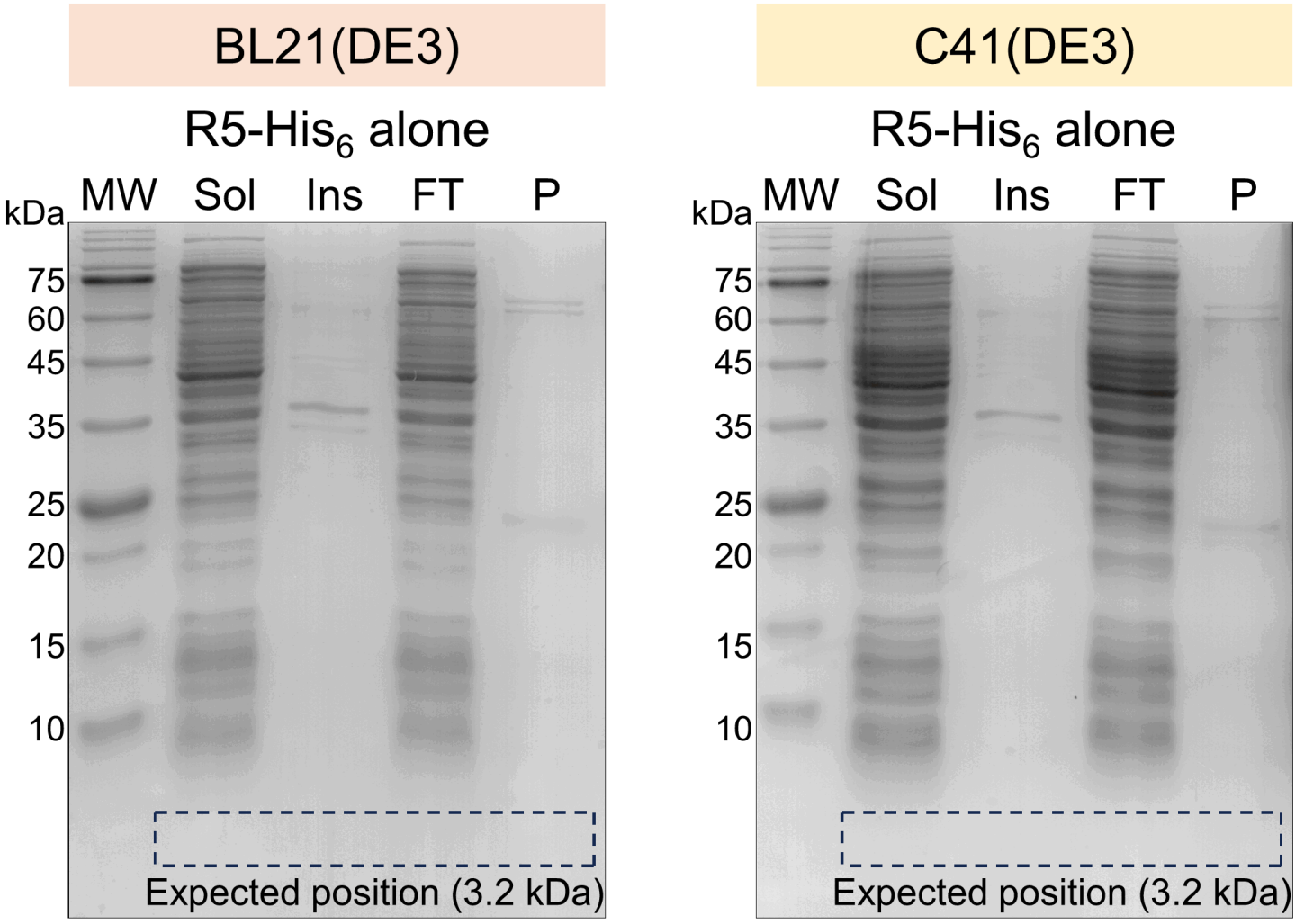


Figure 2

Direct expression of R5-His_6 in *E. coli* BL21(DE3) and C41(DE3). Coomassie-stained SDS-PAGE images of cellular fractions and Ni-NTA purification fractions from each strain. Dashed boxes indicate the expected migration region of R5-His_6 (3.2 kDa). No distinct band corresponding to R5-His_6 was detected in either strain under the conditions tested. MW, molecular mass marker; Sol, soluble fraction; Ins, insoluble fraction; FT, flow-through; P, purified fraction (Ni-NTA eluate).

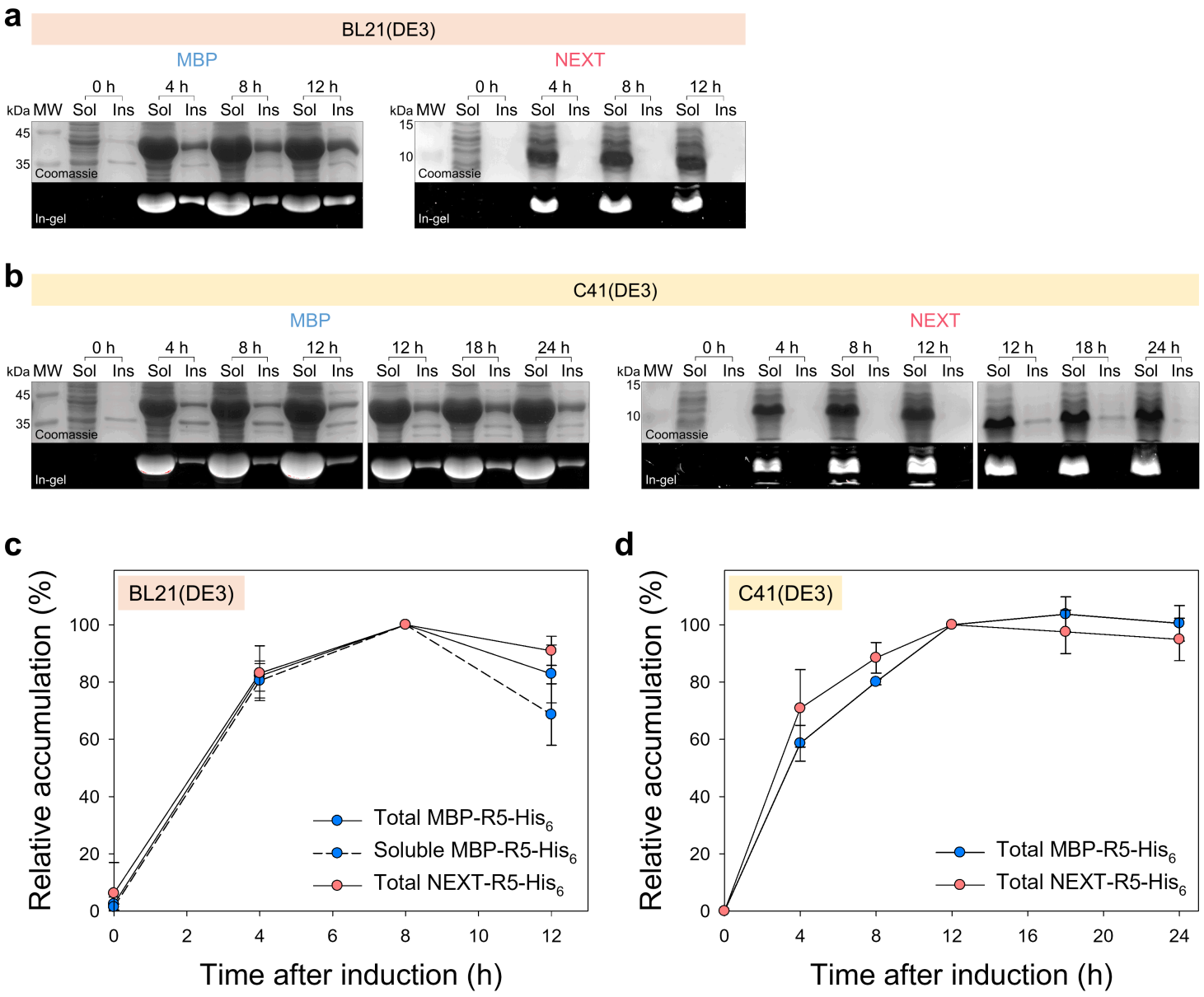


Figure 3

Time-dependent accumulation and solubility of R5-His$_6$ fusion proteins in *E. coli*. (a, b) SDS-PAGE images of soluble (Sol) and insoluble (Ins) fractions from BL21(DE3) **(a)** and C41(DE3) **(b)** expressing MBP-R5-His$_6$ or NEXT-R5-His$_6$, abbreviated as MBP and NEXT, respectively. Representative Coomassie-stained (upper) and in-gel His-tag-stained (lower) gels are shown. Samples were loaded on an equivalent culture-volume basis. C41(DE3) samples at 0–12 h and 12–24 h were obtained from separate cultures and analyzed on separate gels. MW, molecular mass marker. **(c, d)** Relative fusion-protein accumulation quantified by in-gel His-tag staining in BL21(DE3) **(c)** and C41(DE3) **(d)**. Total accumulation was calculated by summing the target-band intensities in the soluble and insoluble fractions. Soluble MBP-R5-His$_6$ accumulation is additionally shown in **(c)**. NEXT-R5-His$_6$ was detected exclusively in the soluble fraction. Each series was normalized separately within each experiment to the corresponding value at 8 h for BL21(DE3) or at 12 h for C41(DE3), which were set to 100%. The two C41(DE3) time courses were

normalized to their respective 12 h values. All quantitative data are presented as mean ± SD from three independent experiments.

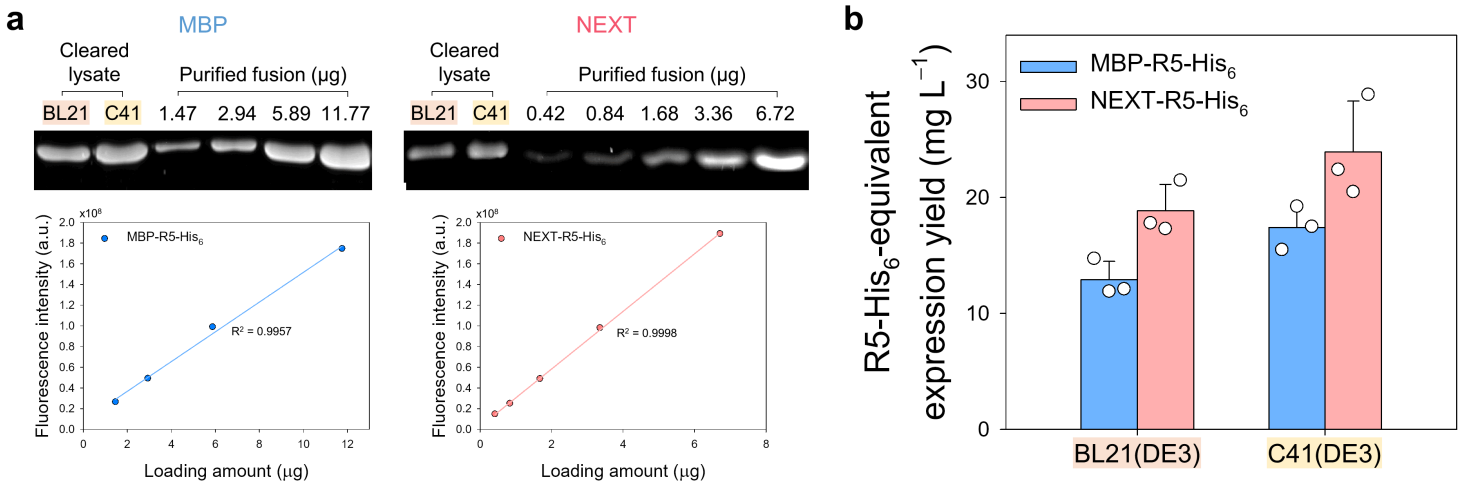


Figure 4

Quantification of soluble R5-His$_6$ fusion-protein expression yields in *E. coli*. (a) In-gel His-tag staining of cleared lysates from BL21(DE3) and C41(DE3) harvested at 8 and 12 h after induction, respectively, alongside known amounts of the corresponding purified fusion proteins and the resulting linear calibration curves. Each fusion protein was calibrated separately in each independent experiment; representative images and curves are shown. a.u., arbitrary units. **(b)** R5-His$_6$-equivalent expression yields per liter of culture, calculated by multiplying soluble fusion-protein yields by the ratio of the molecular mass of R5-His$_6$ to that of the corresponding fusion protein. These values reflect the R5-His$_6$ content of soluble fusions before cleavage and purification. Bars and error bars represent mean ± SD, and individual points represent three independent experiments.

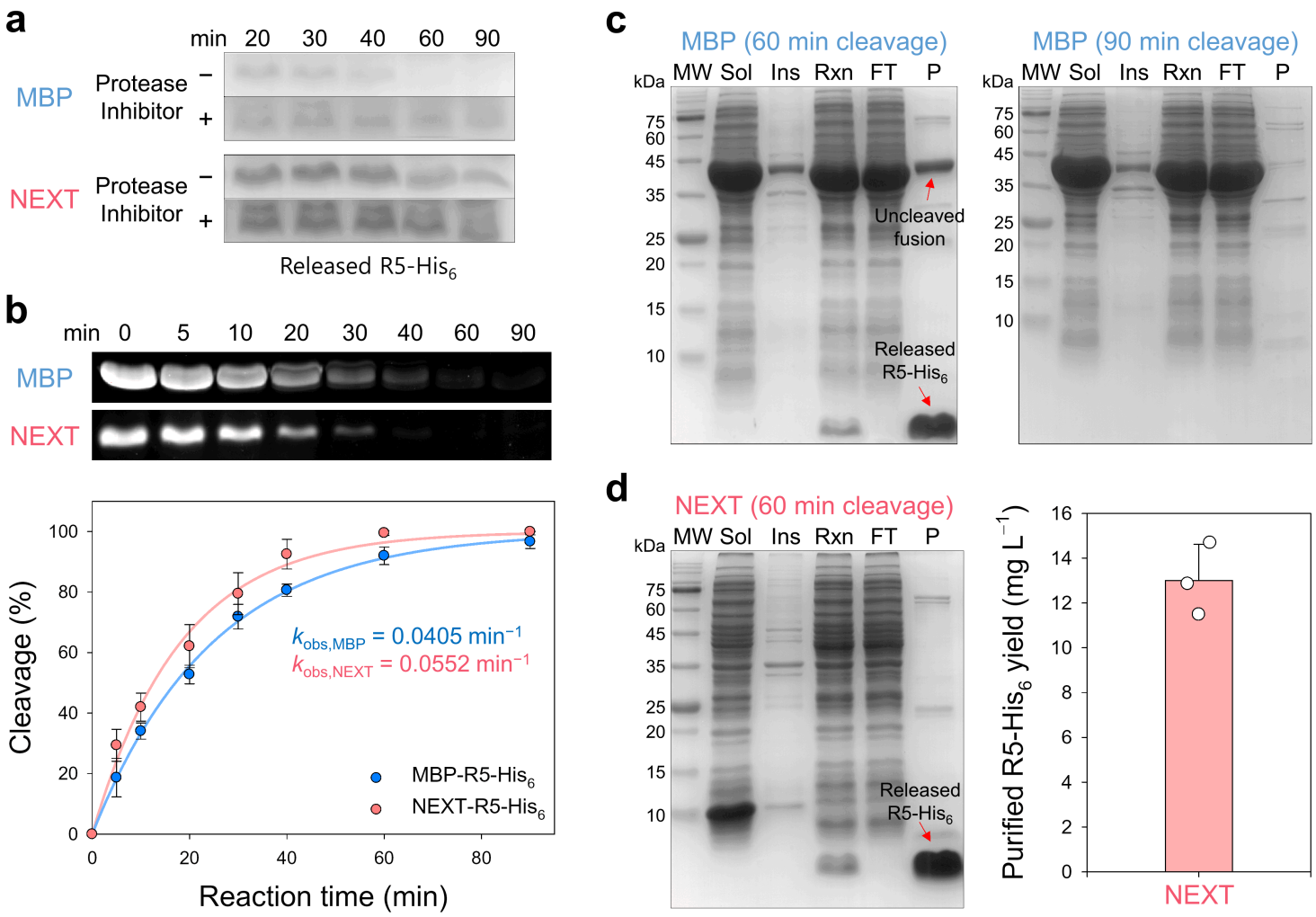


Figure 5

Lysate-based cleavage and purification of R5-His$_6$ from MBP and NEXT fusions. Lysates containing fusion proteins produced in *E. coli* C41(DE3) were mixed with a separate lysate containing His-tag-free TEV protease produced in *E. coli* BL21(DE3) at a 9:1 (v/v) ratio. **(a)** SDS-PAGE analysis of released R5-His$_6$ during cleavage with (+) or without (−) a protease inhibitor cocktail. **(b)** Time-dependent disappearance of uncleaved fusions visualized by in-gel His-tag staining and corresponding cleavage percentages calculated relative to the initial fusion-band intensity. Curves show single-exponential fits, with observed rate constants (k_{obs}) indicated. **(c)** Coomassie-stained SDS-PAGE analysis of Ni-NTA purification following 60 or 90 min of cleavage for the MBP fusion. **(d)** Corresponding purification analysis following 60 min of cleavage for the NEXT fusion (left) and purified R5-His$_6$ yield per liter of culture (right). Yields refer to the Ni-NTA eluate before dialysis. A protease inhibitor cocktail was included in **(b-d)**. Arrows indicate uncleaved fusion and released R5-His$_6$. Quantitative data are presented as mean ± SD from five independent experiments for MBP and four for NEXT in **(b)**, and three independent experiments in **(d)**. MW, molecular mass marker; Sol, soluble fraction; Ins, insoluble fraction; Rxn, reaction mixture after cleavage; FT, flow-through; P, purified fraction (Ni-NTA eluate).

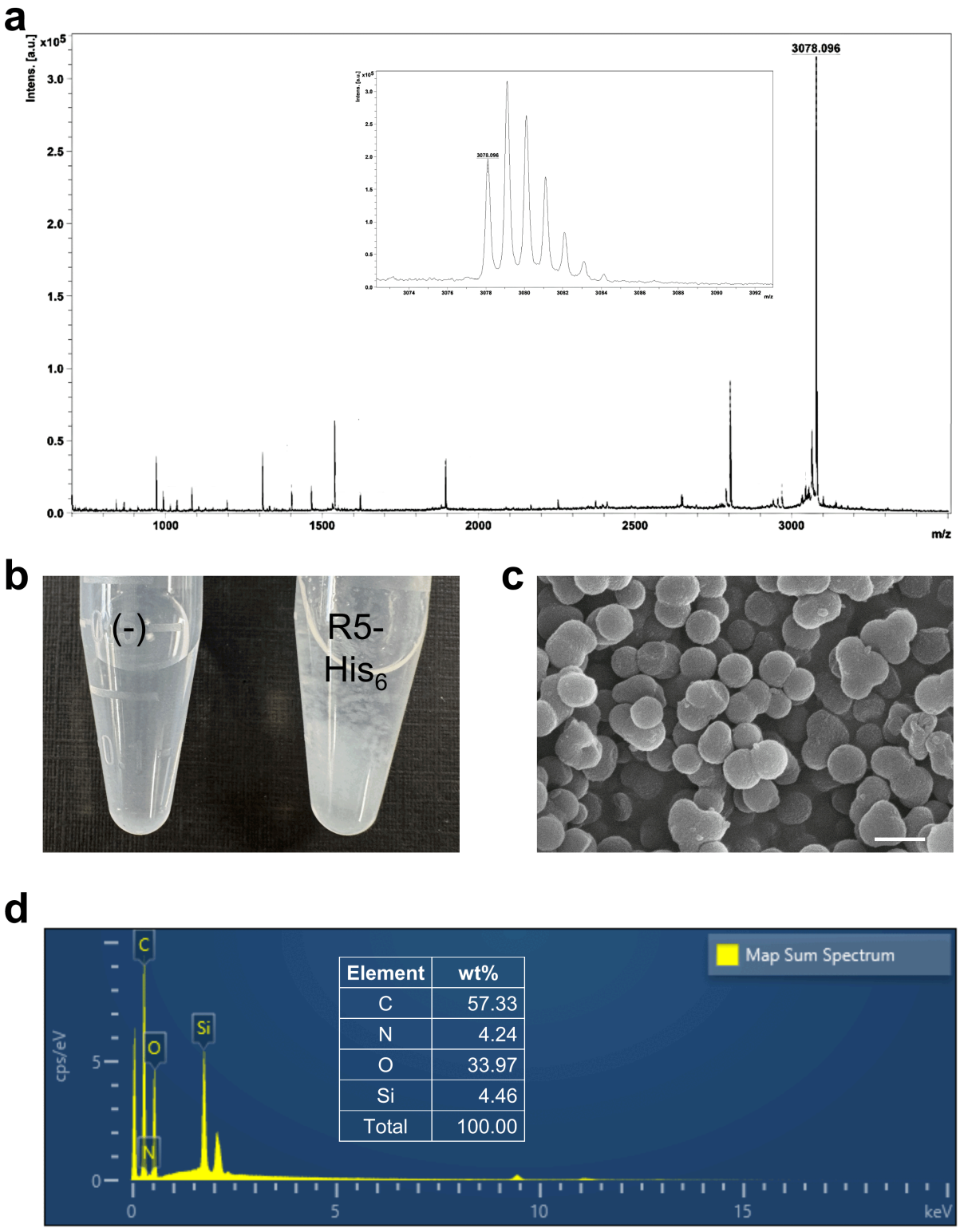


Figure 6

Mass spectrometric characterization and silica-forming activity of purified R5-His_6. (a) MALDI-TOF mass spectrum of purified R5-His_6. The inset shows an expanded view of the dominant isotopic envelope, with its first resolved peak at *m/z* 3078.096 (calculated monoisotopic *m/z* for $[M + H]^+$, 3077.600). **(b)** Photograph of reaction mixtures containing acid-hydrolyzed tetramethyl orthosilicate (TMOS) without peptide (−) or with purified R5-His_6. **(c)** Scanning electron microscopy (SEM) image of

the silica particles. Scale bar, 500 nm. **(d)** Energy-dispersive X-ray spectroscopy (EDX) spectrum of the silica particles. The inset lists the elemental weight percentages (wt%) determined by EDX for the analyzed area.

Supplementary Files

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

- Additionalfile1.pdf