MethodsX 8 (2021) 101512

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Protocol Article

The data of heterologous expression protocol for synthesis of ^{15}N, ^{13}C-labeled SEM1(68-107) peptide fragment of homo sapiens semenogelin 1

Aydar Bikmullin, Evelina Klochkova, Filipp Krasnovid, Dmitriy Blokhin*

Kazan Federal University, 18 Kremlevskaya, Kazan 420008, Russia

A B S T R A C T

The semenogelin 1 protein is secreted in the seminal vesicles. After ejaculation it is split into small peptide fragments using internal proteases. It was shown that the fragments SEM1(45-107), SEM1(49-107), SEM1(68-107) (SEM1(86-107) form amyloid fibrils, which increase the possibility of HIV infection. The article presents a protocol for the synthesis and purification of a ^{15}N, ^{13}C-labeled SEM1(68–107) peptide for further structural studies by high-resolution NMR spectroscopy. The work describes cloning, expression of fusion protein GB1-SEM1(68-107) in *E.coli*, its purification, removal of GB1 and purification of SEM1(68-107). The purity of SEM1(68-107) samples on each purification steps was evaluated by polyacrylamide gel electrophoresis under denaturing conditions (SDS-PAGE) and tricine-SDS-PAGE. The developed protocol allows to obtain SEM1(68-107) peptide for NMR studies (using 3D experiments), instead of costly solid-phase synthesis.

A R T I C L E I N F O

Method name: Heterologous expression and purification of ^{15}N,^{13}C-labeled peptides using a fusion with a partner protein followed by TEV cleavage

Keywords: Semenogelin 1, GB1 protein, SEM1(68-107) peptide, Protein expression, Protein purification, TEV cleavage

Article history: Received 4 May 2021; Accepted 6 September 2021; Available online 8 September 2021

* Corresponding author.
E-mail address: dblohin@kpfu.ru (D. Blokhin).

https://doi.org/10.1016/j.mex.2021.101512

Specifications table

Subject Area	Biochemistry, Genetics and Molecular Biology
More specific subject area	Heterologous expression of semenogelin 1 peptide fragments
Protocol name	Heterologous expression and purification of 15N,13C-labeled peptides using a fusion with a partner protein followed by TEV cleavage
Reagents/tools	Reagents: Phusion Green High Fidelity DNA Polymerase (Thermo Scientific, USA); GeneJET Gel Extraction Kit (Thermo Fisher, USA); T4 DNA Ligase Kit (Thermo Fisher, USA); GeneJET Plasmid MiniPrep kit; minimal medium (M9) (Na_2HPO_4 33.7 mM; KH_2PO_4 22.0 mM; NaCl 8.55 mM; $MgSO_4$ 1 mM; $CaCl_2$ 0.3 mM; D-biotin (1 mg/l); thiamin (1 mg/l); D-glucose 0.4 %; NH_4Cl 9.35 mM; EDTA 50 mg /L (134 uM); $FeCl_3{*}6H_2O$ 8.3 mg/L (31 uM); $ZnCl_2$ 0.84 mg/L (6.2 uM); $CuCl_2{*}2H_2O$ 0.13 mg/L (0.76 μM); $CoCl_2{*}2H_2O$ 0.1 mg/L (0.42 μM); H_3BO_3 0.1 mg/L (1.62 μM); $MnCl_2{*}4H_2O$ 16 ug/L (0.081 μM)) containing antibiotics (kanamycin, 50 μg/mL; chloramphenicol, 25 μg/mL); isopropy-β-D-1-thiogalactopyranoside (IPTG); buffer 1 (50 mM Tris-HCl, pH 8.8, 0.3 M NaCl); buffer 2 (50 mM Tris-HCl, pH 8.8; 1 M NaCl); buffer 3 (50 mM Tris-HCl, pH 8.8; 0.3 M NaCl; 150 mM imidazole); DTT (1 mM); PMSF (1 mM); EDTA (0,5 mM); SDS-PAGE; Tris-glycine buffer; tricine-SDS-PAGE Instruments: BioRad NGC T100 chromatographic system; a Beckman Coulter centrifuge set (Avanti JXN-26, Optima XPN-80); HD2070 ultrasonic homogenizer (Bandelin, Germany); Amicon spin-concentrators (Millipore, Ireland); a Milli-Q Millipore water filtration system; a sterile SafeFAST Elite laminar; cultivation shakers Inforce HT; PCR thermal cycler BioRad T100; NMR spectrometer 700 MHz (Bruker, AVANCE III-700)
Experimental design	The construction of expression vector carrying the GB1 gene with a six-histidine tag at the N-terminus (His-tag) and SEM1(68-107) gene was made. Further heterologous expression was made on minimal medium (M9) with isotope labeled glucose (^{13}C) and ammonium chloride (^{15}N) as a source of carbon and nitrogen, respectively. Fusion protein was digested with homemade recombinant his-tagged TEV protease. After the reaction mix was loaded on NiNTA resin for protein purification. As a result, we were performed pure ^{15}N,^{13}C-labeled SEM1(68-107) peptide.
Trial registration	N/A
Ethics	N/A.
Value of the Protocol	• The obtained protocols of heterologous expression make it possible to synthesize ^{15}N, ^{13}C-labeled SEM1(68-107) peptide, which is necessary for a complete structural NMR study. • The obtained protocols of protein expression (^{15}N, ^{13}C-labeled) for NMR experiments can reduce the cost of peptide synthesis in comparison with the solid-phase method.

Description of protocol

SEM1(86–107) peptide forms semen amyloid fibrils, which increase the infectious activity of the human immunodeficiency virus (HIV). NMR spectroscopy is frequently used for revealing the spatial structure of peptides in solution, which can be used to understand the process of fibril-formation by SEM1(86-107). To carry out structural NMR studies labeling of proteins or peptides with ^{15}N, ^{13}C is necessary. Since solid-phase synthesis of labeled peptides is expensive *in vivo* production of peptides with ^{15}N, ^{13}C-labeling is an attractive alternative.

Construction of expression vector

The SEM1(68-107) part (SEM40 fragment) of *Homo sapiens* semenogelin 1 gene (NCBI gen ID: 6406) was amplified from human genomic DNA. Sites of *Nco*I and *Xho*I restriction endonucleases as well as stop codon TGA were introduced by constructing following forward and reverse primers SEM40-f (5'-TTTTTTCCATGGGCACATATCATGTAGATGCCAATGATCATGACC-3') and SEM40-r (5'-TTTTTTCTCGAGTCAGAGCAGTTGTTGACTTCCACCTAGA-3') which were purchased by Evrogen (Moscow, Russia). PCR reaction was performed in 100 μL volume using Phusion Green High Fidelity DNA Polymerase (Thermo Scientific, USA) according the manufacturer recommendations. The thermal profile was as follows: 95 °C, 3 min; followed by 35 cycles of [95 °C, 15 s; 59 °C, 30 s; 72 °C, 30 s with a final extension at 72 °C for 5 min. PCR product was separated by a 2% agarose gel electrophoresis and purified with CleanUp Mini (Evrogen, Moscow, Russia) according the manufacturer recommendations. pET vector (pET-GB1) encoding N-terminal 6xHis-Tag followed by the IgG-binding

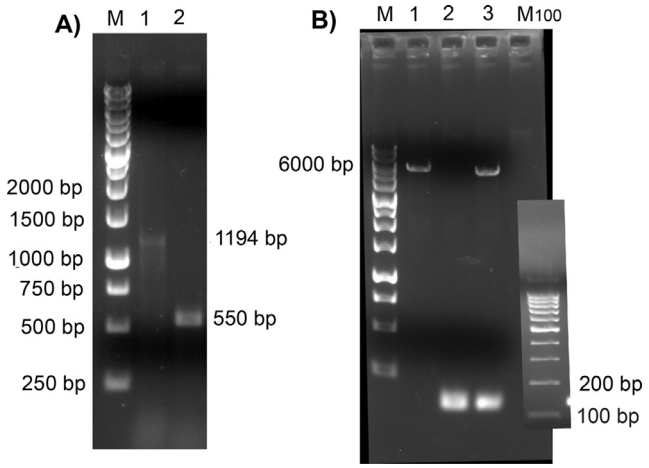

Fig. 1. Electrophoregrams of the nucleotide components for the expression vector cloning. (A) – Clone PCR with T7 primers. M – 1kb DNA Ladder (Evrogen); 1 – pET-GB1 without insert (but with GFP precloned); 2 – Clone with sem40 insert (pET-GB1::sem40). (B) – Restriction analysis with NcoI, XhoI. M – 1kb DNA ladder (Evrogen); 1 – pET-GB1 (5400 bp), separated from GFP fragment; 2 – sem40 (120bp); 3 – pET-GB1::sem40; M100 – 100bp DNA ladder (Evrogen, Russia).

Table 1
Features of the GB1-SEM40 and SEM40.

Protein	AA sequence	Number of AA	Mol. Weight, kDa
GB1-SEM40	MKHHHHHHPMKQYKLILNGKTLKGETTTEAVDAATAEKVFKQYAND NGVDGEWTYDDATKTFTVTEGSGSGSENLYFQGAMGTYHVDANDH DQSRKSQQYDLNALHKTTKSQRHLGGSQQLL	122	13,7
SEM40	GAMGTYHVDANDHDQSRKSQQYDLNALHKTTKSQRHLGGSQQLL	44	4,9

B1 domain of *Streptococcus* (GB1) with a TEV cleavage site expresses cloned genes under the control of a T7 promoter [1]. GB1 is a highly soluble and stable partner-protein for protein expression. Six-histidine tagged GB1 was linked with SEM40 protein fragment via the cleavage site of high efficient sequence-specific protease from Tobacco Etch virus [2]. pET-GB1 and the amplified SEM40 PCR fragment were digested with *Nco*I and *Xho*I, separated in agarose gel electrophoresis and isolated from gel. Following ligation with T4 DNA Ligase Kit by Thermo Fisher (USA) and transformation into *E. coli* strain DH5α resulted in obtaining of pET-GB1:SEM40 construct which was checked with restriction analyze (Fig. 1). The insertion, upstream and downstream regions of pET-GB1:SEM40 construct were sequenced with universal T7univ and T7 rev primers in Evrogen (Moscow, Russia).

Amino acid sequences of fusion protein GB1-SEM40, and SEM40 after cleavage and their molecular weights are shown in Table 1.

SEM40 expression and purification

Protein expression

Heterologous protein expression was performed in *E. coli* BL21 (DE3)pLysS on selective synthetic minimal medium (M9) containing antibiotics (kanamycin, 50 μg/mL; chloramphenicol, 25 μg/mL) with isotope labeled glucose (^{13}C) and ammonium chloride (^{15}N) (Cambridge Isotope Laboratories, UK) as a source of carbon and nitrogen [3]. You should use sterile 1M $MgSO_4$, 1M $CaCl_2$, 4M NH_4Cl, 0,5M EDTA (pH 8,0) stock solutions, 10 mg/ml D-biotin and 10 mg/ml thiamine stock solutions, 10X salt stock solution (Na_2HPO_4, KH_2PO_4, NaCl), 100X trace elements stock solution ($FeCl_3$*$6H_2O$, $ZnCl_2$, $CuCl_2$*$2H_2O$, $CoCl_2$*$2H_2O$, H_3BO_3, $MnCl_2$*$4H_2O$), 20% D-glucose (sterilized by 0,22 uM filtration) and

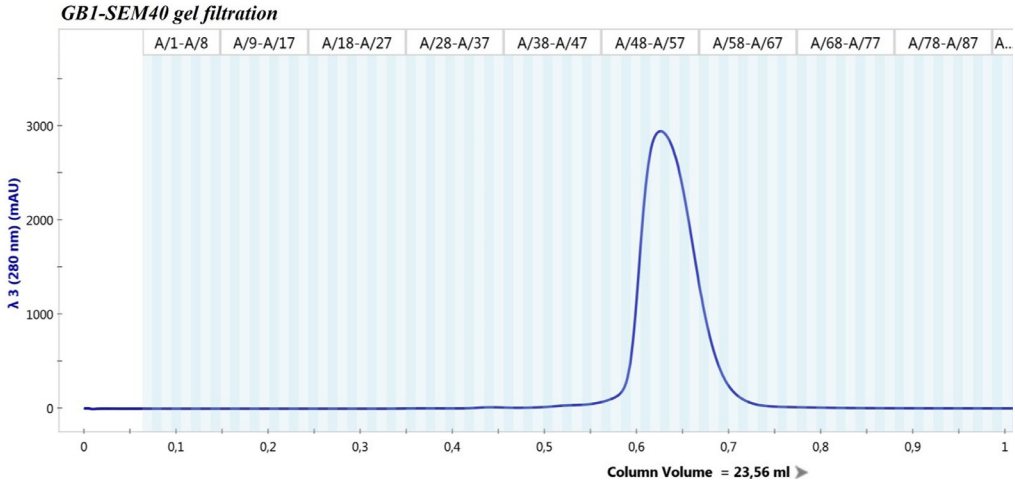

Fig. 2. Chromatogram of the GB1-SEM40 gel filtration.

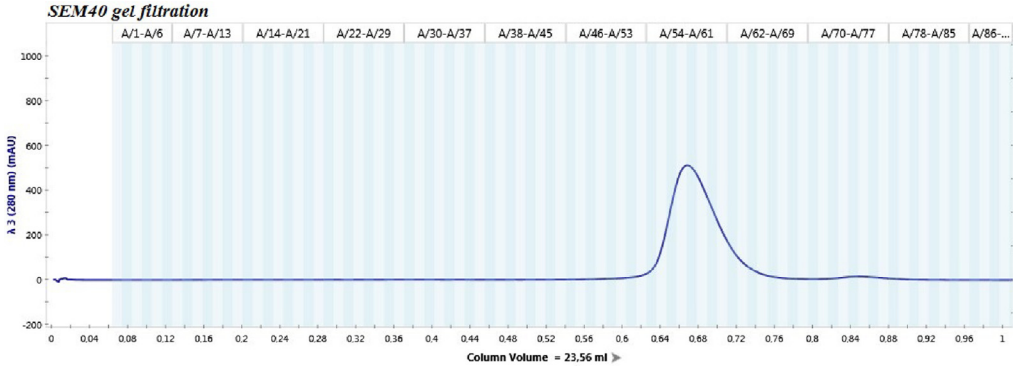

Fig. 3. Chromatogram of the SEM40 gel filtration.

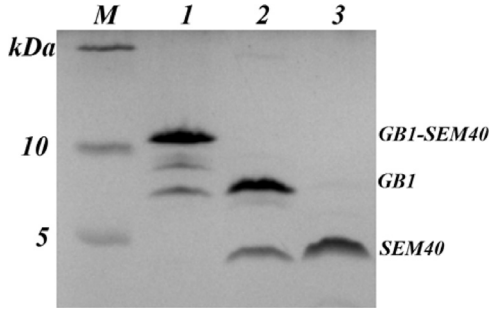

Fig. 4. Tricine SDS-PAGE of the intermidate components of the SEM40 purification. M – protein ladder; 1 – GB1-SEM40 before TEV cleavage; 2 – GB1 + SEM40 after TEV cleavage; 3 – pure SEM40.

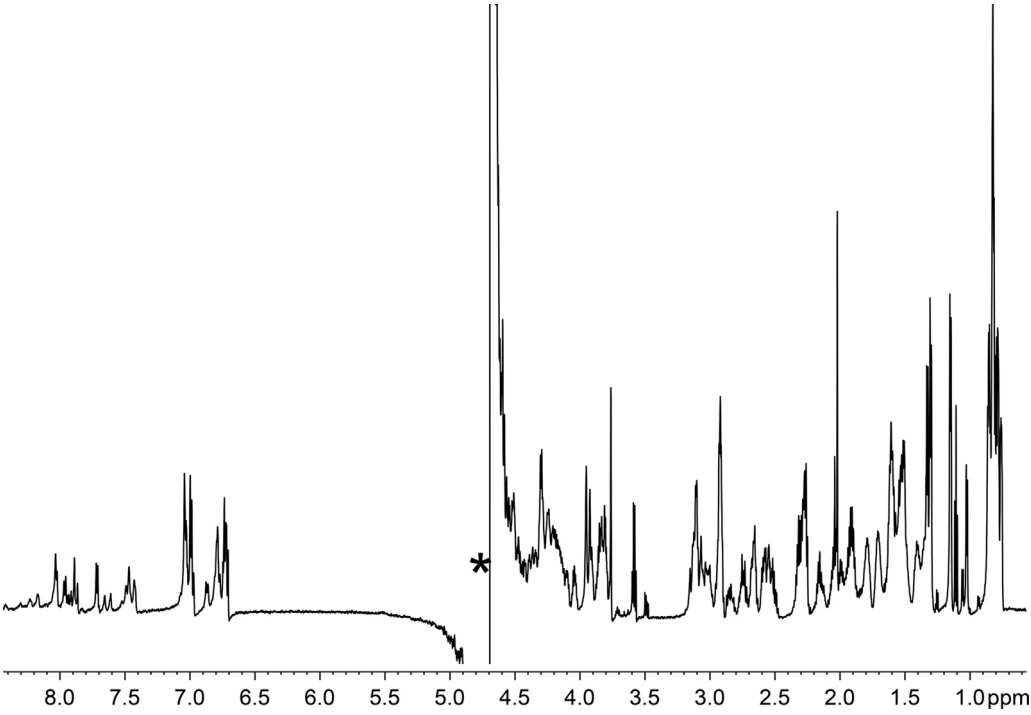

Fig. 5. ^{1}H NMR spectrum (^{1}H 700 MHz) of SEM40 peptide (1.6 mg/mL) in PBS 20 mM, pH 7.0 at T = 25 °C. (* is signal of water).

distilled water for M9 medium preparing [4]. Primarily the cells were grown in LB rich nutrient medium at 37 °C with shaking at the rate of 180 rpm. At optical density OD600 equal to 0.6 cells were harvested in sterile environment, washed and transferred into M9 medium. Then we grew the cells with same conditions for 40 min. The expression of the fusion protein GB1-SEM40 was induced by adding isopropy-β-D-1-thiogalactopyranoside (IPTG OD600=0.6–0.8) to a final concentration of 1 mM. Expression was carried out for 16h at the temperature of 37 °C and shaking rate of 180 rpm. Then cells were collected by centrifugation (5000 rpm, 10 min, 4 °C) using Avanti JXN-26, rotor JLA-9.1000 (Beckman, USA), the pellet was frozen in a liquid nitrogen and stored at –20 °C.

The cells were disrupted at 4 °C in buffer 1 (50 mM Tris-HCl, pH 8.8, 0.3 M NaCl) by endogenous T7 lysozyme (pLysS) and sonication with a HD2070 ultrasonic homogenizer (Bandelin, Germany) in the presence of protease inhibitor cocktail (Mini Protease Inhibitor Cocktail (Roche, Switzerland)), including metalloproteinases. Cell debris was pelleted by centrifugation at 100,000g for 1 h at 4°C using Optima XPN (45Ti rotor) centrifuge (Beckman Coulter, USA). Supernatant was cleared by 5 μm membrane filters and loaded on NiNTA-agarose resin equilibrated in buffer 1.

GB1-SEM40 purification

IMAC was carried out in buffer 1 with intermediate step of salt wash using buffer 2 (50 mM Tris-HCl, pH 8.8; 1 M NaCl). Protein was eluted with buffer 3 (50 mM Tris-HCl, pH 8.8; 0.3 M NaCl; 150 mM imidazole). Fractions after elution was pooled and concentrated by Amicon Ultra-4 (10 K) spin-concentrators (Millipore, Ireland) to the concentration allowed for loading on gel filtration column.

Gel filtration was performed using an NGC Discover chromatographic system and Enrich SEC75 column (BioRad, USA) in buffer 1 with 1 mL/min flowrate. The chromatogram had one peak in the ~0,65 CV region (Fig. 2). Peak fractions were pooled and concentrated to 1, 2 mL.

TEV-cleavage

Fusion protein was digested with homemade recombinant his-tagged TEV protease [5] at ratio TEV:GB1-SEM40 equal 1:100 (w/ w). Overnight reaction was made in presence of DTT (1 mM), PMSF (1 mM) and EDTA (0,5 mM) at 4 °C [2].

SEM40 purification

After TEV cleavage the reaction mix was loaded on NiNTA resin (in buffer 1) to trap His-tagged GB1 and TEV protease. We obtained pure SEM40 protein in a flow through fraction and concentrated the sample to 1, 2 mM by 3 kDa Amicon Ultra-0.5 (3K) spin-concentrators (Millipore, Ireland) for NMR experiment. We performed the final gel filtration of SEM40 (Fig. 3) to check purity and stability of the sample.

The purity of the samples on each purification steps were evaluated by polyacrylamide gel electrophoresis under denaturing conditions (SDS-PAGE) in pH 8.3 Tris-glycine buffer [6] and tricine-SDS-PAGE [7]. The electropherogram is shown at Fig. 4.

The ^{1}H NMR spectrum of SEM40 with water suppression is shown at Fig. 5. It corresponds to the type of protein spectrum, there are no narrow signals characteristic of low molecular weight compounds. These data correlate with electropherogram (Fig. 4).

We uploaded in Mendeley database following files [8]: nucleotide sequence of insertion GB1-SEM40 with his-tag (seq_gb1-sem40); amino acid sequences of GB1-SEM40 (file aa_seq_gb1-sem40) and SEM40 after TEV cleavage (file aa_seq_sem40_after_TEV); pET-GB1::sem40 construction, annotated sequence (file pet-GB1-sem40); sequence of SEM40 amplificated from the Human genome, with restriction sites (NcoI, XhoI) and stabilizing TTTTTT tails (file sem40); fragment amplificated with T7-primers from pET-GB1::sem40, It contains T7 promotor, GB1-SEM40 coding sequence, and T7 terminator (file T7sem40).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the Russian Science Foundation (D.S. Blokhin, project no. 20-73-10034).

References

[1] J. Bogomolovas, B. Simon, M. Sattler, G. Stier, Screening of fusion partners for high yield expression and purification of bioactive viscotoxins, Protein Expr. Purif. 64 (1) (2009) 16–23, doi:10.1016/j.pep.2008.10.003.
[2] S. Raran-Kurussi, S. Cherry, D. Zhang, D.S. Waugh, Removal of affinity tags with TEV protease, Methods Mol. Biol. 1586 (2017) 221–230, doi:10.1007/978-1-4939-6887-9_14.
[3] K.S. Usachev, A.A. Golubev, S.Z. Validov, V.V. Klochkov, A.V. Aganov, I.Sh. Khusainov, M.M. Yusupov, Backbone and side chain NMR assignments for the ribosome elongation factor P (EF-P) from staphylococcus aureus, Biomol. NMR Assign. 12 (2) (2018) 351–355, doi:10.1007/s12104-018-9838-z.
[4] C.R. Harwood, S.M. Cutting, Chemically defined growth media and supplements, in: C.R. Harwood, S.M. Cutting (Eds.), Molecular Biological Methods for Bacillus, Wiley, Chichester, United Kingdom, 1990, p. 548.
[5] J.E. Tropea, S. Cherry, D.S. Waugh, Expression and purification of soluble His(6)-tagged TEV protease, High Throughput Protein Expr. Purif. MIMB 498 (2009) 297–307, doi:10.1007/978-1-59745-196-3_19.
[6] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 68–685, doi:10.1038/227680a0.
[7] H. Schagger, Tricine–SDS-PAGE, Nat. Protoc. 1 (2006) 16–22, doi:10.1038/nprot.2006.4.
[8] D. Blokhin, Sequences GB1-SEM1(68–107) and SEM1(68-107), Mendeley Data (2021) v2, doi:10.17632/mh36zhbv3y.2.