Materials Science & Engineering C 108 (2020) 110426

Contents lists available at ScienceDirect

Materials Science & Engineering C

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

A vector-enzymatic DNA fragment amplification-expression technology for construction of artificial, concatemeric DNA, RNA and proteins for novel biomaterials, biomedical and industrial applications

Piotr M. Skowron[a,b,*,1], Natalia Krawczun[a,b], Joanna Zebrowska[a,b], Daria Krefft[a,b], Olga Zołnierkiewicz[a], Marta Bielawa[b], Joanna Jezewska-Frackowiak[a,b], Lukasz Janus[a,b], Malgorzata Witkowska[a], Malgorzata Palczewska[a], Adriana Schumacher[c,d,2], Anna Wardowska[d,e], Milena Deptula[e], Artur Czupryn[f], Piotr Mucha[g], Arkadiusz Piotrowski[h,i], Pawel Sachadyn[j], Sylwia Rodziewicz-Motowidlo[k], Michal Pikula[d,e], Agnieszka Zylicz-Stachula[a,b,1]

[a] Department of Molecular Biotechnology, Faculty of Chemistry, University of Gdansk, Gdansk 80-308, Poland
[b] BioVentures Institute Ltd., Poznan 60-141, Poland
[c] Department of Embryology, Faculty of Medicine, Medical University of Gdansk, Gdansk 80-211, Poland
[d] Department of Clinical Immunology and Transplantology, Faculty of Medicine, Medical University of Gdansk, 80-210, Poland
[e] Laboratory of Tissue Engineering and Regenerative Medicine, Department of Embryology, Faculty of Medicine, Medical University of Gdansk, Gdansk 80-211, Poland
[f] Nencki Institute of Experimental Biology, Warsaw 02-093, Poland
[g] Department of Molecular Biochemistry, Faculty of Chemistry, University of Gdansk, 80-308, Poland
[h] Department of Biology and Pharmaceutical Botany, Faculty of Pharmacy, Medical University of Gdansk, 80-416, Poland
[i] International Research Agenda - 3P Medicine Lab, Medical University of Gdansk, 80-416, Poland
[j] Department of Molecular Biotechnology and Microbiology, Faculty of Chemistry, Gdansk University of Technology, Gdansk 80-233, Poland
[k] Department of Biomedical Chemistry, Faculty of Chemistry, University of Gdansk, Gdansk 80-308, Poland

ABSTRACT

A DNA fragment amplification/expression technology for the production of new generation biomaterials for scientific, industrial and biomedical applications is described. The technology enables the formation of artificial Open Reading Frames (ORFs) encoding concatemeric RNAs and proteins. It recruits the Type IIS *Sap*I restriction endonuclease (REase) for an assembling of DNA fragments in an ordered head-to-tail-orientation. The technology employs a vector-enzymatic system, dedicated to the expression of newly formed, concatemeric ORFs from strong promoters. Four vector series were constructed to suit specialised needs. As a proof of concept, a model amplification of a 7-amino acid (aa) epitope from the S protein of HBV virus was performed, resulting in 500 copies of the epitope-coding DNA segment, consecutively linked and expressed in *Escherichia coli* (*E. coli*). Furthermore, a peptide with potential pro-regenerative properties (derived from an angiopoietin-related growth factor) was designed. Its aa sequence was back-translated, codon usage optimized and synthesized as a continuous ORF 10-mer. The 10-mer was cloned into the amplification vector, enabling the N-terminal fusion and multiplication of the encoded protein with MalE signal sequence. The obtained genes were expressed, and the proteins were purified. Conclusively, we show that the proteins are neither cytotoxic nor immunogenic and they have a very low allergic potential.

1. Introduction

Recently peptides and polypeptides have been increasingly employed in different fields of applied science such as medicine or material engineering. Intense investigation on peptide-based biomaterials as biologically active tools has led to the development of the wide range of compounds with enhanced properties, including peptidomimetics. Both, native and engineered derivatives of peptides have been successfully applied in regenerative medicine, biosensor construction or drug-delivery systems, among others [1–4].

In an attempt to conduct further investigation on these biomaterials and to ensure sufficient supply of the desired peptides, a reliable method is needed to obtain a high yield of gene expression product. With general advances in molecular biology, several approaches have been utilized to acquire a large quantity of specific DNA fragment or gene expression product. Some of these methods rely on the

* Corresponding author at: Department of Molecular Biotechnology, Faculty of Chemistry, University of Gdansk, Gdansk 80-308, Poland.
E-mail address: piotr.skowron@ug.edu.pl (P.M. Skowron).
[1] These authors contributed equally to this work.
[2] Present address: Department of Pharmacology, Faculty of Medicine, Medical University of Gdansk, Poland.

https://doi.org/10.1016/j.msec.2019.110426
Received 26 April 2019; Received in revised form 17 October 2019; Accepted 11 November 2019
Available online 13 November 2019
0928-4931/

construction of the plasmids which are capable of carrying multiple joined genes. Strategies enabling head-to-tail arrangement of the monomers are preferred over the head-to-head or random arrangement. The head-to tail arrangement of the cloned multimers ensures stability of the recombinant DNA plasmid and a proper operon expression. Thus, it allows for the construction of artificial, concatemeric DNAs, RNAs and proteins, containing multiple copies of ordered segments with pre-programmed function.

Several methods require specific cleavage sites to be included in a DNA concatemer sequence, which imposes limitations on a DNA fragment to be used as a monomer [5–8]. Kim and Szybalski demonstrated that multimers of the DNA fragments can be obtained by the means of Type IIS REase cleavage followed by DNA ligation [6]. However, due to the use of 4-nt overhang REase *Bsp*MI and partial DNA digestion, it was impossible to maintain functional ORFs. Another approach was presented by Wang et al. and Lennick et al. [9,10]. A DNA concatemer was created by ligation of the sticky-ended DNA fragments with addition of the synthetic DNA linkers (oligodeoxynucleotides (oligos)). These methods, however, are not versatile. The monomeric units require tedious *de novo* creation of the constructs. Moreover, additional aa had to be introduced to a target aa sequence and only 8–12 copies within a concatemer were obtained.

A development of the seamless cloning-based methods eliminated the additional restriction sites requirement [11–14]. Meyer and Chilkoti [15] introduced the recursive directional ligation method, which employed the seamless cloning approach. Chu et al. [16] demonstrated the amplification of DNA multimers by non-template PCR. Nonetheless, synthesis of long, protein encoding DNA multimers with these methods still pose a challenge.

This paper presents a development of the straightforward and highly efficient technology for obtaining concatemeric proteins of the desired length using four series of the specific amplification-expression DNA vectors, equipped with the universal DNA amplification module. The module contains two convergent DNA recognition sequences of the Type IIS REase *Sap*I, separated with a *Sma*I site for insertion of any DNA fragment. The presented method has numerous potential applications, especially in the pharmaceutical industry and tissue engineering, including vaccines and drug delivery systems production as well as mass-production of the peptide-derived biomaterials. The technology described has been protected by Polish and PCT patent applications, filed in 2014–15, while in 2018 Polish patent has been issued [17,18] and the global patent applications are currently being processed.

2. Material and methods

2.1. Bacterial strains, cell lines, plasmid, media and reagents

All REases were from New England Biolabs (Ipswich, MA, USA). Marathon DNA polymerase was from A&A Biotechnology (Gdynia, Poland). T4 DNA Ligase, T4 DNA Polymerase, Shrimp Alkaline Phosphatase (SAP), 100 bp and 1 kb DNA and protein ladders were from Thermo Fisher Scientific Baltics UAB (Vilnus, Lithuania) and GE Healthcare (Upsalla, Sweden). Ni Sepharose 6 Fast Flow resin and chromatographic columns: Resource Q, Resource S, HisTrap FF, HiLoad 16/600 Superdex 200 pg were form GE Healthcare. *E. coli* TOP10 [F-*mcr*A Δ (*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G] from Life Technologies (Gaithersburg, MD, USA) was used for plasmid DNA purification. *E. coli* strains (Invitrogen™ BL21(DE3), Invitrogen™ BL21 Star (DE3), Agilent Technologies BL21-Gold (DE3), 1 ml Pierce High-Capacity Endotoxin Removal spin columns and horseradish peroxidase substrate were from Thermo Fisher Scientific (Waltham, MA, USA). The HaCaT cell line was from DKFZ (Heidelberg, Germany). Primaria flasks (cat. no 353810) were from Becton Dickinson (Franklin Lakes, NJ, USA). The 46BR.1N line was from ECACC. Horseradish peroxidase, Sigma FAST™ Protease Inhibitor, media (DMEM, RPMI), serum (FCS) for human cell culture, anti-polyHistidine antibodies, DMEM HG (Dulbecco's Modified Eagle Medium High Glucose), cat. no D6429, 10% Fetal Bovine Serum, cat. no F9665, Cell Proliferation Kit II, cat no 11465015001, and Histopaque® were from Sigma-Aldrich (St. Louis, MO, USA). Bacterial media components were from BTL (Lodz, Poland). Agarose was from Bioshop (Burlington, Canada). DNA purification kits were from Blirt (Gdansk, Poland) and Thermo Fisher Scientific Baltics UAB. High resolution agarose (cat. AG42-005) was from Blirt. DNA sequencing was performed at Genomed (Warsaw, Poland) or Eurofins Genomics (Ebersberg, Germany). Proteus NoEndo M spin columns were from Protein ARK (Sheffield, UK). Trans-Blot Turbo Mini PVDF Transfer Pack was from Bio-Rad (Hercules, CA, USA). The oligos chemical synthesis was performed at Genomed (Warsaw, Poland) or Sigma Aldrich. The gene synthesis was performed at Bio Basic (Markham ON, Canada) and Genescript (Piscataway NJ, USA). Other reagents were from Avantor Performance Materials Poland S.A. (Gliwice, Poland), AppliChem Inc. (St. Louis Missouri, MO, USA) or Fluka Chemie GmbH (Buchs, Switzerland). Cell cytotoxicity test based on lactate dehydrogenase (LDH), cat. no MK401 was from Takara (Japan). Silicone culture inserts, cat no 81176 were from Ibidi GmbH (Planegg, Germany). Reagent kit Flow CAST® Highsens was from Bühlmann Laboratories (Schönenbuch, Switzerland). The genetic maps of the DNA vectors and recombinant constructs were performed using SnapGene software version 4.1 (http://www.snapgene.com). STATISTICA software was from StatSoft (Krakow, Poland). Prism 5 software was from GraphPad Software (San Diego, CA, USA).

2.2. 1st series of pR promoter/p15A ori DNA amplification-expression pAMP vectors

The DNA amplification-expression vectors were designed on the basis *p15A* origin vector pACYC184 and pRZ4737 [19] (Data in Brief: Supplemental data) (GenBank: MK606505, MK606506, MK606507, MK606519, MK606520, MK651654). The pAMP DNA vectors contain: (i) a strong, temperature-regulated bacteriophage lambda *pR* transcription promoter, (ii) a bacteriophage lambda *cI857ts* repressor gene for host-independence of the vector, (iii) a DNA fragment amplification module, with two convergent *Sap*I sites for in-frame, head-to-tail amplification of DNA fragments, resulting in assembly of an artificial, continuous, multimeric ORF and (iv) antibiotic resistance gene. DNA sequences of all PCR primers used for vectors construction, are provided in Data in Brief: Supplemental data.

Firstly, a basic *p15ori*-PR DNA vector was created. For this purpose, two DNA fragments were PCR amplified: the first PCR fragment was amplified from pRZ4737 DNA vector using PR-Sac-MCS and PR-Cla primers and contained *pR* promoter and gene encoding *cI857ts* repressor. A multiple cloning site (MCS) was designed and introduced using PR-Sac-MCS primer. The MCS consisted of *Bsp*HI, *Nco*I, *Nde*I and *Sac*I recognition sequences. The second PCR fragment was amplified from pACYC184 DNA vector using FpACYC-Sac and RpACYC-Cla and contained an antibiotic resistance gene against chloramphenicol and *p15A* origin. The obtained PCR fragments were cleaved with *Sac*I and *Cla*I REases. The pRZ4737-derived DNA fragment was dephosphorylated. Thereafter, both DNA fragments were purified and ligated. Secondly, the 5′-phosphorylated, synthetic oligos, encoding the designed amplifying modules (Data in Brief, Fig. 1) were cloned into dephosphorylated *p15ori*-PR DNA vector, previously cleaved with *Nco*I and SacI. All the amplifying modules possessed the appropriate sticky ends, which facilitates DNA cloning. The amplifying modules (Data in Brief, Fig. 1) may be easily introduced and modified, as desired, to various DNA vectors, containing alternative origins of replication, antibiotic resistance genes, transcriptional promoters and translation initiation signals.

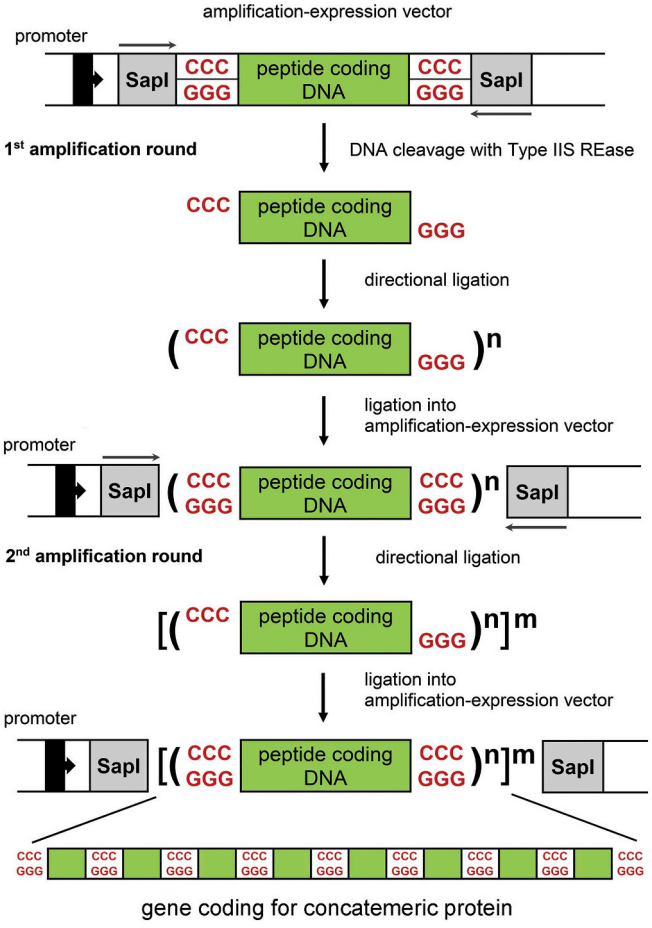

Fig. 1. Diagram showing the principle of the developed DNA fragment amplification-expression method.

2.3. 2nd series of T7-lac promoter/colE1 ori DNA amplification-expression pET21AMP vectors

The widely used pET-21d(+) expression vector was utilized as the template for plasmid modification [20]. First, the existing *Sap*I site (position 3046) was removed by site-specific mutagenesis. The synthetic DNA modules with cohesive ends for *Nco*I and *Sac*I (Data in Brief, Fig. 1a, d) were cloned into the pET-21d(+)del*Sap*I vector, linearized with these enzymes. The obtained pET21AMP-HisA vector possesses the amplifying module which was inserted into the original vector's MCS (Fig. 1) (Data in Brief: Supplemental data) (GenBank MK606521).

2.4. 3rd series of T7-lac promoter/colE1 ori/ubiquitin fusion DNA amplification-expression pET28AMP_SapI-Ubq vectors

The pET-28a(+) expression vector was used as the template for plasmid modification. The native *Sap*I recognition sequence (position 3108) was eliminated from pET-28a(+) by site-specific DNA mutagenesis. The obtained pET-28a(+)del*Sap*I vector was digested with *Nco*I and *Xho*I, dephosphorylated with SAP and gel purified. A DNA cassette with the modified His6_c-Myc_WYY_ubiquitin_*Sap*I-*Sma*I-*Sap*I amplification module enabling the ubiquitin gene fusion was designed and synthetized. The 5′ and 3′ ends of the synthetic DNA cassette contained the *Nco*I and *Xho*I restriction sites. DNA cleavage with *Nco*I/*Xho*I followed by ligation with T4 DNA ligase yielded the complete pET28AMP_*Sap*I-Ubq amplification-expression vector (Data in Brief, Fig. 2) (GenBank MK606527) that was transformed into competent *E. coli* TOP10. The obtained bacterial clones were used for plasmid DNA purification. The purified plasmid DNAs were screened by restriction mapping and DNA sequencing. DNA sequences of the PCR primers used for vectors mutagenesis and construction are provided in Data in Brief.

2.5. 4th series of T7-lac promoter/colE1 ori/PhoA/MalE DNA amplification-expression-secretion pET28AMP_PhoA and pET28AMP_MalE vectors

The pET-28a(+)-del*Sap*I expression vector was digested with *Nco*I and *Xho*I, dephosphorylated with SAP and gel purified. Two alternative DNA cassettes with modified His6_PhoA_*Sap*I-*Sma*I-*Sap*I and His6_MalE_*Sap*I-*Sma*I-*Sap*I amplification modules were designed, synthetized and cloned into *Sma*I site of pUC59 DNA vector. The recombinant pUC59 plasmid DNA was digested with *Bsa*I and *Xho*I enzymes. The DNA restriction fragment containing the modified DNA cassette was gel purified and ligated with the *Nco*I/*Xho*I linearized pET-28a(+)del*Sap*I vector. The ligated DNA was transformed into competent *E. coli* TOP10. The recombinant plasmid DNA isolated from the obtained bacterial clones was screened by restriction mapping and DNA sequencing. Positive bacterial colonies revealed the correct construction of pET28AMP_PhoA (GenBank 606526) and pET28AMP_MalE (GenBank MK606522) amplification-secretion DNA vectors (Data in Brief: Fig. 3 and Supplemental data).

2.6. General scheme for directional DNA fragment amplification

In the following sections three variants of the DNA fragment amplification methodology are presented: i) using just one amplification round specified in Section 2.7.2, ii) additionally performing a second amplification round as specified in Section 2.7.3 and iii) an alternative amplification method specified in Section 2.7.4. All of them follow the same general protocol. Reaction steps include: (i) design or selection of the DNA fragment (monomer) to be amplified; (ii) chemical synthesis of DNA, PCR amplification or excision of the monomer using restriction endonucleases; (iii) addition of asymmetric *Sap*I 3-nt cohesive ends at 5′ and 3′ termini of the monomer. This can be achieved by introducing *Sap*I recognition sequences during chemical synthesis of the monomer, PCR amplification or with the use of the vector's built-in *Sap*I sites, after cloning the restriction fragment; (iv) purification of the DNA monomer equipped with *Sap*I 3-nt cohesive ends; (v) directional self-ligation of DNA monomers in head-to-tail orientation; (vi) ligation of a mixture of the formed concatemers or a selected gel-purified concatemer back into a *Sap*I-cleaved amplification vector; (vii) selection of bacterial clones containing a concatemer with the desired number of monomers; (viii) direct biosynthesis of the protein encoded by the obtained DNA concatemer using the vector's strong promoters or excision of the concatemer, equipped with *Sap*I-cohesive ends and repetition of steps (iv)–(viii), until a desired number of monomer copies within a concatemer is obtained.

2.7. Model amplification of 7-aa epitope HBV S protein gene

2.7.1. Cloning of a DNA fragment encoding an HBV epitope

For a proof of concept studies, a 7-aa epitope from S protein of HBV was selected [21]. The synthetic DNA fragment encoding the HBV epitope (TKPTDGN) was designed, optimized for *E. coli* expression, synthetized and subjected to a pilot amplification experiment in the vector pAMP1-HisA. For this purpose, a synthetic 5′-phosphorylated, 21 bp dsDNA fragment 5′-ACCAAACCGACCGACGGTAAC-3′ was cloned into the amplifying vector pAMP1-HisA and into *Sma*I restriction site (Data in Brief, Fig. 4). The obtained bacterial clones were used for plasmid DNA isolation. A correct orientation of the cloned insert was verified by DNA sequencing of the recombinant plasmid DNAs.

2.7.2. First round of amplification

The obtained DNA construct pAMP1-HisA_HBVepitope (GenBank MK606508) was used as a template DNA for PCR amplification. The

insert was amplified using Marathon DNA polymerase and starters Fnest3 and Rnest2 (Data in Brief: Supplemental data). The obtained PCR fragment was purified and subjected to *Sap*I cleavage. DNA cleavage was carried out for 2 h at 37 °C, in a reaction volume of 40 µl. Composition of the reaction mixture: 20 mM Tris-acetate (pH 7.9 at 25 °C), 10 mM magnesium acetate, 50 mM potassium acetate, 100 µg/ml BSA, 20 µg of the DNA fragment, 5 units of *Sap*I. The resulting *Sap*I restriction fragments were separated using polyacrylamide gel electrophoresis (PAGE) in 10% polyacrylamide gel in TBE buffer. The *Sap*I restriction fragment, encoding a modified HBV epitope, was cut out from the gel and purified. The *Sap*I modification comprised the 3-nt, single-stranded 5′ cohesive ends CCC/GGG. The purified DNA fragment (1.5 µg) was subjected to autoligation *in vitro* at 16 °C using 0.01 Weiss units of T4 DNA ligase with a reaction volume of 72 µl and samples taken at reaction time intervals of 5, 10, 20, 40, 80 and 160 min. The reaction buffer was: 40 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP (pH 7.8 at 25 °C). Ligation products were analysed electrophoretically, yielding a series of DNA segments of increasing length, forming directional concatemers of the polymerised epitope gene. The resulting concatemers were re-cloned into *Sap*I-linearized and SAP dephosphorylated pAMP1-HisA DNA vector, where they could be subjected to another amplification cycle or directly to expression of the artificial gene, a coding for poly-HBV_epitope protein. For DNA cloning, the following reaction composition and conditions were used: 40 mM Tris-HCl (pH 7.8 at 25 °C), 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, 540 ng of the autoligated insert DNA mixture, 270 ng of the *Sap*I-linearized and dephosphorylated pAMP1-HisA vector DNA, 1 Weiss unit of T4 DNA Ligase. The reaction was carried out at 16 °C for 16 h in a 90 µl reaction volume. Ligation products were purified and transformed into electrocompetent *E. coli* cells.

2.7.3. Second round of amplification

The 2nd round of amplification was performed similarly to the 1st round, except that: (i) pAMP1-HisA_HBVepitope_5, encoding five repeats of the HBV epitope, was used as a template DNA for PCR, (ii) *Sap*I restriction fragments were separated in 4% HR agarose gel. For DNA cloning, the following reaction composition and conditions were used: 40 mM Tris-HCl (pH 7.8 at 25 °C), 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, 100 ng of the autoligated insert DNA mixture, 100 ng of the *Sap*I-linearized and dephosphorylated pAMP1-HisA vector DNA, 0.5 Weiss unit of T4 DNA Ligase. The reaction was carried out at 16 °C for 16 h in a 30 µl reaction volume. Ligation products were purified and transformed into electrocompetent *E. coli* cells.

2.7.4. Alternative amplification

Alternatively, in order to boost DNA synthesis capability, a gene, encoding 25 copies of the epitope, was designed, chemically synthetized and cloned into the *Sap*I-linearized and dephosphorylated pAMP1-HisA (Data in Brief, Fig. 5). The obtained recombinant constructs contain DNA fragments, encoding ether 25 or 50 HBV epitope copies (Data in Brief: Supplemental data) (GenBank MK606514, MK606515). Alternative amplification was performed in two stages using the recombinant pAMP1-HisA_TKPTDGNGP_50 DNA construct. In the first stage, the selected DNA construct was cleaved with *Sap*I. The excised DNA fragment (2 µg), encoding 50 copies of the epitope, was purified and subjected to autoligation *in vitro* at 16 °C using 0.01 or 0.1 Weiss units of T4 DNA ligase with a reaction volume of 20 µl. The reaction buffer contained 40 mM Tris-HCl (pH 7.8 at 25 °C), 150 mM NaCl, 20% PEG4000, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP. Samples were taken at reaction time intervals of 15, 30, 45, 60 min. All the reaction products were combined and purified. In the second stage, the obtained DNA mixture was cloned into *Sap*I-linearized and dephosphorylated pAMP1-HisA DNA vector. For this purpose, the ligation reaction was performed in a volume of 20 µl containing: 40 mM Tris-HCl (pH 7.8 at 25 °C), 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, 400 ng of the autoligated insert DNA mixture, 50 ng of the *Sap*I-linearized and dephosphorylated pAMP1-HisA vector DNA, 1 Weiss unit of T4 DNA Ligase. Ligation was performed at 22 °C for 1 h. The reaction mixture was then incubated at 70 °C for 5 min, DNA ligation products were purified, ethanol precipitated, and used for *E. coli* transformation.

2.8. Construction of concatemeric genes encoding polyproteins with potential pro-regenerative properties

2.8.1. Cloning of DNA concatemers encoding poly-TSRGDHELLGGGAAPVGG

A synthetic 562 bp DNA cassette encoding the AGF fragment 10-mer TSRGDHELLGGGAAPVGG (AGF_10) was designed, optimized for expression in *E. coli*, chemically synthetized and cloned into pUC57 in *Eco*RV site. The cassette contained two convergent *Sap*I recognition sequences flanking the AGF_10 coding region (Data in Brief, Fig. 6). The recombinant pUC57_AGF_10 plasmid DNA was cleaved with *Sap*I. The restriction fragments have been separated using electrophoresis in 1.5% agarose gel, in TBE buffer and isolated from the gel using electroelution [22]. Thereafter, 200 ng of the purified insert DNA was pre-ligated for 1 h at 16 °C, using 1 Weiss unit of T4 DNA ligase. Then, 25 ng of the *Sap*I linearized and SAP dephosphorylated DNA vector (either pET28AMP_PhoA or pET28AMP_MalE) (Data in Brief, Fig. 3) was added to the reaction mixture and ligation reaction as continued for the next 2 h. The molar vector to insert ratio was 1 to 10. The ligated DNA was transformed into competent *E. coli* TOP10. The recombinant plasmid DNAs isolated from the obtained bacterial clones were screened by *Sap*I restriction analysis, colony PCR and DNA sequencing. Then, the resultant artificial fusion genes (Data in Brief: Supplemental data) were subjected to expression in *E. coli* BL21 Star (DE3) and production of poly-AGF protein variants was evaluated.

2.9. Protein purification and endotoxin removal

The detailed purification procedures of individual concatemeric proteins are provided in the Data in Brief. The identity of the obtained concatemeric proteins was confirmed by LC-MS-MS/MS analysis and immunodetection using anti-polyHistidine antibodies (not shown). For endotoxin removal from the obtained polyepitopic protein samples, 1 ml Pierce High-Capacity Endotoxin Removal spin columns were used. The resin was regenerated by washing with 0.2 N NaOH in 95% ethanol for 1–2 h at room temperature, centrifuged at 500 × *g* for 2 min and then, equilibrated with 20 ml of endotoxin-free PBS buffer [22]. The protein samples dialyzed against PBS buffer were applied to the equilibrated column and eluted with the same buffer.

2.10. LC–MS–MS/MS analysis

LC–MS–MS/MS analysis (liquid chromatography coupled to tandem mass spectrometry) was performed at a Mass Spectrometry Laboratory (IBB PAS, Warsaw, Poland) as described in the Data in Brief. The raw data were processed using Mascot Distiller followed by Mascot Search (Matrix Science, UK) against the predicted derived reference peptide masses. The search parameters are provided in the Data in Brief. Peptides with a Mascot Score exceeding the 5% False Positive Rate threshold and with a Mascot Score exceeding 30 were considered to be positively identified.

2.11. Western blotting

Purified proteins were separated by SDS-PAGE and electroblotted onto a PVDF membrane [22]. The membrane was probed with murine monoclonal anti-polyHistidine antibodies, conjugated with horseradish peroxidase. A specific protein was visualized by adding solution of peroxidase substrate. The detailed procedure is provided in the Data in Brief.

2.12. Culture of human fibroblasts and keratinocytes

In this study, two cell types were used: HaCaT keratinocytes from DKFZ and the 46BR.1N fibroblasts from ECACC. HaCaT (Human Adult Calcium High Temperature) cell line is originally derived from spontaneously transformed adult keratinocytes. Cells after > 140 passages were immortalized using the SV40 gene, which enabled them to maintain the possibility of differentiation and proliferation. The studies showed that this line is characterized by an altered and unlimited potential for growth and strongly resembles naturally occurring keratinocytes in the epidermis [23]. The 46BR.1N cell line is derived from spontaneously transformed human fibroblast from a patient with hypergammaglobulinemia. The immortalized cell line was obtained by transformation with the pSV3neo plasmid expressing the SV40 antigen [24]. Cells were routinely cultured on Primaria flasks in DMEM HG (Dulbecco's Modified Eagle Medium High Glucose - with 4.5 mg/ml of glucose), supplemented with 10% Fetal Bovine Serum in an incubator (37 °C, 5% CO_2). Culture medium was changed every 2–3 days.

2.13. Cell proliferation assay (XTT)

The XTT cell proliferation assay was done using Cell Proliferation Kit II. Cells were seeded at a density of 4000 cells per well into 96-well plates in DMEM HG medium supplemented with 10% of FBS. After 24 h, the media were exchanged for a serum-free DMEM containing appropriate concentrations of tested proteins. The cells were incubated with proteins for 48 or 72 h. Thereafter, XTT reagent was added and the plates were incubated at 37 °C for 4 h in the presence of 5% CO_2. Plates were then read using a standard plate reader at OD 490 nm. Cell proliferation was normalized with respect to the non-treated control (100%).

2.14. Cell cytotoxicity assay (LDH)

Cell cytotoxicity was quantified by measuring the lactate dehydrogenase (LDH) activity in culture supernatants. Cells were added into 96-well plates in DMEM HG medium supplemented with 10% FBS. After 24 h, the medium was changed to a serum-free DMEM HG containing appropriate concentrations of the tested proteins. All solutions used in the experiments were prepared with a phosphate buffered saline (PBS) solution (pH 7.4) under sterile conditions. After 48 h of incubation supernatants were collected for LDH analysis which was made following the manufacturer's instructions. Cell death was normalized with respect to the non-treated control (0%). Triton X-100 detergent (1%) was used as a positive control for maximum LDH release (maximum cytotoxicity).

2.15. Migration assay

Proteins effect on cell migration after 24 h was assessed using silicone Ibidi culture inserts. The cells were cultured in inserts for 24 h in DMEM HG medium supplemented with 10% FBS. After this time, medium was serum-free, and proliferation was blocked by adding mitomycin C (5 μg/ml) for 2 h. Furthermore, medium was changed, and the cells were stimulated to migrate at selected concentrations (0.1 and 1.0 μg/ml) of tested proteins. After 24 h cells were fixed with 3.7% paraformaldehyde, stained with 0.05% crystal violet and the migratory effect was measured with light microscope. Four representative photographs of each scratch were taken. The surface areas were measured and analysed by Imaging Software NIS-Elements Basic Research program.

2.16. Basophil activation test (BAT)

The reagent kit Flow CAST® Highsens was used to perform basophil activation tests. Blood obtained from healthy volunteers (n = 8) was collected in EDTA tubes, and the test was carried out within 24 h of blood sampling. A total of 100 μl blood was incubated separately with 20 μl of the concatemeric protein (1 μg/ml) and 20 μl of PBS. Samples were stained using 40 μl of CD63, CD203c and CCR3 antibodies mixture. Then, the samples were incubated for 15 min at 37 °C in a water bath. After the incubation, lysis and wash procedures were done. The samples were analysed up to 2 h using BD FACSCanto II flow cytometer. For each patient sample, negative and positive (anti-FceRI mAb and fMLP) stimulation controls were prepared. The cut-off values for analysed allergen were estimated as 5% of negative stimulation control sample.

2.17. Immunogenicity assays

Tests were conducted on human peripheral blood mononuclear cells (PBMC), isolated from 'buffy coats' using a ficoll density gradient. Following two wash steps (in PBS), erythrocyte lysis and cell counting (cell counter, Bio-Rad), the PBMCs were seeded onto a 24-well culture plate at a density of 1 mln cells/1 ml of RPMI 1640 (P/S, 10% FBS)/well. The cells were then allowed to adapt to the culture conditions for the next 24 h. After cell resting, the concatemeric proteins were added to the wells in a final concentration of 1 μg/ml, and the cells were incubated for the next 48 h under controlled conditions (37 °C, 5% CO_2). The untreated cells constituted a negative control, whereas cells stimulated with lipopolysaccharide (LPS) (1 μg/ml) and phytohemagglutinin (PHA) (2.5 μg/ml) were the positive control. After incubation, the PBMCs were collected, washed with PBS, counted and prepared for flow cytometry analysis under the following conditions: 104 cells/100 μl were stained with monoclonal, fluorochrome-labelled antibodies (anti-CD3, CD4, CD8, CD16, CD56, CD25, CD69, CD71, HLA-DR). Following 30 min of incubation at RT in the dark, the cells were analysed using a flow cytometer (LSRFortessa, BD). Statistical significance was checked with the Mann-Whitney *U* test ($p < 0.05$) using STATISTICA software and graphs were prepared with GraphPad Prism 5 software.

2.18. Capillary electrophoresis

Analysis of AGF_30 (30-mer fragment of AGF) stability in human plasma was performed using capillary electrophoresis (CE). A sample of blood was collected from a healthy volunteer and centrifuged in a lithium/heparin tube with EDTA as an anticoagulant. The plasma fraction was separated, mixed with AGF_30 (1.5 mg/ml, 1:1 v/v ratio) and incubated at 37 °C. A small sample (30 μl) was collected in a time-dependent manner and analysed by CE. CE analysis was performed on a P/ACE MDQ System (Beckman Coulter, USA), controlled by Karat software. An uncoated, fused silica capillary (Postnova Analytics GmbH, Germany), 40 cm (30 cm to detector) × 75 μm, thermostated at 25 °C, was used. Separations were performed at 20 kV with the following background electrolyte (BGE) composition: 25 mM phosphate (pH 7.0), 50 mM SDS, normal electrode polarization (cathode at the detector end). Samples were injected into the capillary at its anodic end *via* hydrodynamic injection at 0.5 psi for 8 s. The capillary was rinsed with 0.1 M NaOH and the BGE solution for 2 min between runs. The separation process was monitored with a UV detector at 214 nm.

3. Results

3.1. Development of the DNA fragment amplification-expression technology

To improve the existing techniques of DNA fragment amplification and enhance advanced RNA and protein expression capability [5–15], we designed and constructed four series of amplification-expression vectors (Data in Brief: Fig. 1 and Supplemental data), dedicated to the formation of artificial, continuous, multimeric ORFs (Fig. 1). The vectors based on two widely used vector groups: (i) the *p15A* origin vector pACYC184 (containing chloramphenicol and tetracycline resistance

genes; the latter marker further removed) [19] and (ii) the *colE1* origin pET21d(+) and pET28a(+) vectors from Novagen's pET system [24], containing IPTG-inducible bacteriophage T7-lac hybrid promoter and ampicillin or kanamycin resistance gene.

3.1.1. 1st series pR promoter/p15A ori DNA amplification-expression pAMP vectors

Firstly, we designed and constructed 6 variants of amplification module, containing two convergent *Sap*I DNA recognition sequences for straightforward, in-frame, head-to-tale amplification of any DNA fragment, resulting in the assembly of an artificial ORF (Data in Brief, Fig. 1). The vectors included options for ligation of a DNA fragment to be amplified in 3 ORFs phases as well as inclusion of His6 tag. We used the Type IIS *Sap*I REase. The REase recognises a relatively long DNA sequence of 7 bp: 5′-GCTCTTC-3′, which is unique in the vectors and in all the relatively short amplified DNA segments, as the cleavage probability is one per 16,384 bp. The enzyme cuts DNA 1/4 nt downstream from the recognition sequence. It generates 3-nt cohesive ends, corresponding to a single codon length. Thus, it is uniquely suited for maintaining continuity of newly formed, concatemeric ORFs, as tandem ligation of the DNA fragments, containing 3-nt 5′ overhangs, does not lead to DNA frameshifts and enables a straightforward method for formation of the artificial ORFs. Additionally, both DNA scission and ligation of the DNA ends generated by *Sap*I is very efficient in comparison to other investigated Type IIS REases, including those used in previous versions of a DNA concatemer construction methods (not shown) [25]. Out of thousands REases known, *Sap*I is the only one ideal enzyme available for cloning of multiple DNA fragments using the developed technology. From the designed and constructed pACYC184-based vectors (Data in Brief: Fig. 1 and Suplemental data), we selected pAMP1-HisA for the technology development (Fig. 2) (Data in Brief: Supplemental data). The pAMP are medium copy number vectors and were designed for a high expression of the 'artificial' multimeric genes, placed under the control of the strong, temperature-inducible lambda bacteriophage *pR* promoter. The medium copy number pAMP1-HisA vector carries chloramphenicol resistance gene (GenBank MK606507). The vector is designated for cloning into +1 frame and enables N-terminal fusion of the polyepitopic protein with the His6 tag, for efficient protein purification using immobilized metal affinity chromatography (IMAC) (Fig. 2). Preferably, we used the vector's *Sap*I-generated 3-nt cohesive ends for cloning of the insert DNA, equipped with compatible 3-nt cohesive ends (Fig. 1), obtained during annealing of synthetic, single-stranded oligos. The oligos were designed in such a way that they form double-stranded insert DNA, while leaving 5′, 3-nt single-stranded termini (Fig. 1). In the pAMP1 vectors (GenBank: MK606505, MK606506, MK606507, MK606519, MK606520, MK651654), the *Sap*I sites are separated by the *Sma*I recognition sequence, which was introduced as an alternative variant of insert cloning (Data in Brief, Fig. 1). *Sma*I is an orthodox Type IIP REase, which cuts DNA within the recognition sequence, generating 'blunt' ends. Alternatively, the vector cut with *Sap*I can be filled-in with dGTP and dCTP for the purpose of blunt end cloning. During *Sap*I/dGTP/dCTP cloning only 1 codon (for proline) is introduced in between the monomeric unit, just like during *Sap*I cohesive ends cloning, while during *Sma*I cloning, 2 codons were introduced (proline-glycine). A vector prepared with *Sap*I or *Sma*I may be cloned with any arbitrary DNA segment, synthetic or natural, which could then be amplified. The amplified DNA segment can encode an antigen or aa sequence of desired biological or chemical function, encompassing several identical or differing peptides. The only limit is the length of the amplified fragment, as dictated by the length of the insert DNA accepted by a given class of DNA vector. The amplifying module may be transferred to different classes of vectors using molecular cloning.

3.1.2. 2nd series of T7-lac promoter/colE1 ori DNA amplification-expression pET21AMP vectors

The second type of DNA amplification-expression vectors was based on pET21d(+) vector, containing a strong, IPTG inducible T7-lac promoter and ampicillin resistance gene. The pET system is powerful and ideal for high-level expression of genes. It enables protein production at the temperature range from 10 °C to 42 °C, which can be further boosted by several °C, depending on *E. coli* host. The recombinant product can

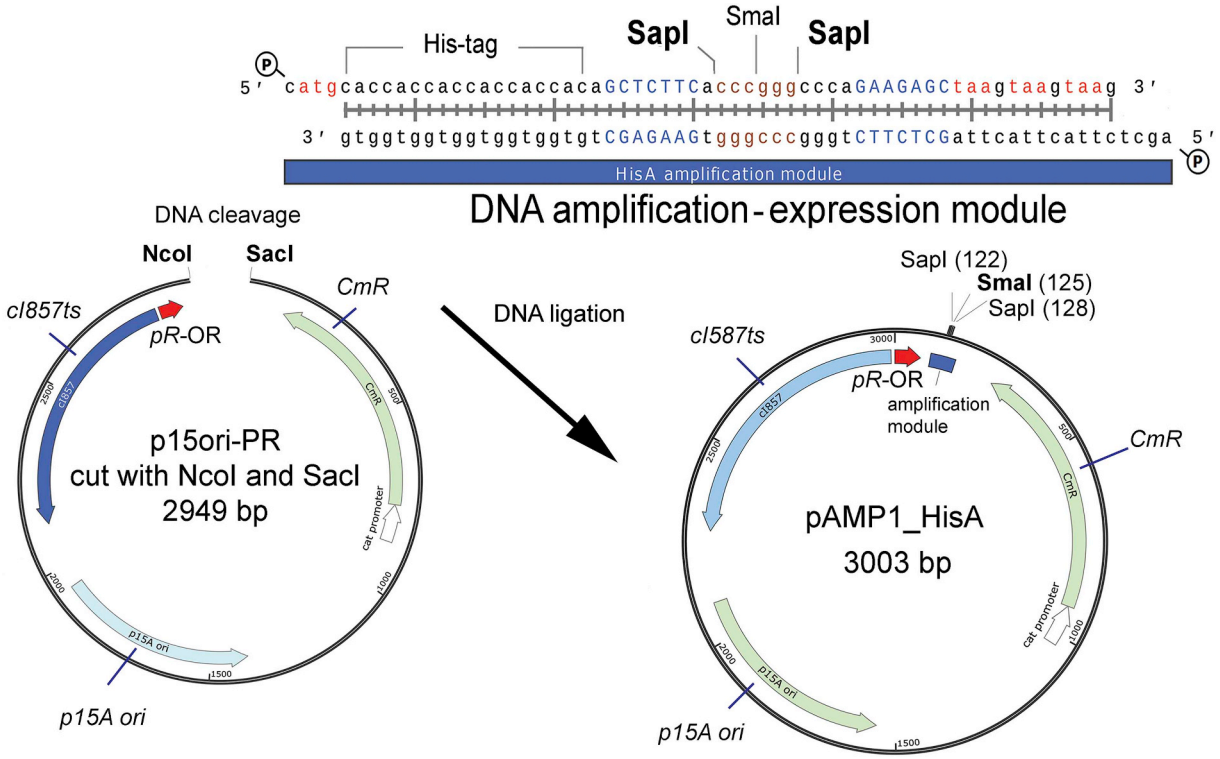

Fig. 2. Diagram showing construction of the pAMP1-HisA amplification-expression vector and DNA sequence of the HisA amplification module.

comprise > 50% of the total cell protein. Thus, the obtained pET21AMP vectors (amplification modules presented in Data in Brief, Fig. 1) (pET21AMP_HisA: GenBank 606521) are an ideal choice for biosynthesis of the concatemeric proteins, which are non-toxic for *E. coli*. Moreover, one can easily attenuate the concatemeric gene expression level by lowering the concentration of inducer or decreasing the temperature of bacterial growth after induction. The reduction in the temperature can also reduce the activity of bacterial proteases and potential fragmentation of the recombinant concatemeric protein.

3.1.3. 3rd series of T7-lac promoter/colE1 ori/ubiquitin fusion DNA amplification-expression pET28AMP_SapI-Ubq vectors

For the third type of DNA amplification-expression vectors, pET28a (+) vector was used. A specialised DNA cassette was designed, chemically synthetized and cloned into the *Nco*I and *Xho*I sites of the pET28a(+) with previously eliminated internal *Sap*I restriction site. The cassette contained His6 tag, c-Myc tag, WYY tag and ubiquitin as well as amplification module with two convergent *Sap*I recognition sequences (Data in Brief, Fig. 2). This vector series, exemplified by pET28AMP_*Sap*I-Ubq (GenBank MK606527) (Data in Brief: Supplemental data), was designed for construction of the concatemeric proteins, N-terminally fused with ubiquitin. The aromatic WYY tag was introduced to facilitate the detection of recombinant protein using UV light and Coomassie Blue staining of electrophoretic gels. The c-Myc tag was added for immunodetection of the protein with dedicated anti-c-Myc antibodies. The His tag was added for metal-affinity chromatography and Western blotting detection of the protein with dedicated anti-His6 tag. The ubiquitin domain can improve the level of gene expression and protein biosynthesis, decrease protein toxicity and improve its stability *in vivo*.

3.1.4. 4th series of T7-lac promoter/colE1 ori/PhoA/MalE DNA amplification-expression-secretion pET28AMP_PhoA and pET28AMP_MalE vectors

This type of DNA amplification-expression vectors is based on the pET28a(+) with eliminated internal *Sap*I site, similar to the 3rd series (Data in Brief: Fig. 3 and Supplemental data) (GenBank MK606522, MK606526). It enables secretion of the concatemeric proteins into the periplasmic space. The secretion is facilitated either by PhoA or MalE secretion leaders. Additionally, the leaders tend to improve the expression level of recombinant genes, which may apply to the concatemeric genes as well. The secreted, concatemeric proteins were of full-length, as expected from their coding genes sequences. According to our observations, the amplification-expression-secretion vectors are a good choice for the biosynthesis of acidic proteins with low pI, such as the AGF-derived concatemeric proteins.

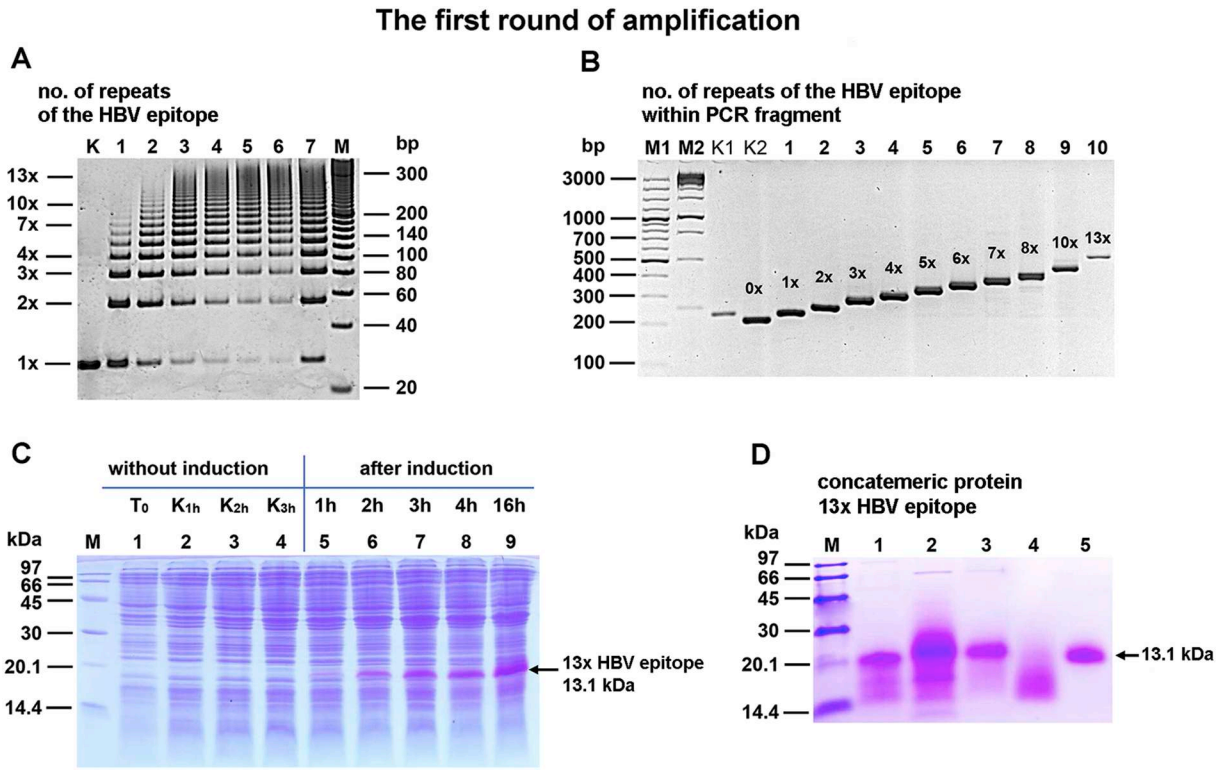

Fig. 3. Results of the first round of amplification of the model 7-aa HBV epitope. Panel A. Electrophoretic analysis of the DNA autoligation products. Lane K, non-ligated DNA insert encoding one HBV epitope; lanes 1–6, DNA autoligation products after incubation with T4 DNA ligase for 5, 10, 20, 40, 80 and 160 min, respectively; lane 7, a mixture of all the DNA autoligation products from lanes 1–6; lane M, Sigma PCR 20-bp Low Ladder (selected bands marked). Panel B. PCR amplification of the obtained artificial genes encoding repeats of the HBV epitope. Lane M1, GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific); lane M2, GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific); lane K1, control PCR fragment 230 bp; K2, control PCR fragment (template: pAMP1-HisA); lanes 1–10, colony PCR fragments (templates: pAMP1-HisA recombinant constructs), encoding 1, 2, 3, 4, 5, 6, 7, 8, 10 and 13 repeats of the HBV epitope, respectively. Panel C. Biosynthesis of the recombinant 13-mer in *E. coli* BL21(DE3). The proteins were separated by SDS-PAGE and stained with Coomassie Blue R250. Lane M, protein ladder (LMW Calibration Kit, GE Healthcare); lanes 1–4: recombinant *E. coli* culture, containing a 13-mer HBV construct prior to induction, and sampled at 0, 1, 2 and 3 h. lanes 5–9: recombinant *E. coli* cultures, containing the 13-mer HBV construct after thermal induction, sampled at 1, 2, 3, 4 and 16 h. The characteristic pink-violet protein band representing the poly-HBV epitope protein (13-mer) is marked with an arrow. Panel D. Purification of the recombinant poly-HBV epitope protein (13-mer; 13.1 kDa). Lane M, Protein Ladder (GE Healthcare); lane 1, after CM Sepharose chromatography; lane 2, after Immobilized Metal Affinity Chromatography (IMAC); lane 3, after Resource S cation exchange chromatography; lane 4, after size exclusion chromatography (SEC) - fraction of the protein with lower molecular weight; lane 5, after SEC – fraction of the protein with higher molecular weight. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Biosynthesis of a polyepitopic protein containing a model 7 aa epitope of the HBV antigen S

3.2.1. The first round of DNA amplification

A model, synthetic DNA fragment, encoding a 7 aa epitope of the HBV surface antigen [21], was cloned and subjected to a pilot amplification experiment in the vector pAMP1-HisA as described in Material and methods (Data in Brief, Fig. 4). For this purpose, the pAMP1-HisA-HBV epitope construct (GenBank MK606508) was digested with *Sap*I. The resulted, excised DNA fragment contained the modified epitope gene. The modification consisted of the flanking 3-nt, single-stranded 5′ cohesive ends. Aside from the amplification enabling function, these ends are responsible for the additional proline residues (the so-called 'helical breakers'), separating the epitopes in the final polymeric, concatemeric protein. These prolines facilitate the independent folding of the multiplied epitopes into tertiary structures, thereby assist in maintaining their natural spatial structure. The number of added 'helical breakers' can be regulated arbitrarily at the DNA level. If required, a suitable number of codons for proline and/or glycine could be added to the end of the synthetic DNA fragment, encoding the epitope. The excised HBV epitope mini-gene (24-bp in length of *Sap*I restriction fragment) was subjected to autoligation *in vitro* and the ligation products were analysed by PAGE electrophoresis (Fig. 3A). As shown in Fig. 3A, the ligation products form a ladder of DNA segments with regularly increasing length. These DNA segments are directional polymers of the original gene. The autoligation reaction kinetics is reflected in the lanes 1–6 (Fig. 3A). Clearly, after only 20 min (Fig. 3A, lane 3), the reaction is nearly completed, reaching plateau of tens of monomeric units present within the formed concatemers. A mixture of the autoligation products (Fig. 3A, lane 7) was re-cloned into pAMP1-HisA backbone, previously cleaved with *Sap*I and dephosphorylated. As a result, we obtained multiple monomer copies within the formed DNA concatemers *in vitro*, where a mixture of long concatemers exceeded resolution capacity of the PAGE used. This indicates the presence of tens or more copies of monomeric units within concatemers (Fig. 3A). Analysis of just a few bacterial clones revealed 13 copies of the epitope gene (Fig. 3B). The resultant gene encoded thirteen linked in frame copies of the epitope, forming recombinant, poly-HBV epitopic protein, which is further biosynthesized *in vivo*. Fig. 3B shows an electrophoretic analysis of the colony PCR products, amplified using a series of positive recombinant bacterial clones, containing inserts with 2–13 copies of the epitope gene. The DNA fragments migrate more and more slowly in the agarose gel, due to their increasing length and as a direct effect of the number of linked epitope copies. DNA sequencing of the resulted recombinant DNA constructs confirmed the continuity of the ORFs. The recombinant *E. coli* clone, containing a pAMP1_HisA_HBVepitope_13 construct

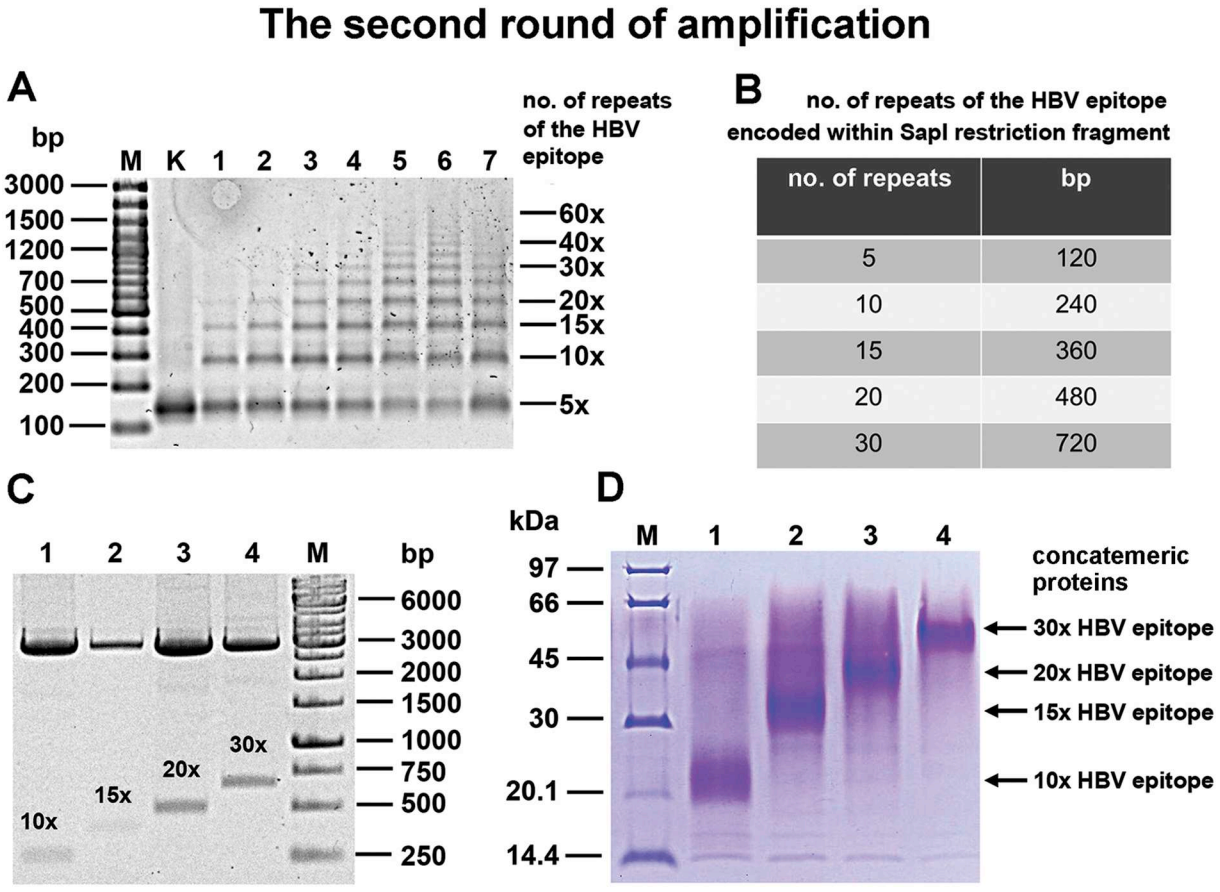

Fig. 4. Results of the second round of amplification of the 5-mer of the 7-aa HBV epitope. Panel A. Electrophoretic analysis of the DNA autoligation products. Lane M, GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific); K, non-ligated DNA insert encoding concatemeric protein containing five repeats of the HBV epitope; lanes 1–6, DNA autoligation products after incubation with T4 DNA ligase for 5, 10, 20, 40, 80 and 160 min, respectively; lane 7, a mixture of all the DNA autoligation products from lanes 1–6. Panel B. The predicted restriction fragment lengths following excision of the inserts with *Sap*I. Panel C. *Sap*I restriction analysis of the obtained recombinant pAMP1-HisA constructs encoding polyepitopic proteins of different length. Lane 1, pAMP1-HisA_HBVep_10; lane 2, pAMP1-HisA_HBVep_15; lane 3, pAMP1-HisA_HBVep_20; lane 4, pAMP1-HisA_HBVep_30; lane M, GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific). Panel D. Purified recombinant polyepitopic protein variants after HiTrap™ IMAC FF chromatography. The proteins were separated by SDS-PAGE and stained with Coomassie Blue R250. Lane M, protein ladder (LMW Calibration Kit, GE Healthcare); lane 1, the 10-mer of the HBV epitope (10.5 kDa); lane 2, the 15-mer (14.9 kDa); lane 3, the 20-mer (19.2 kDa); lane 4, the 30-mer (28.7 kDa). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(GenBank 606510) (Data in Brief: Supplemental data), was selected for a pilot poly-HBV epitope gene expression analysis. The presence of the recombinant HBV poly-epitope protein (13-mer) was investigated in bacterial samples collected from recombinant *E. coli* culture before and after thermal induction of the recombinant gene expression (Fig. 3C, D). The analysis was performed by SDS-PAGE and Coomassie brilliant blue (CBB) R250 staining [21]. One could note a characteristic, purple-pink colour of the protein band, resulting from distinctive CBB R250 staining of the proline-rich [26,27] 13-mer (TKPTDGNGP_13) (Fig. 3C, D). The other *E. coli* proteins stain blue. This unique feature of the poly-HBV epitope protein variants facilitated their detection during the purification process. The purification scheme for the HBV poly-epitope proteins included the following stages: cell lysis, heat treatment, precipitation of nucleic acids with polyethyleneimine (PEI), ammonium sulfate (AmS) precipitation, HiTrap™ CM FF chromatography, HiTrap™ IMAC FF chromatography, Resource™ S chromatography, size exclusion chromatography (SEC) using HiLoad 16/600 Superdex 200 pg column. The detailed purification procedure is outlined in the Data in Brief. Fig. 3D shows SDS-PAGE analysis of the protein purity after the chromatographic steps.

3.2.2. The second round of DNA amplification

The 2nd round of amplification was performed as described above, except that a concatemer encoding 5 epitope copies was used as the initial DNA 'monomer'. The autoligation products of the pentamer are shown in Fig. 4A, B. The largest concatemer, visible as separate band at the edge of agarose gel resolution, contains 12 copies of the pentamer, constituting a 60-fold directionally polymerised HBV epitope gene. Larger concatemers are evidently present, although not separated into distinct bands (Fig. 4A, B). The autoligation products were re-cloned into *Sap*I cleaved pAMP1-HisA vector (Fig. 4C) (Data in Brief: Supplemental data) (GenBank MK 606509, MK606511, MK606512, Mk606513) and subjected to analytical expression in order to obtain variants of epitope multiplication within the polyepitopic protein. The resulted proteins were purified as described for the recombinant poly-HBV epitope 13-mer (Fig. 4D). Both the 1st and the 2nd amplification round were highly efficient, resulting in the conversion of the majority of monomers into concatemers within just a few minutes, accumulating steadily until over 90% of monomers were consecutively linked (Data in Brief, Fig. 7). This proves the stability and anti-interference ability of the system.

3.2.3. Alternative method for DNA fragment amplification

This scheme was based on amplification of the longest concatemeric DNA oligomer, obtainable by chemical synthesis. The amplification was performed as described above, except that a synthetic gene encoding 25 epitope copies (Data in Brief: Fig. 5 and Supplemental data) was used as the initial DNA 'monomer' (Fig. 5A). The DNA sequence of the synthetic gene was codon optimized for efficient protein biosynthesis in *E. coli* (Data in Brief, Fig. 5). The chemical synthesis was at its limits due to the repeats content, thus of very low efficiency. Nevertheless, the gene was cloned initially to pUC57mini vector, from which it was further excised (not shown) and used for the amplification reaction. The autoligation products were cloned into *Sap*I-cleaved pAMP1-HisA vector and the obtained recombinant DNA constructs were analysed by restriction mapping and DNA sequencing (Fig. 5B). Using this strategy, we obtained a series of recombinant bacterial clones, containing 25 to 500 copies of the HBV epitope gene cloned-in as single continuous ORFs within the pAMP1-HisA vector (Fig. 5B) (GenBank MK 06514, MK606515, MK60616, MK 606517, MK606518). The resulted positive bacterial clones were subjected to analytical expression in *E. coli* BL21(DE3) gold (Fig. 5C, D). The TKPTDGNGP_25 polyepitopic protein was purified to homogeneity (Fig. 5C) and its immunological activity was investigated (not shown). The resulted variants of the polyepitopic proteins were thermostable and exhibited the characteristic, purple-pink colour of the protein band (Fig. 5C, D), as described for 13-mer (TKPTDGNGP_13). Somewhat surprisingly, the produced recombinant proteins did not form sharp protein bands when separated using SDS-PAGE (Fig. 5C, D). Instead of sharp protein bands, they formed characteristic purple-pink 'smears'. A range of molecular weights of the separated recombinant polypeptides depended on the number of the HBV epitope repeats within the investigated concatemeric protein (Fig. 5D). The observed phenomenon probably results from: (i) the unstructured character of the polyepitopic proteins, (ii) their tendency to form dimers or larger aggregates and (iii) various length of the biosynthesized polypeptides due to reduction in transcription/translation efficiency of the repetitive DNA/mRNA sequences in the higher molecular weight range, which results in a mixture of polyepitopic proteins of very similar, but not identical length. Regardless of the selected variant of the amplification methodology and in contrast to PCR reactions, the accumulation of formed concatemers occurs linearly as no additional substrate (template) is generated during DNA ligation. One should also note that the amplification reaction is self-limiting because: (i) free 3-nt cohesive ends of the monomeric substrate are gradually converted into dsDNA CCC/GGG segments, located between the ligated monomeric units and (ii) the availability of the substrate monomer containing 3-nt cohesive ends decreases during the course of the reaction (Data in Brief, Fig. 7). To summarize, the DNA fragment amplification scheme for each of the presented variants of the technology consists of the following steps: (1) selection/design of a DNA fragment coding for a peptide with a desired function; (2) obtaining the monomeric DNA fragment by: (a) chemical synthesis, (b) PCR amplification or (c) restriction endonucleases excision; (3) equipping the monomer with asymmetric *Sap*I 3-nt cohesive ends. Options (a) and (b) are suitable for directly introducing *Sap*I sites at the 5′ and 3′ termini of the monomer DNA fragment. Option (c) is suitable for blunt end cloning into the *Sma*I site of the amplification vector and further equipping the cloned monomer with 3-nt cohesive ends using the vector's *Sap*I sites; (4) purification of the DNA monomer containing 3-nt *Sap*I cohesive ends; (5) autoligation of DNA monomeric fragments in a directional fashion and head-to-tail orientation; (6) cloning of either a mixture of the formed concatemers or a gel-purified concatemer of a desired length back into a *Sap*I-cleaved amplification vector; (7) selection of positive bacterial clones containing concatemers with the desired number of monomers; (8) direct expression of the cloned concatemeric gene or its re-excision with *Sap*I and repetition of steps [4–8] in order to obtain a desired number of monomer copies within the resulting concatemer.

3.3. Biosynthesis and purification of concatemeric proteins, containing repeated fragments of an angiopoietin-related growth factor

3.3.1. Designing of a synthetic gene, cloning, expression and purification of the AGF-derived concatemeric protein

A potentially bioactive, 9-aa peptide TSRGDHELL had been selected as a target for polyepitopic protein construction. The peptide originated from the angiopoietin-related growth factor (AGF), which is known to promote epidermal proliferation, new blood vessel formation and wound healing in the skin [28,29]. The peptide was fused with the GG helical breakers and the AAPV peptide, which is being efficiently digested by human elastase. Such a strategy was expected to result in a gradual, enzymatic release of the bioactive peptide from the polyepitopic protein, when applied to a wound. A dsDNA fragment encoding ten repeats of the TSRGDHELLGGAAPVGG (AGF_10) was designed (Fig. 6A; Data in Brief, Fig. 6) and their DNA sequence was optimized for expression in *E. coli*. The fragment was flanked with *Sap*I recognition sequences, enabling its amplification using the amplification-expression vectors. The designed DNA fragment was chemically synthetized and cloned into pUC57mini. The DNA cassette was excised from the recombinant pUC57mini by *Sap*I cleavage and subjected to autoligation. Furthermore, the concatemeric ligation products were cloned into pET28AMP_PhoA or pET28AMP_MalE, previously cleaved with *Sap*I (Data in Brief: Figs. 3 and 6, Supplemental data). These DNA

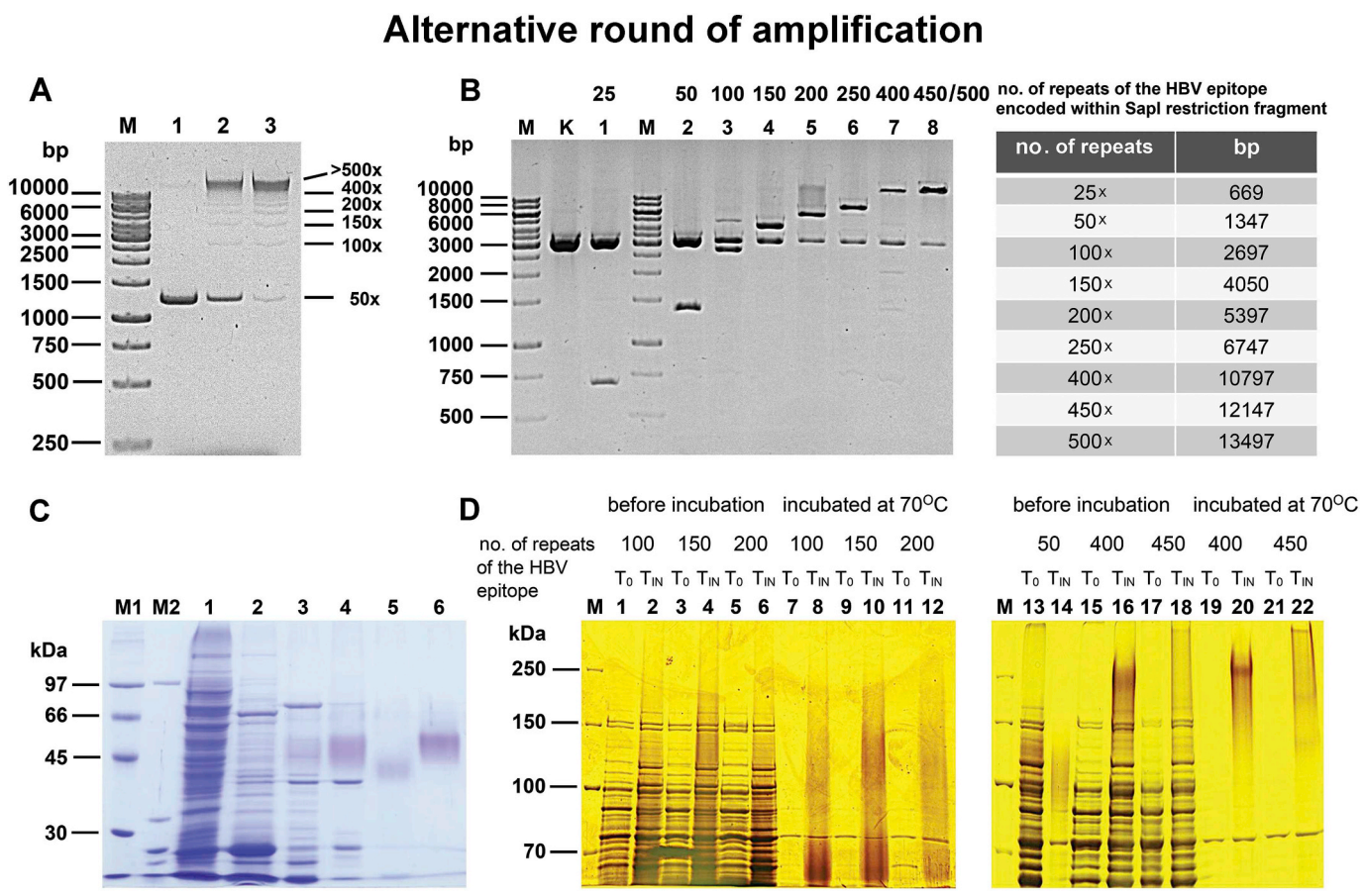

Fig. 5. Results of the alternative round of amplification of the model HBV epitope. Panel A. Electrophoretic analysis of the DNA autoligation products. Lane M, GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific); lane 1, non-ligated DNA insert encoding concatemeric protein with 50 repeats of the HBV epitope; lane 2, mixed DNA autoligation products obtained after 15-, 30-, 45- and 60-min ligation. Incubation with 0.01 Weiss unit of the T4 DNA ligase at 16 °C; lane 3, as in lane 2, but with 0.1 Weiss unit of the T4 DNA Ligase. Panel B. *Sap*I restriction analysis of the obtained recombinant pAMP1-HisA constructs. The predicted sizes of the *Sap*I-excised inserts encoding polyepitopic protein variants are shown in the right column of the table. Panel C. Purification of the recombinant poly-HBV epitope protein (25-mer; 23.5 kDa). Lane M1, Protein Ladder (GE Healthcare); lane M2, PageRuler Unstained Low Range Protein Ladder (Thermo Fischer Scientific); lane 1, crude lysate, lane 2, after 30 min incubation at 70 °C; lane 3, after AMS fractionation; lane 4, after IMAC; lane 5, after Resource Q anion exchange chromatography; lane 6, after SEC. Panel D. Expression screening of the obtained, recombinant *E. coli* BL21(DE3) gold clones, producing poly-HBV epitope protein variants. The bacterial crude lysates were analysed by SDS/PAGE. Lane M, Protein Ladder (Thermo Fischer Scientific); lane 1, bacteria producing the 100-mer (88.6 kDa) of the HBV epitope, before induction; lane 2, bacteria producing the 100-mer after thermal induction; lane 3, the 150-mer (132 kDa), before induction; lane 4, the 150-mer after induction; lane 5, the 200-mer (175.4 kDa), before induction; lane 6, the 200-mer after induction; lanes 7–12, as in lanes 1–6, except that the crude lysates were incubated at 70 °C for 30 min and centrifuged; lane 13, bacteria producing the 50-mer (45.2 kDa), before induction; lane 14, bacteria producing the 50-mer after induction; lane 15, the 400-mer (32.4 kDa), before induction; lane 16, the 400-mer, after induction; lane 17, the 500-mer, before induction; lane 18, the 500-mer, after induction; lanes 19–22, as in lanes 15–18, but after heat treatment.

vectors enabled the secretion of polyepitopic protein to the periplasmic space. This strategy was selected to reduce a possible proteolytic degradation of the recombinant polyepitopic proteins. The resulted recombinant bacterial clones were analysed using the colony PCR, *Sap*I restriction mapping and DNA sequencing. The presence of concatemeric genes, containing 10, 20, 30, and 40 repeats of the TSRGDHELLGGAAPVGG was confirmed. For gene expression experiments, bacterial clones producing concatemeric proteins with 10, 20, 30, 40 repeats of the TSRGDHELLGGAAPVGG were selected (Fig. 6B, C) (Data in Brief: Supplemental data) (GenBank MK606523, MK606524, MK659889, MK606525). The recombinant genes were expressed in *E. coli* BL21(DE3). The resulted polyepitopic protein variants were purified to homogeneity (Fig. 6C) and their sensitivity to human elastase was confirmed (Fig. 6D, lane 2). Interestingly, a spontaneous fragmentation of the concatemeric proteins was observed even without elastase. A characteristic ladder of the degraded protein fragments was observed below the major protein band, corresponding to the full-length TSRGDHELLGGAAPVGG_20 (AGF_20) polypeptide (Fig. 6D, lane 1). Although, this phenomenon was observed for all the AGF-derived concatemeric proteins, it was more visible for the AGF_10 (Fig. 6C, lane 1) and AGF_20 (Fig. 6D, lane 1) variants.

3.4. Biological effects of the AGF-derived concatemeric proteins on human skin keratinocytes and fibroblasts

3.4.1. Proliferation and viability assays

To exclude a potential cytotoxicity of the AGF-derived concatemeric protein variants both the XTT and LDH tests were used. For this purpose, keratinocytes and fibroblasts were cultured in the presence of various concentrations of the proteins. The cytotoxicity was evaluated by measurement of LDH release after 48 h (Data in Brief, Fig. 8a, b). The level of LDH release after treatment with a strong non-ionic detergent Triton X-100 was used as a positive control (maximum toxicity). The cytotoxicity assays showed that the investigated concatemeric proteins did not increase the LDH release by 46BR.1N fibroblasts and HaCaT keratinocytes, regardless of the concentration used (Data in Brief, Fig. 8a, b). It indicates that the poly-TSRGDHELLGGAAPVGG concatemeric proteins are not cytotoxic to these cells. Cell proliferation

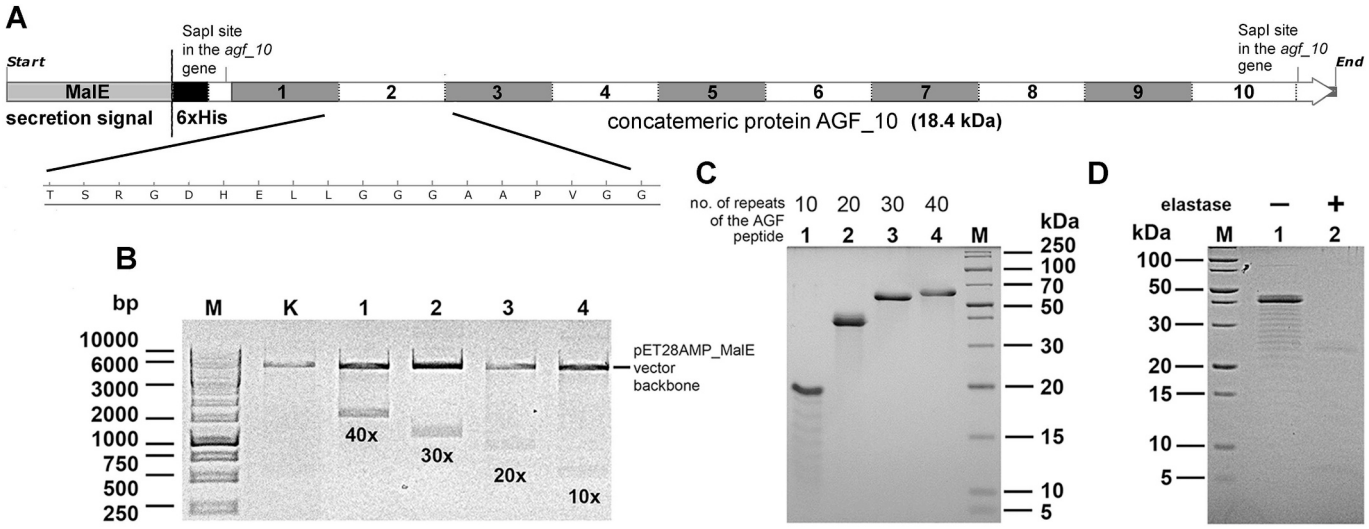

Fig. 6. Obtaining the AGF-derived concatemeric protein variants with a potential pro-regenerative activity. Panel A. Diagram representing the AGF-derived concatemeric protein, containing 10 repeats of the TSRGDHELLGGAAPVGG peptide. Panel B. *Sap*I restriction analysis of the obtained recombinant pET28AMP-MalE constructs. The number of the AGF-derived peptide repeats encoded by *Sap*I-excised DNA fragment is shown under the corresponding DNA band. Panel C. SDS-PAGE analysis of the purified AGF-derived concatemeric protein variants: lane 1, the 10-mer; lane 2, the 20-mer; lane 3, the 30-mer; lane 4, the 40-mer. Panel D. Cleavage of the AGF-derived 20-mer with human elastase. Lane M, PageRuler Unstained Broad Range Protein Ladder (Thermo Fischer Scientific); lane 1, non-treated the 20-mer; lane 2, the 20-mer cleaved with elastase. Note that the AGF-derived concatemeric proteins undergo a gradual, spontaneous fragmentation.

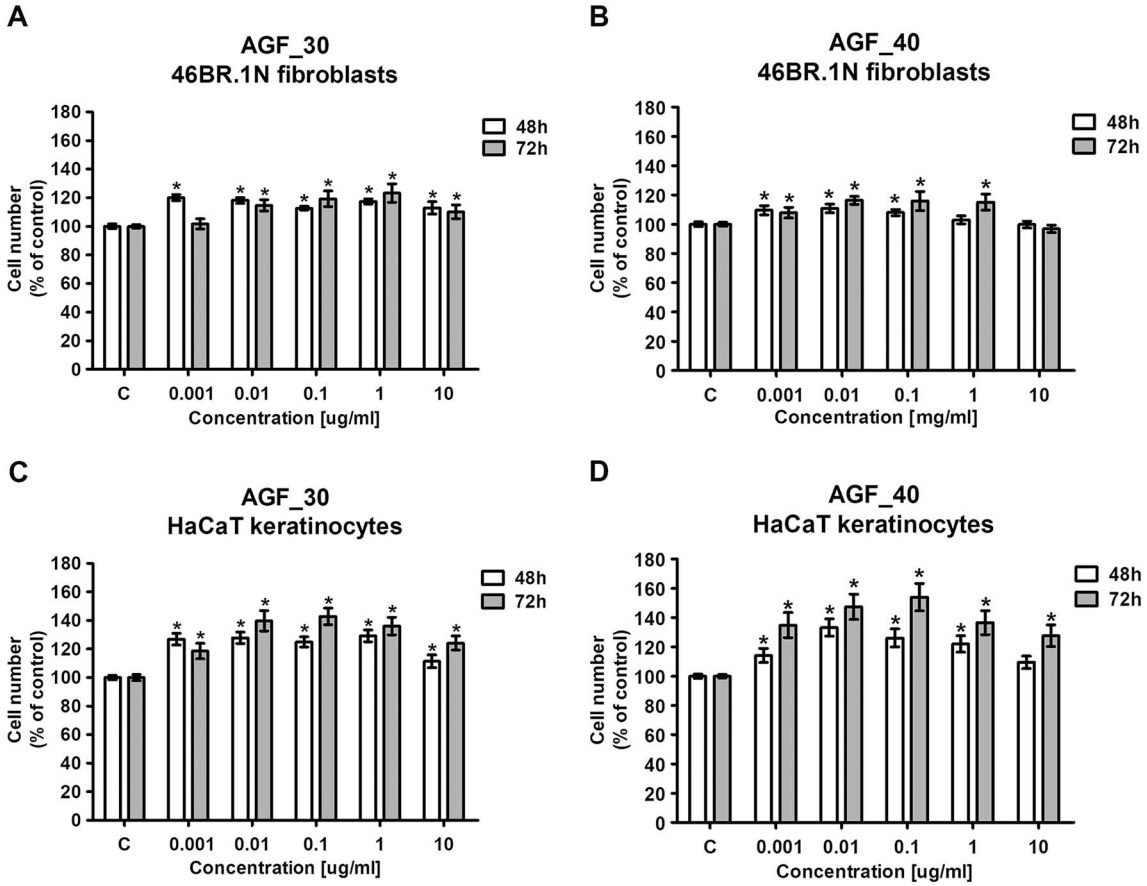

Fig. 7. Effects of the AGF-derived concatemeric protein variants on proliferation and viability of human fibroblasts and keratinocytes. Results obtained for the AGF_30 protein (TSRGDHELLGGAAPVGG_30) and for the AGF_40 protein (TSRGDHELLGGAAPVGG_40). The influence of these protein variants on the proliferation of 46BR.1N fibroblasts (panels A and B) and HaCaT keratinocytes (panels C and D) after 48 or 72 h of incubation. Results are presented as the mean ± SEM. All results come from four independent experiments performed in quadruple. *$p < 0.05$ relative to control, Mann-Whitney *U* test.

plays a crucial role in wound healing and tissue regeneration. Thus, we decided to check the investigated concatemeric proteins' effect on skin cells using this process. Proliferation of cells was estimated by colourimetric XTT test. In case of the TSRGDHELLGGAAPVGG_30 (AGF_30) concatemeric protein, XTT assay showed a slight increase (10–20% compared to the control) in the proliferation of 46BR.1 N fibroblasts in all tested concentrations (Fig. 7A). Similarly, a pro-proliferative effect on HaCaT keratinocytes was noticeable (Fig. 7C). The observed effect was stronger (30–40% compared to the control) than for fibroblasts. A significant increase in proliferation was observed after 72 h of incubation, at the concatemeric protein concentration range from 1 ng/ml to 1 μg/ml (Fig. 7C). In the XTT assay, the TSRGDHELLGGAAPVGG_40 (AGF_40) concatemeric protein showed a slight effect (10–15% compared to the control) on proliferation of 46BR.1N fibroblasts in concentrations of 1 ng/ml–1 μg/ml (Fig. 7B). Similarly to the AGF_30, it significantly (30–50% compared to the control) increased proliferation of HaCaT keratinocytes in all tested concentrations. The strongest stimulation was visible for concentrations of 10 ng/ml and 100 ng/ml (Fig. 7D).

3.4.2. Cell migration

The migration assay showed that the AGF_40 stimulates migration of 46BR.1N fibroblasts. Scratch area was smaller by approximately 30% and 20% in concentrations of 100 ng/ml and 1 μg/ml respectively compared to the control. There is also a slight (10% compared to the control) stimulation of HaCaT cells migration in 100 ng/ml concentration. The results are shown in Data in Brief, Fig. 9a.

3.4.3. Cell chemotaxis

The chemotaxis assay showed that the AGF_40 concatemeric protein slightly (about 10% compared to the control) stimulates chemotaxis of HaCaT keratinocytes in 1 μg/ml concentration after 24 h incubation. Interestingly, the AGF_30 does not affect chemotaxis (Data in Brief, Fig. 9b).

3.5. The AGF-derived concatemeric proteins immunogenicity and allergic potential

In order to assess the potential immunogenic influence and the allergic potential of the analysed compounds on the human immune system, peripheral blood mononuclear cells were challenged with the presence of the AGF_30 and AGF_40 concatemeric proteins (Fig. 8). Among the tests for preclinical immunogenicity assessment of the potential biological drugs are leukocyte/lymphocyte activation assay and basophil activation test [30]. PBMCs were incubated with the investigated concatemeric proteins (1.0 μg/ml) for 48 h and subsequently analysed with flow cytometry to assess the activation level of selected immune cell subpopulations. We analysed three crucial subsets of the immunocompetent cells: T helper cells – Th (CD3, CD4), T cytotoxic lymphocytes - CTL (CD3, CD8) and natural killer cells – NK (CD16, CD56) (Fig. 8). The expression of the following activation markers on each cell subpopulations was evaluated: CD69, CD71, CD25 and HLA-DR. The obtained data showed no activation of immune cells in the presence of two examined proteins. The levels of activation markers on all analysed immune cells subpopulations (CTL, Th, NK) were comparable to negative control regardless of the concatemeric protein used in the test, either the AGF_30 or AGF_40 (Fig. 8; Data in Brief, Fig. 10). The statistically significant decrease in PBMCs activation level was noted for cells stimulated with both proteins ($p < 0.05$) compared to the positive control (LPS/PHA) (Fig. 8). The BAT assay results showed no basophils activation in the presence of the investigated concatemeric proteins (1.0 μg/ml), as they reach even lower values than the negative control (not statistically significant) (Data in Brief, Fig. 11). The degranulation level of basophils challenged with both analysed proteins remained below or around the cut off value (5%) set by the kit manufacturer for the tests considered positive. The performed tests proved

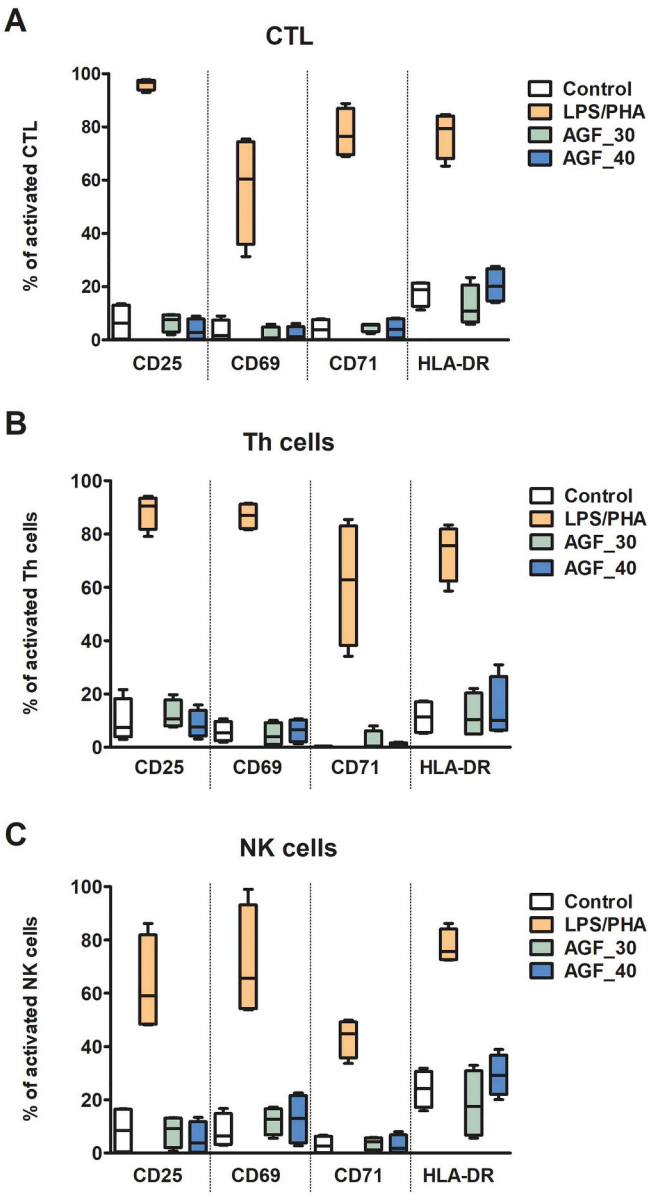

Fig. 8. *In vitro* immunogenicity assay was performed on human PBMCs exposed to the AGF-derived polyprotein variants: AGF_30 (TSRGDHELLGGAAPVGG_30) and AGF_40 (TSRGDHELLGGAAPVGG_40) for 48 h. The final protein concentration was 1 μg/ml. The analysis was performed with flow cytometry in order to evaluate the activation level of selected immune cell subpopulations. Panel A. Expression level of activation markers on CTLs (T cytotoxic lymphocytes). Panel B. Expression level of activation markers on Th cells (T helper cells). Panel C. Expression level of activation markers on NK cells (natural killer cells). Results were obtained from four independent experiments performed in quadruple and are presented as a median with min = max. p-value was evaluated with Mann-Whitney *U* test (in comparison to positive control). Untreated cells constituted a negative control, whereas cells stimulated with LPS (1.0 μg/ml, Lipopolysaccharide) and PHA (2.5 μg/ml, phytohemagglutinin) were a positive control.

that both investigated concatemeric proteins can be considered immunologically safe, as they do not induce immune cells response.

3.6. Stability of the AGF-derived concatemeric proteins in plasma

The stability of AGF_30 in human plasma was investigated by CE analysis. After optimizing CE separation conditions, the sample was analysed in a time-dependent manner and without any additional

preparation after mixing. However, because of the high viscosity of human plasma and its tendency to stick to the inner wall of the capillary, it was necessary to rinse the capillary with 0.1 M sodium hydroxide and separation buffer after each run. For the separation of sample ingredients, the separation buffer was enriched with SDS. Under the applied experimental conditions, a signal of plasma proteins was clearly separated from the AGF_30 signal. Therefore, the peak corresponding to AGF_30 could be easily assigned (Data in Brief, Fig. 12). AGF_30 was characterized by a longer migration time in comparison to plasma proteins. A progressive decrease of the AGF_30 peak (peak area) was observed. The half-life of AGF_30 in human plasma was estimated to be approximately 3 and a half hours. Complete degradation of the protein was observed after 21 h.

4. Discussion

Current chemical DNA synthesis methods have great difficulties in providing multimeric genes, containing repeats of a peptide coding DNA sequences. Most often, it is possible to synthesize genes with only a few repeats of such DNA fragment. In 'cell friendly' cases (such as the DNA fragment encoding HBV epitope) one can get up to 20–25 synthetic repeats.

In this paper we present a new, simple to use technology to produce artificial, repetitive genes encoding concatemeric RNAs and proteins of any nucleotide and aa sequence. The technology allows the acquisition of hundreds of DNA repeats within a concatemer. We have presented three approaches to construct long concatemer DNA. One of the approaches aimed at obtaining the largest possible monomer copy number. For this purpose, we used a pre-synthesized concatemer as a 'monomer', which was further subjected to amplification, resulting in much longer concatemers. The longest, commercially available synthetic concatemer was the 25-mer of 21-bp HBV epitope-coding DNA. Thus, we used this 'monomer' for further evaluation. The maximum length of the obtained concatemer corresponded to 500 copies of the monomer. Further increase was limited by several factors: (i) circularization of concatemers; (ii) vector capacity for accepting long DNA inserts and (iii) negative selection pressure of *E. coli* against long repetitive DNA molecules. Thus, we believe that we have reached the upper limit of concatemer length, using a combination of vectors and *E. coli* host described in this paper. Possibly, moving to other expression systems, with new DNA vectors, would allow one to push this limit further.

The key to the presented technology is: (i) a universal amplification vector module with two convergent DNA recognition sequences for a robust Type IIS REase *Sap*I (recognising long, 7 bp sequence and yielding cohesive 5′ ends of a codon length), (ii) coupled capabilities of high expression, fusion and secretion built in four versatile series of DNA vectors constructed. The vectors can be further modified to suit particular applications. Instead of *Sap*I, one can use its isoschizomers: *Bsp*QI (the recommended reaction temperature is 50 °C) or *Lgu*I (the recommended reaction temperature is 37 °C). However, according to our experience, *Sap*I appeared to be the most efficient enzyme for the technology. The designed *Sap*I restriction sites arrangement enables construction of the ORFs containing the desired number of head-to-tail repeats of any arbitrary DNA sequence. The DNA fragment subjected to oligomerization can either be obtained from a chemically synthetized origin or of a natural origin. Chemical synthesis approach allows also the development of co-polymeric, polyepitopic proteins. This can be obtained by joining the epitope-coding DNA segments close to each other. A source of the epitopes may be various proteins or fragments thereof. Co-polymeric, polyepitopic proteins can also be formed by random incorporation of the DNA fragments, equipped with *Sap*I cohesive ends. Such DNA fragments can be excised with *Sap*I from the previously obtained recombinant DNA constructs (peptide coding DNA fragments cloned into the *Sap*I amplification module of the amplification-expression vectors), purified, and self-ligated. Then, the resulted, new concatemeric genes can be recloned to the suitable amplification-expression vector. For maximum flexibility of the concatemeric gene expression options, we have designed and constructed four series of amplification-expression DNA vectors. For this purpose, we selected popular, commercially available expression vector backbones. Thus, potential end users, familiar with pACYC184 and pET vectors, can easily use the developed technology. The constructed vector series included: (1) a temperature-inducible *pR* promoter for cytoplasmic expression at 42 °C, which may help to maintain a polyepitopic protein solubility as well as improve transcription of GC-rich genes. Important feature of this vector series is its host independence. Both transcription promoter and its repressor gene are located within the same vector backbone; (2) IPTG inducible, T7-lac promoter for a versatile, wide temperature range and high-level expression of genes; (3) IPTG inducible, T7-lac promoter combined with a DNA cassette enabling ubiquitin-polyepitopic protein fusion. This vector series is dedicated to difficult aa sequences and recommended to be used when the concatemeric protein is insoluble, easily degraded or not effectively biosynthesized by recombinant bacterial host. The ubiquitin moiety can be removed by deubiquitinase [31,32] without leaving any additional aa residues in the polyepitopic protein; (4) IPTG inducible T7-lac promoter expression combined with the PhoA or MalE directed secretion of the polyepitopic proteins into periplasmic space. This strategy is useful for 'toxic' polyepitopic proteins and/or proteins requiring bisulfide bridges formation.

Moreover, all the vector series provide option for fusion of the polyepitopic protein N-terminus with His6 tag. As the pACYC184 vector derivatives are compatible with pET derivatives, they are suitable for simultaneous, independently controlled expression of two concatemeric genes in the same cell, if desired.

All vectors are constructed in such a way that *Sap*I excision/ligation/cloning cycles can be repeated, until a desired number of monomeric units is obtained within a constructed concatemeric gene. We also exemplified versatile use of the amplification scheme. The user can choose between two alternative variants of a substrate to be subjected for amplification: (i) a monomeric gene, coding for a single peptide or various peptides or (ii) chemically synthetized DNA (contain up to 25 repeats) used as 'monomer'. One should note that a capacity of the DNA vector determines the length of the resulted artificial gene. To overcome this limitation, the amplifying module can be easily transferred to any cloning or expression vector, if necessary. The maximum molecular weight of the obtainable, recombinant concatemeric proteins depends on: (i) the selected aa sequence of the monomeric peptide, (ii) limitation of the expression system and recombinant host used, and (iii) efficiency of a cellular transcription-translation machinery in selected host. For some applications, especially biomedical, the repeated aa segments offer an opportunity to use polyepitopic proteins as a sort of 'molecular dispenser', gradually releasing active mono- or oligomeric peptides. The release of the peptides can be conducted by naturally occurring human proteases. Human proteases of a lower specificity will fragment the concatemeric protein more or less randomly. Alternatively, several specific human proteases can be recruited to the peptide release simply by inclusion of a selected protease aa recognition sequence into a monomeric peptide unit. The proposed strategy of a therapeutic peptide delivery would be a clear advantage over using short, synthetic, bioactive peptides, which diffuse out rapidly from an application area and are quickly degraded by proteases. As shown in example development for the AGF-derived polyepitopic proteins, certain aa sequences are more prone to spontaneous hydrolysis or degradation by specialised proteases, for example human neutrophil elastase [33,34], providing alternative route for gradual release. This could be useful, for example for advanced therapy of wound healing. Regeneration of an injured tissue would be controlled by a gradual release of biologically active peptides by matrix metalloproteinases or serine proteases. The concatemeric proteins could also be immobilized on macromolecular carriers (such as microorganisms, cells, bacteria,

bacteriophages, viruses, defective virions, auto-aggregating proteins, or nanoparticles) to enhance their activity or to suit a specialised application. The immobilization may be performed using genetic or chemical means. Such a strategy may strengthen their potential therapeutic or biotechnological effect. It is assumed that proteins have a greater risk of causing an immunologic reaction in comparison to smaller chemical particles. Immunogenicity as well as allergies to biopharmaceuticals affect not only the safety of the patient (various types of side effects), but also the effectiveness of the therapy and the economic aspects of introducing new biodrugs [35,36]. Therefore, pre-clinical research of the assessment of the immunological safety constitutes an essential medical, pharmaceutical and economic issue [30]. Lack of immunogenicity of new AGF polyepitopic proteins as well as a low allergic potential and a low cytotoxicity presented in the study are advantages of the presented method of obtaining such artificial proteins. Pro-proliferative and pro-migratory effects show that AGF polyepitopic protein can be taken into consideration as a potential wound healing stimulant. The AGF and other polyepitopic proteins (constructed using the technology described in this paper), aimed at tissue regeneration, have been covered by European and Republic of Poland patent applications (EP18000319.6 and P.425131, respectively) and this investigation is being continued. In order to confirm the safe use of any other concatemeric protein obtained with the presented method, similar biological studies would be required.

4.1. Conclusions

A new DNA fragment amplification technology was developed for the production of artificial, repetitive genes, encoding concatemeric RNAs and proteins of any nucleotide and aa sequence. The amplification reaction is linear. The technology allows for the formation of ordered polymers, containing 500 or more copies of DNA, RNA or peptide repeats within a concatemer.

Four series of amplification-expression vectors were constructed for specialised applications.

The amplification of a DNA fragment, encoding a peptide with a biological or chemical function can lead to great enhancement in desirable interaction of the resulting (poly)peptide with a specific ligand. Such concatemeric proteins can be a basis for construction of: (i) artificial antigens - a new generation of vaccines with a magnified potential stimulation of the immune system; (ii) concatemeric proteins containing modules for rare and/or toxic metal ions chelation for their industrial production, environmental remediation or an organism detoxification; (iii) high capacity reservoirs for necessary enzyme cofactors (such as cations, anions, organic molecules) that can regulate or inhibit a particular enzymatic activity; (iv) reservoirs for peptide hormones; (v) reservoirs for developed bioactive peptides or derivative from the signalling proteins, which can stimulate tissue regeneration; (vi) protective, therapeutic concatemeric proteins, containing peptide activators or inhibitors of biological functions, for treatment of molecular, viral and bacterial diseases; (vii) poly-micro RNA, poly-antisense nucleic acids for the treatment of genetic, molecular, viral and bacterial diseases.

Polyepitopic proteins, composed of the HBV S protein epitope and AGF peptide, were constructed. The HBV derivative is a prototype of a novel type vaccine, whereas the AGF derivative is a prototype of a novel type of pro-regenerative biological drug, additionally acting as a 'molecular dispenser'. AGF_30 is relatively stable in human plasma ($t1/2 = 3.5$ h). This shows that concatemerization of the bioactive peptide (TSRGDHELL) may increase its stability in human fluids. The effect of progressive degradation of AGF_30 shows its ability to gradually release bioactive peptides. However, the identity of degradation products has not been established so far.

Accession numbers

GenBank numbers: MK651654, MK606505, MK606506, MK606507, MK606508, MK606509, MK606510, MK606511, MK606512, MK606513, MK606514, MK606515, MK606516, MK606517, MK606518, MK606519, MK606520, MK606521, MK606522, MK606523, MK606524, MK659889, MK606525, MK606526, MK606527.

Supplementary data

Supplementary data are available as Data in Brief.

Authors' contributions

PMS conceived the project, coordinated its execution and designed the DNA vectors, amplification modules and obtained supporting grants and other funds. PMS, AZS, MP (Michal Pikula), MW, SRM, PS, AC, AP, PM, LJ participated in the design and interpretation of the experimental analyses. PM and PS designed the AGF-derived peptide. JZ expressed and purified the AGF-derived protein variants. AZS, OZ prepared a series of pAMP DNA vectors and cloned the DNA fragment encoding single HBV epitope. JFF designed and conducted site-specific mutagenesis of pET21d(+) and pET28a(+). NK constructed pET28AMP-PhoA and pET28AMP-MalE DNA vectors. MP (Malgorzata Palczewska) designed and constructed pET28AMP_*Sap*I-Ubq DNA vector. AS performed proliferation, cytotoxicity, migration and chemotaxis experiments. AW designed and performed the immunological experiments, analysed and interpreted the immunological data. MD performed proliferation and cytotoxicity experiments, collected and analysed data. MP (Michal Pikula) conceived and supervised cell culture experiments and provided financial support for these experiments. OZ performed the first round of amplification and preliminary expression experiments concerning the 13-mer of the HBV epitope. NK performed the second round of amplification, expressed the resultant concatemeric genes. MN purified the 10–30 mers. AZS, DK performed the alternative route of amplification, expressed all the resultant, concatemeric genes and purified the 25-mer of the HBV epitope. AZS and PMS prepared the manuscript; MP (Michal Pikula), AS, AW, MD prepared the section describing the cell culture experiments, MW contributed to Introduction and Material and methods section writing. All the authors read and approved the final manuscript.

Acknowledgements

We thank Prof. Miroslawa Cichorek (Department of Embryology, Medical University of Gdansk), Prof. Janusz Morys (Chairperson of Anatomy, Medical University of Gdansk, Poland) and Prof. Piotr Trzonkowski (Department of Medical Immunology, Medical University of Gdansk) for granting access to the equipment needed. We would also like to thank Maciej Zielinski, PhD and the personnel of the Laboratory of Immunology and Clinical Transplantology from The University Clinical Centre in Gdansk for carrying out BAT assays. We thank Agnieszka Ozog for excellent project administration. Marta Skowron Volponi is highly appreciated for critical reading of the manuscript and English corrections.

Funding

This work was supported by National Centre for Research and Development – Poland [grant numbers POIG.01.04.00-22-140/12 to BioVentures Institute Ltd. (Poznan, Poland), STRATEGMED1/235077/9/NCBR/2014 to University of Gdansk, Faculty of Chemistry, Molecular Biotechnology Department] and BioVentures Institute Ltd. Funding for open access charge: University of Gdansk.

Declaration of competing interest

None declared. The technology is owned by BioVentures Institute Ltd. (Poland) and has been protected by Polish patent no 228341 [17] (filed in 2014), PCT patent application filed in 2015 and the global patent applications [18] are currently being processed.

Data availability

The raw data required to reproduce these findings are available as Data in Brief. The processed data required to reproduce these findings are available as Data in Brief.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.msec.2019.110426.

References

[1] K. Fosgerau, T. Hoffmann, Peptide therapeutics: current status and future directions, Drug Discov. Today 20 (1) (2015) 122–128.

[2] J. Lau, M. Dunn, Therapeutic peptides: historical perspectives, current development trends, and future directions, Bioorganic and Medicinal Chemistry 26 (10) (2018) 2700–2707.

[3] D. Chow, M.L. Nunalee, D.W. Lim, A.J. Simnick, A. Chilkoti, Peptide-based biopolymers in biomedicine and biotechnology, Materials Science and Engineering. R, Reports: A Review Journal 62 (4) (2008) 125–155.

[4] H. Lee, D. Piao, J. Lee, J. Choi, J. Bok, C. Cho, S. Kang, Y. Choi, Artificially designed recombinant protein composed of multiple epitopes of foot-and-mouth disease virus as a vaccine candidate, Microb. Cell Factories 16 (1) (2017) 33, https://doi.org/10.1186/s12934-017-0648-2.

[5] T. Kempe, S. Kent, F. Chow, S. Peterson, W. Sundquist, J. L'Italien, D. Harbrecht, D. Plunkett, W. DeLorbe, Multiple-copy genes: production and modification of monomeric peptides from large multimeric fusion proteins, Gene 39 (2–3) (1985) 239–245.

[6] S. Kim, W. Szybalski, Amplification of cloned DNA as tandem multimers using BspMI-generated asymmetric cohesive ends, Gene 71 (1) (1988) 1–8.

[7] J. Lee, P. Skowron, S. Rutkowska, S. Hong, S. Kim, Sequential amplification of cloned DNA as tandem multimers using class-IIS restriction enzymes, Genetic Analysis: Biomolecular Engineering 13 (6) (1996) 139–145.

[8] J. Lee, I. Minn, C. Park, S. Kim, Acidic peptide-mediated expression of the antimicrobial peptide buforin II as tandem repeats in *Escherichia coli*, Protein Expr. Purif. 12 (1) (1998) 53–60.

[9] Y. Wang, J. Cai, High-level expression of acidic partner-mediated antimicrobial peptide from tandem genes in *Escherichia coli*, Appl. Biochem. Biotechnol. 141 (2–3) (2007) 203–213.

[10] M. Lennick, J.R. Haynes, S.H. Shen, High-level expression of alpha-human atrial natriuretic peptide from multiple joined genes in *Escherichia coli*, Gene 61 (1) (1987) 103–112.

[11] J.R. McDaniel, J.A. Mackay, F.G. Quiroz, A. Chilkoti, Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes, Biomacromolecules 11 (4) (2010) 944–952.

[12] J. Won, A. Barron, A new cloning method for the preparation of long repetitive polypeptides without a sequence requirement, Macromolecules 35 (22) (2002) 8281–8287.

[13] S. Shen, Multiple joined genes prevent product degradation in *Escherichia coli*, Proc. Natl. Acad. Sci. 81 (15) (1984) 4627–4631.

[14] M. Li, D. Shukla, B. Rhodes, A. Lall, J. Shu, B. Moriarity, D. Largaespada, HomeRun Vector Assembly System: a flexible and standardized cloning system for assembly of multi-modular DNA constructs, PLoS One 9 (6) (2014) e100948, https://doi.org/10.1371/journal.pone.0100948 (eCollection 2014).

[15] D. Meyer, A. Chilkoti, Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system, Biomacromolecules 3 (2) (2002) 357–367.

[16] H. Chu, J. Ryum, S. Park, B. Kim, D. Kim, J. Won, A new cloning strategy for generating multiple repeats of a repetitive polypeptide based on non-template PCR, Biotechnol. Lett. 33 (5) (2011) 977–983.

[17] Skowron, P., Żylicz-Stachula, A., Żolnierkiewicz, O., Skowron, M., Janus, L., Jeżewska-Frąckowiak, J. and Szymańska, A. (2018) Patent PL 228341-B1.

[18] Skowron, P, Zylicz-Stachula, A, Zolnierkiewicz, O, Skowron, M, Janus, L, Jezewska-Frackowiak, J, Krefft, D, Nidzworski, D., Szemiako, K., Maciejewska, N., Nowak, M., Szymanska, A. (2015) PCT Patent Application WO 2015162560 A1.

[19] A.C. Chang, S.N. Cohen, Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid, J. Bacteriol. 134 (1978) 1141–1156.

[20] Studier, F.W., Davanloo, P., Rosenberg, A.H., Moffatt, B.A. and Dunn, J.J. (1986) Patent US 07002408.

[21] D. Kunikowska, B. Dera-Tomaszewska, H. Dziadziuszko, J. Wysocki, I. Wolańska, Z. Maćkiewicz, R. Głośnicka, G. Kupryszewski, Enantiomers of hepatitis B virus surface antigen (140–146) S region fragment: synthesis and studies on immunological properties, Polish J Chem. 73 (1999) 1803–1809.

[22] M.R. Green, J. Sambrook, Molecular Cloning: A Laboratory Manual, 4th edn, Cold Spring Harbor Laboratory Press, NY, 2012.

[23] P. Boukamp, R.T. Petrussevska, D. Breitkreutz, J. Hornung, A. Markham, N.E. Fusenig, Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line, JCB 106 (3) (1988) 761–771.

[24] L.M. Henderson, C.F. Arlett, S.A. Harcourt, A.R. Lehmann, B.C. Broughton, Cells from an immunodeficient patient (46BR) with a defect in DNA ligation are hypomutable but hypersensitive to the induction of sister chromatid exchanges, PNAS 82 (7) (1985) 2044–2048.

[25] S. Lundin, A. Jemt, F. Terje-Hegge, N. Foam, E. Pettersson, M. Käller, V. Wirta, P. Lexow, J. Lundeberg, Endonuclease specificity and sequence dependence of type IIS restriction enzymes, PLoS One 10 (1) (2015) e0117059, https://doi.org/10.1371/journal.pone.0117059 (eCollection 2015).

[26] J.A. Beeley, D. Sweeney, J.C. Lindsay, M.L. Buchanan, L. Sarna, K.S. Khoo, Sodium dodecyl sulphate-polyacrylamide gel electrophoresis of human parotid salivary proteins, Electrophoresis 12 (1991) 1032–1041.

[27] M.G. Humphreys-Beher, D.J. Wells, Metachromatic staining patterns of basic proline-rich proteins from rat and human saliva in sodium dodecyl sulfate-polyacrylamide gels, J. Appl. Biochem. 6 (5–6) (1984) 353–360.

[28] Y. Oike, K. Yasunaga, Y. Ito, S. Matsumoto, H. Maekawa, T. Morisada, F. Arai, N. Nakagata, M. Takeya, Y. Masuho, T. Suda, Angiopoietin-related growth factor (AGF) promotes epidermal proliferation, remodeling, and regeneration, Proc. Natl. Acad. Sci. U. S. A. 100 (16) (2003) 9494–9499.

[29] H. Okazaki, S. Hirakawa, M. Shudou, Y. Nakaoka, Y. Shirakata, K. Miyata, Y. Oike, K. Hashimoto, K. Sayama, Targeted overexpression of Angptl6/angiopoietin-related growth factor in the skin promotes angiogenesis and lymphatic vessel enlargement in response to ultraviolet B, J. Dermatol. 39 (4) (2012) 366–374.

[30] M. Deptuła, A. Wardowska, M. Dzierzynska, S. Rodziewicz-Motowidło, M. Pikuła, Antibacterial peptides in dermatology-strategies for evaluation of allergic potential, Molecules 23 (2) (2018), https://doi.org/10.3390/molecules23020414.

[31] A. Wojtowicz-Krawiec, I. Sokolowska, M. Smorawinska, L. Chojnacka-Puchta, D. Mikiewicz, N. Lukasiewicz, A. Marciniak-Rusek, R. Wolinowska, A. Bierczynska-Krzysik, A.J. Porebska, J. Kuthan-Styczen, L. Gurba, P. Borowicz, A. Mazurkiewicz, G. Plucienniczak, A. Plucienniczak, Use of Ubp1 protease analog to produce recombinant human growth hormone in *Escherichia coli*, Microb. Cell Factories 13 (1) (2014) 113, https://doi.org/10.1186/s12934-014-0113-4.

[32] H.A. Yu, S.G. Kim, E.J. Kim, W.J. Lee, D.O. Kim, K. Park, Y.C. Park, J.H. Seo, Characterization of ubiquitin C-terminal hydrolase 1 (YUH1) from *Saccharomyces cerevisiae* expressed in recombinant *Escherichia coli*, Protein Expr. Purif. 56 (2007) 20–26.

[33] O. Wiesner, R.D. Litwiller, A.M. Hummel, M.A. Viss, C.J. McDonald, D.E. Jenne, D.N. Fass, U. Specks, Differences between human proteinase 3 and neutrophil elastase and their murine homologues are relevant for murine model experiments, FEBS Lett. 579 (24) (2005) 5305–5312.

[34] A.A. Aimetti, M.W. Tibbitt, K.S. Anseth, Human neutrophil elastase responsive delivery from poly(ethylene glycol) hydrogels, Biomacromolecules 10 (6) (2009) 1484–1489, https://doi.org/10.1021/bm9000926.

[35] K.D. Ratanji, J.P. Derrick, R.J. Dearman, I. Kimber, Immunogenicity of therapeutic proteins: influence of aggregation, J. Immunotoxicol. 11 (2) (2014) 99–109.

[36] A. Kuriakose, N. Chirmule, P. Nair, Immunogenicity of biotherapeutics: causes and association with posttranslational modifications, J Immunol Res 2016 (2016) 1298473, https://doi.org/10.1155/2016/1298473.