



OPEN

Expression and purification of a native Thy1-single-chain variable fragment for use in molecular imaging

Natacha Jugniot, Rakesh Bam & Ramasamy Paulmurugan

Molecular imaging using singlechain variable fragments (scFv) of antibodies targeting cancer specific antigens have been considered a non-immunogenic approach for early diagnosis in the clinic. Usually, production of proteins is performed within *Escherichia coli*. Recombinant proteins are either expressed in *E. coli* cytoplasm as insoluble inclusion bodies, that often need cumbersome denaturation and refolding processes, or secreted toward the periplasm as soluble proteins that highly reduce the overall yield. However, production of active scFvs in their native form, without any heterologous fusion, is required for clinical applications. In this study, we expressed an anti-thymocyte differentiation antigen-scFv (Thy1-scFv) as a fusion protein with a N-terminal sequence including 3 × hexa-histidines, as purification tags, together with a Trx-tag and a S-tag for enhanced-solubility. Our strategy allowed to recover ~ 35% of Thy1-scFv in the soluble cytoplasmic fraction. An enterokinase cleavage site in between Thy1-scFv and the upstream tags was used to regenerate the protein with 97.7 ± 2.3% purity without any tags. Thy1-scFv showed functionality towards its target on flow cytometry assays. Finally, in vivo molecular imaging using Thy1-scFv conjugated to an ultrasound contrast agent (MB_{Thy1-scFv}) demonstrated signal enhancement on a transgenic pancreatic ductal adenocarcinoma (PDAC) mouse model (3.1 ± 1.2 a.u.) compared to non-targeted control (0.4 ± 0.4 a.u.) suggesting potential for PDAC early diagnosis. Overall, our strategy facilitates the expression and purification of Thy1-scFv while introducing its ability for diagnostic molecular imaging of pancreatic cancer. The presented methodology could be expanded to other important eukaryotic proteins for various applications, including but not limited to molecular imaging.

Molecular imaging techniques play a central role in clinical oncology by enhancing diagnostic and therapeutic approaches. Ultrasound (US) molecular imaging (USMI) is a recent modality with greatly improved detection accuracy compared to conventional US. When using molecularly targeted US contrast agents, USMI has shown combined advantages of US modality and molecular imaging. With the recent first molecular US contrast agent to enter clinical trials, innovation in human cancer imaging by USMI lies ahead. Targeted US contrast agents called ‘microbubbles’ (MBs) are gas-filled microparticles synthesized by conjugating specific ligands onto the MB shell, making them to bind tumor neovascular targets thus enabling enhanced-tumor detection. However, producing targeted-ligands as recombinant proteins/antibodies without heterologous tags is needed to generate non-immunogenic agents for translational applications.

With the introduction of recombinant DNA technology in 1974¹, faster, easier, and more efficient production of heterologous proteins have been performed compared to purifying proteins from their natural sources. Nowadays, diverse expression systems such as bacteria, yeast, insect cells, mammalian cells, cell-free systems, transgenic animals and plants have been used for the expression of recombinant proteins. Choice in host system plays a critical role not only in the expression of the protein of interest but also in the way it can be subsequently purified².

Escherichia coli remains the most attractive host for recombinant protein expression^{3,4}. However, as a prokaryotic system, *E. coli* may not be able to produce some eukaryotic proteins in their native form⁵. To overcome such issues, strategies include the use of mutated host strains, mRNA enhanced-stability, use of optimal codons,

Department of Radiology, Molecular Imaging Program at Stanford (MIPS), Canary Center for Cancer Early Detection at Stanford, Stanford University School of Medicine, Stanford University, 3155 Porter Drive, Palo Alto, CA 94304, USA. ✉ email: paulmur8@stanford.edu

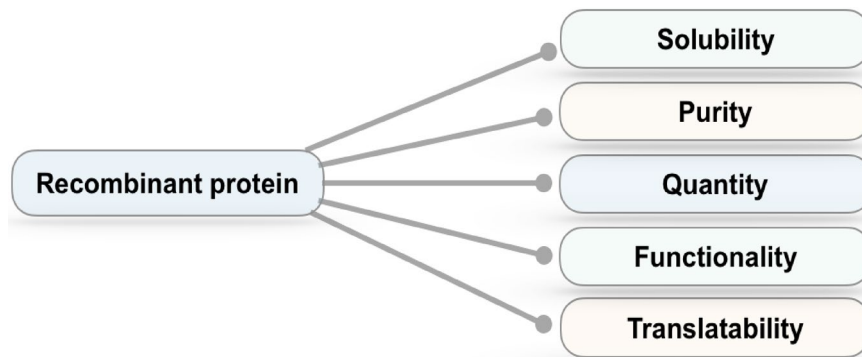


Figure 1. Recombinant protein checkpoints.

co-expression of molecular chaperones such as foldases and post-translational protein modifying enzymes, as well as optimization of growth conditions^{6,7}. Despite plethora of methodologies available, challenges remain in providing recombinant proteins with relevant quantity, purity, solubility, functionality, and translatability toward clinical applications (Fig. 1). Periplasmic protein expression is considered a favorable approach for disulfide bond formation in *E. coli*, and was initially the most frequently used methodology⁸. Alternatively, expressing proteins in the cytoplasm leads to much higher expression levels but results in aggregation into insoluble inclusion bodies (IBs)^{9,10}. Although useful in many cases where large amounts of proteins are needed, denaturation/refolding protocols of IBs necessitate the use of strong denaturants and reducing agents¹¹ which can lead to improper refolding and might not give full recovery of the protein with its biological activity¹². On the other hand, a wide range of fusion partners with solubilization tags (e.g., thioredoxin (Trx), poly(NANP) (N-acetylneuraminic acid phosphatase), S-tag) are now available for enhanced expression of cytosoluble recombinant proteins, and affinity tags (e.g., maltose binding protein (MBP), glutathione-S-transferase (GST), hexa-histidine tag) for more efficient purification processes^{13,14}. Although tag sequences are essential for the expression and purification of recombinant proteins, they can interfere with the structure and function of their fusion partner while limiting their application in the clinic due to immunogenicity. Therefore, tag removal should be considered, especially if the protein of interest is intended for clinical applications or structural studies¹⁵. Hence, developing protocols for the recombinant production of eukaryotic proteins in their native form is important.

With the goal of developing US contrast agents for pancreatic ductal adenocarcinoma (PDAC) imaging, we chose to express a single-chain variable fragment (scFv) targeting the thymocyte differentiation antigen (Thy1/CD90)¹⁶. Thy1 is a cell-surface glycoprotein originally described as a mouse thymocyte differentiation marker. Subsequently, Thy1 has been shown to be expressed in other tissues, including overexpression in the neovasculature of various human cancers (e.g., colon cancer¹⁷, glioblastoma¹⁸, hepatocellular carcinoma¹⁹, ovarian cancer²⁰, prostate cancer²¹ and PDAC^{22,23}). Proteomic and immunohistochemical analysis in human tissue samples revealed Thy1 as a molecular marker upregulated on the neovasculature of PDAC^{22,23}. Importantly, 81% of PDAC cases stained positive for Thy1 while in normal pancreas and chronic pancreatitis cases, values were 11% and 7%, respectively. This has suggested a low background pattern of Thy1 in other tissues compared to PDAC, thus suitable for USMI. Overall, Thy1 neovascular immunostaining could distinguish PDAC from normal and chronic pancreatitis tissues with 90% specificity and 81% sensitivity. We have previously engineered an anti-Thy1-scFv (or simply called “Thy1-scFv”) through directed evolution of scFv protein scaffold using a yeast surface display library²⁴. Here, we further engineer Thy1-scFv as a fusion protein with a N-terminal sequence including hexa-histidine tandems (1, 3 or 5 hexa-histidine tags) as purification tags together with a Trx-tag and a S-tag for improving the solubility, and one enterokinase (EK) cleavage site at the junction of Thy1-scFv and the upstream tags. We choose the extensively used Δ gor, Δ trx_B, DsbC⁺ SHuffle T7 *E. coli* system to express the engineered protein with disulfide bonds in the cytoplasm. The method we describe here stands out for its relative simplicity to produce Thy1-scFv using complementary tag sequences, which, when cleaved by EK, resulted in generating the native Thy1-scFv without any tag (Fig. 2). Evaluation of native Thy1-scFv showed successful receptor binding capacity *in vitro*. *In vivo* results illustrated the conserved functionality of Thy1-scFv enabling PDAC molecular imaging by US. The presented method could be beneficial at the outset of many projects implying eukaryotic protein production and could be optimized toward translational applications.

Results

Engineering, expression and purification of Thy1-scFv. *Construction of vectors expressing Thy1-scFv gene for efficient soluble protein production.* We constructed three different expression vectors in the popular pET32b vector backbone with Trx- and S-tags and additional tandem His-Tags, giving: pET32b-1XHis-scFv, pET32b-3XHis-scFv, and pET32b-5XHis-scFv (Fig. 3a). Sequence confirmed, clones were transformed into T7 SHuffle *E. coli* cells to investigate the effect of each variant on the expression and purification of Thy1-scFv. In all constructs, the tagged-Thy1-scFv has a theoretical molecular weight around 50 kDa.

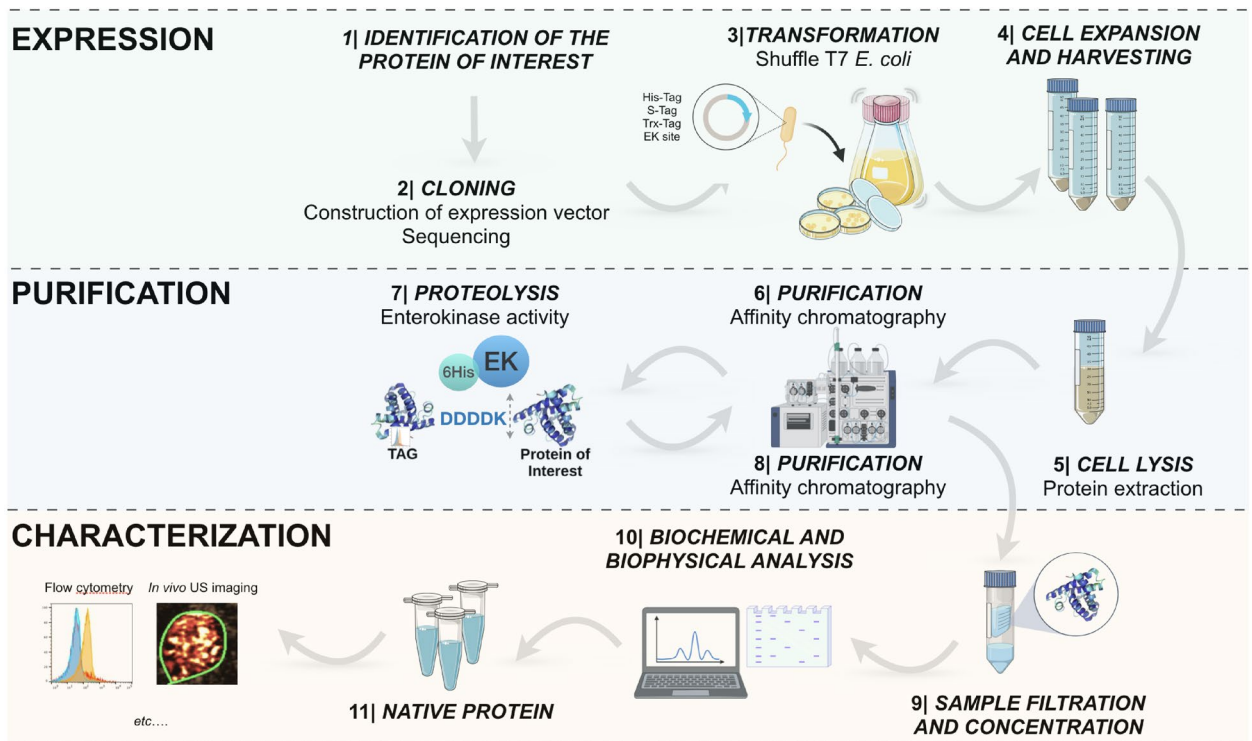


Figure 2. Schematic workflow showing the methodology employed for the construction, expression and purification of recombinant protein in its native form.

Optimization of conditions for Thy1-scFv expression and purification by affinity chromatography. Protein expression in *E. coli* was induced at 30 °C (Fig. 3b; Supplementary Figure S1) since higher temperatures usually result in a rapid decrease in protein yield due to degradation and misfolding²⁵. As expected, at 37 °C, the expression rate of Thy1-scFv was very low (Supplementary Figure S2), in contrast, we observed an increase in our target protein yield when cultured at 30 °C probably because such lower temperature reduces protein degradation, improves folding efficiency, and thus reduces IB formation. Moreover, endogenous proteases have a higher turnover rate when *E. coli* is grown at 37 °C, thus leading to an enhanced-proteolysis of Thy1-scFv into a Thy1-scFv fragment (around 25 kDa). After purification by affinity chromatography, we detected Thy1-scFv in the elution fractions at an apparent molecular weight consistent with its theoretical mass in all three constructs. Importantly, culture expressing pET32b-3XHis-scFv with an induction temperature of 30 °C showed higher protein level compared to the other constructs (Fig. 3b, pET32b-3XHis-scFv, lanes 6 to 8), and Thy1-scFv was isolated with $47.7 \pm 11.5\%$ purity. This is in part due to higher binding of Ni-agarose with the 3X-His-tag used for purification. Apart from the full-length Thy1-scFv, a fragment of around 25 kDa was the major co-purified protein. We suspected it to come from residual endogenous proteolytic activity on Thy1-scFv on a locally weaker structure probably in the multi-histidine tag region. After transfer onto nitrocellulose membranes, an anti-hexa-histidine tag antibody was used to confirm the presence of Thy1-scFv. Analysis of both soluble and insoluble fractions indicates that $\sim 35\%$ of the protein can be recovered in the soluble fraction (Supplementary Figure S3). The tagged-scFv from pET32b-1XHis-scFv bound the column with weak affinity and therefore was co-purified with many contaminants. The use of pET32b-5XHis-scFv generated a longer and flexible tag sequence, more prone to influence protein binding sites and improper column binding, as suggested by the affinity chromatography elution profile. Based on these results, the construct pET32b-3XHis-scFv was chosen for further experiments presented in this study. Elution fractions 6–8 from this construct were pooled and used for further enrichment. Elimination of imidazole salts resulted in a more stable protein. The sample was concentrated (Fig. 4, undigested T_0) yielding to 0.37 ± 0.15 mg of tagged-Thy1-scFv per liter of bacterial culture (Table 1), difficult to reach using other methods of purification.

Optimization of tag removal by EK activity. To optimize the hydrolysis efficacy of EK on the fusion protein, 30 μ g of the purified tagged-protein were incubated with 0.5, 1, 2 and 8 Unit(s) of EK at 25 °C for different times (1, 4, 8 and 24 h(s)) (Fig. 4). Incubation of the tagged-Thy1-scFv with EK resulted in the generation of native scFv (~ 30 kDa; expected molecular weight 27,864.93 Da) and free tags (~ 17 kDa). Product bands appeared in all conditions after 1 h incubation along with undigested substrate. After 4 h, the substrate was almost fully digested (approximately 80% of full-length fusion protein was cleaved with 1, 2 and 8 U, Supplementary Figure S4a). However, an incubation time > 4 h or the use of ≥ 8 U of EK during a time > 1 h triggered non-specific cleavage of Thy1-scFv (Supplementary Figure S4b) and was also not cost effective. No spontaneous hydrolysis was detected after 24 h of incubation.

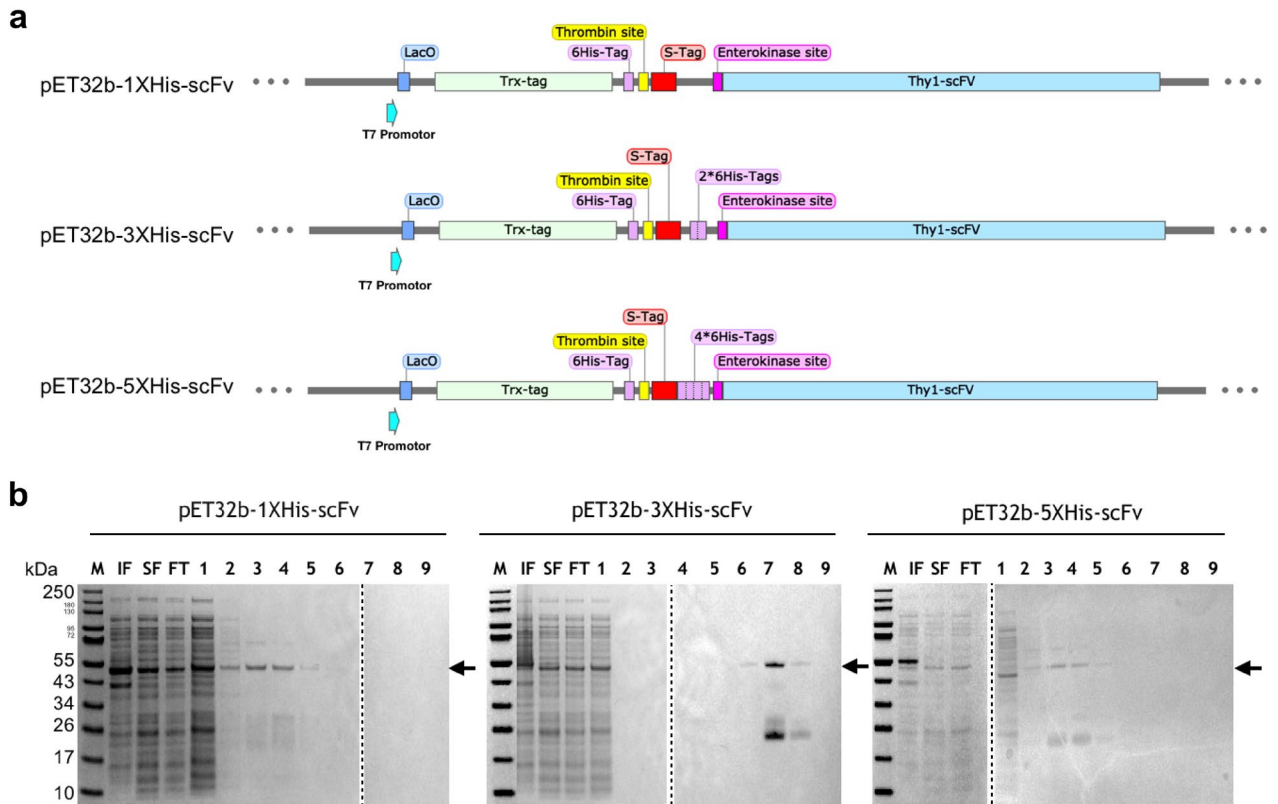


Figure 3. Construction and expression of recombinant Thy1-scFv proteins. (a) Schematic map showing the cloning pattern of pET-32b vector constructed to express recombinant Thy1-scFv in various formats (different number of hexa-histidine tags): pET32b-1XHis-scFv, pET32b-3XHis-scFv, and pET32b-5XHis-scFv. (b) Purity of protein based on the elution fractions after purification by affinity chromatography: Thy1-scFv formats resolved in 4–12% gradient SDS-PAGE. *M* protein molecular weight marker, *IF* insoluble fraction from cell lysate, *SF* soluble fraction from cell lysate, *FT* flow through, lanes 1–9: eluted fractions; arrows indicate the position of the tagged-Thy1-scFvs. The same methodology was applied for all Thy1-scFv formats and gels were processed in parallel. Dotted lines have been used to delineate different gels. Full-length gels are presented in Supplementary Figure S1.

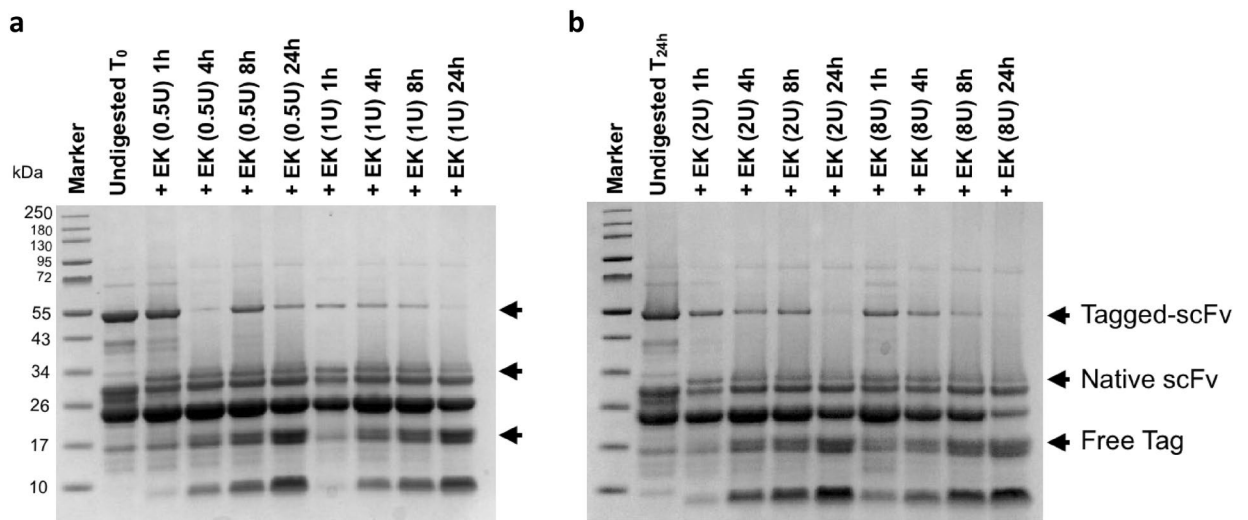


Figure 4. Optimization of EK incubation time and concentrations from (a) 0.5U and 1U to (b) 2U and 8U, for tag removal of the recombinant Thy1-scFv protein. 4–12% gradient SDS-PAGE were processed in parallel.

Purification step	Volume (mL)	[Proteins] _{total} (mg/mL)	Total proteins (mg)	Tagged-Thy1-scFv purity (%) ^a	Native Thy1-scFv purity (%) ^a	Tagged-Thy1-scFv (mg)	Native Thy1-scFv (mg)
Cell lysate (n=5)	20	27.2 ± 4.6	543 ± 92	–	–	–	–
IMAC1 (n=5)	1.1 ± 0.3	0.69 ± 0.1	0.78 ± 0.2	47.7 ± 11	–	0.37 ± 0.1	–
IMAC2 (n=3)	1.2 ± 1.1	0.15 ± 0.04	0.23 ± 0.1	–	97.7 ± 2.3	–	0.22 ± 0.1

Table 1. Summary of Thy1-scFv purification from 1 L bacterial culture. ^aValue based on analysis of Thy1-scFv band intensity using BioRad Gel-Doc system.

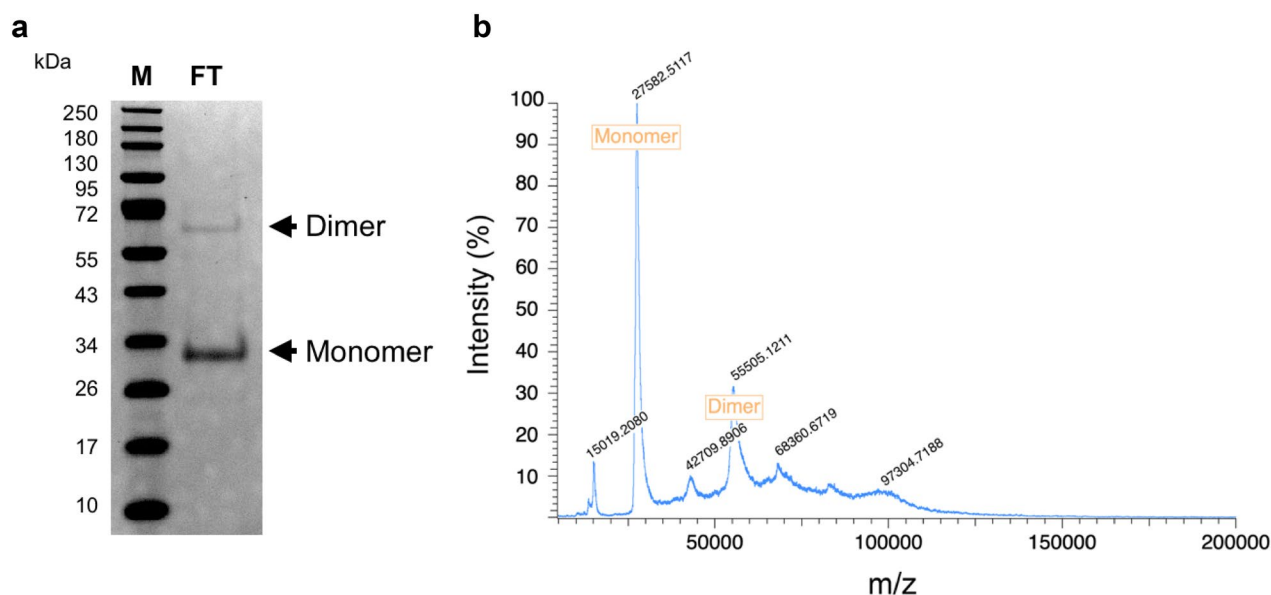


Figure 5. Second step purification of Thy1-scFv after proteolysis. (a) SDS-PAGE showing pure Thy1-scFv. *M* protein molecular weight marker, *FT* flow through. (b). Mass spectrometry on Thy1-scFv flow through fraction.

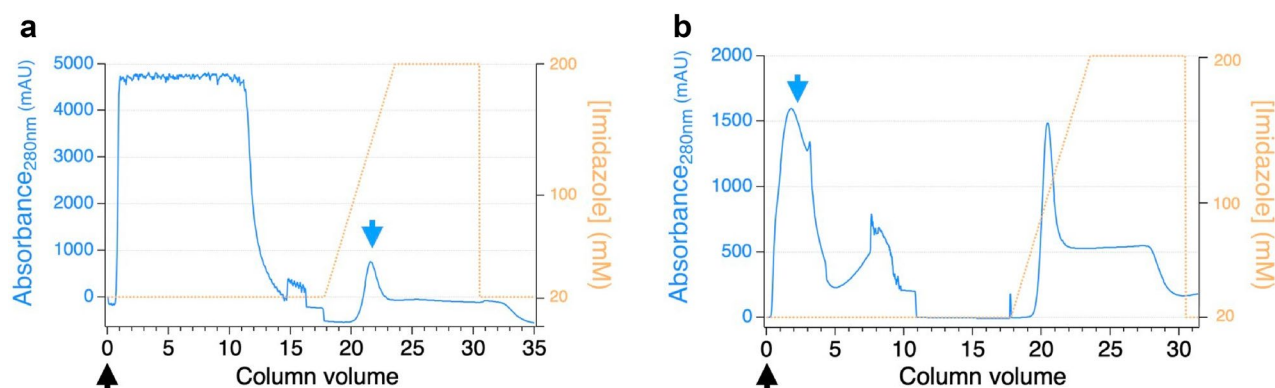


Figure 6. Elution profile of Thy1-scFv on Ni-NTA IMAC. (a) IMAC1: tagged-Thy1-scFv is collected during the elution step. (b) IMAC2: native Thy1-scFv is collected in the flow through fraction. Black arrows represent the time of injection, and blue arrows indicate the peak and elution time of the protein of interest.

Native Thy1-scFv purification. Following a second immobilization by metal affinity chromatography, 0.22 ± 0.11 mg of scFv can be recovered with 97.7% purity as demonstrated by gel electrophoresis (Fig. 5a) and mass spectrometry (Supplementary Figure S5). Further analysis also highlighted a preponderant monomer fraction with a small fraction of dimer (Fig. 5b). The monomer/dimer relative intensity ratio was 10:2 by SDS-PAGE and 10:3 by MALDI analyses. Contrary to the first round of purification (Fig. 6a) where the recombinant protein was recovered in the latest elution fractions, the native protein was isolated in the flow-through fraction (Fig. 6b). Recombinant EK, uncleaved recombinant Thy1-scFv, and fragments containing histidine tags were retained on the column.

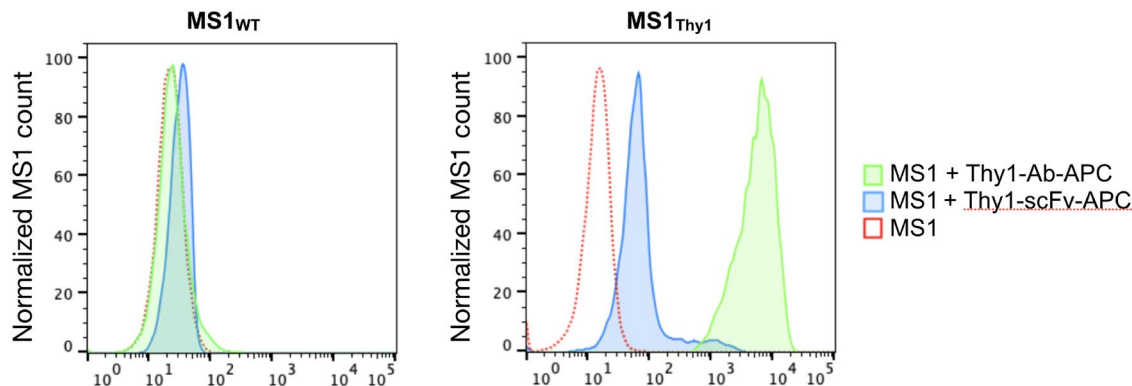


Figure 7. Binding efficiency of Thy1-scFv to MS1 cells.

Thy1-scFv binds to Thy1-expressing cells in vitro. APC conjugated commercial antibody (Thy1-Ab-APC) was used as positive control to confirm cell-surface expression of Thy1 (Fig. 7). Binding capacity of Thy1-scFv-APC was performed against both Thy1 expressing cells (MS1_{Thy1}) and control cells (MS1_{WT}) by flow cytometry. Incubation of cells with Thy1-scFv-APC showed higher binding to MS1_{Thy1} compared to the controls as detected by a shift toward increased fluorescent signal intensity. Thus, Thy1-scFv retained its ability to bind to its target cells in vitro.

In vivo US contrast agents functionalized with Thy1-scFv enhance PDAC molecular imaging in transgenic animal model. To test the functionality of Thy1-scFv in vivo, we used US contrast agents (MBs) covalently bound with Thy1-scFv on their surface using NHS-chemistry, thus giving MB_{Thy1-scFv}. Control MBs, without ligand, were reported as MB_{non-targeted}. Prior to imaging, both MB types were tested for size and concentration changes that may occur due to Thy1-scFv's influence on steric changes or octafluoropropane (C₃F₈) gas dissipation.

Incorporation of Thy1-scFv as part of the MB composition, *i.e.*, MB_{Thy1-scFv}, did not significantly affect the MB mean diameter or size distribution compared to MB_{non-targeted} (mean diameter = 1.2 ± 1.4 μm and 1.1 ± 0.8 μm, respectively) nor the concentration (1.10⁹ particles/mL and 1.2.10⁹ particles/mL, respectively) in agreement with MBs used in the clinic (e.g., Definity (Lantheus Medical Imaging), Sonovue (Bracco Diagnostics)) (Supplementary Figure S6). Tumors were located on B-mode imaging and MBs were injected intravenously starting with MB_{non-targeted} and then MB_{Thy1-scFv} in this specific order (Fig. 8a). On dTE images, mice injected with MB_{Thy1-scFv} showed tumors with significantly increased Thy1 molecular imaging signal (3.1 ± 1.2 a.u.) compared to MB_{non-targeted} (0.4 ± 0.4 a.u.), with a quantitative outcome of ~ 7.8-fold (p < 0.03) (Fig. 8b,c) compared to control. Conversely, imaging of healthy pancreas did not produce any significant differences in imaging signal between the two MB constructs (MB_{Thy1-scFv} = 0.1 ± 0.1 a.u. and MB_{non-targeted} = 0.07 ± 0.06 a.u.) and were significantly lower than with MB_{Thy1-scFv} in PDAC model (39-fold, p < 0.02). Histological analysis of H&E-stained tissues confirmed presence of PDAC (Fig. 8d). Overall, these results demonstrate conserved Thy1-targeting property of Thy1-scFv in vivo and sketches the applicability for PDAC USMI.

Discussion

Molecular imaging has made a considerable contribution to oncology throughout the course of early detection and prognosis, and is an integral part of clinical trials. Biomarkers can be detected using various targeted-probes based on antibodies, peptides or proteins, oligonucleotides, or small molecules conjugated to imaging agents for suitable imaging modalities. Specifically, recombinant protein expression has become an established technique for production of cancer specific antigen-binding ligands in bacterial systems. However, conventional methods can be cumbersome toward meeting criteria for clinical applications. The purpose of this study was to engineer a production model for recombinant protein in their native form through the example of Thy1-scFv, while introducing its potential for early diagnosis of pancreatic cancer.

In vitro refolding technology of IBs has become prevalent to recover insoluble eukaryotic proteins expressed in *E. coli*^{10,26}. IBs are usually denatured with high concentration of chaotropes such as urea or guanidine hydrochloride²⁷. The major drawbacks of this method are its complex and expensive operational process, which further needs optimization at multiple steps. In addition, use of high concentrations of denaturing agents results in complete denaturation of the secondary structure favoring re-aggregation during successive process²⁶. Thus, recovery of soluble and active protein can be greatly reduced. In this study, we proposed an alternative to such extreme procedures and presented a multi-step process for enhanced-soluble protein production in SHuffle T7 *E. coli* cells cytoplasm using gene fusion technology. We used the well-known Trx-tag, improving disulfide bond formation, combined with S-tag. S-tag, commonly used as affinity tag, was here employed to enhance Thy1-scFv solubility thanks to its abundance in charged and polar residues²⁸. Both Trx-tag and S-tag are small and do not interfere with the proper folding or function of a fused target protein. Although essential for the expression and purification of recombinant proteins, tag sequences have the potential to interfere with the structure and the function of their fusion partner. In addition, many tags have interacting partners in mammalian systems which

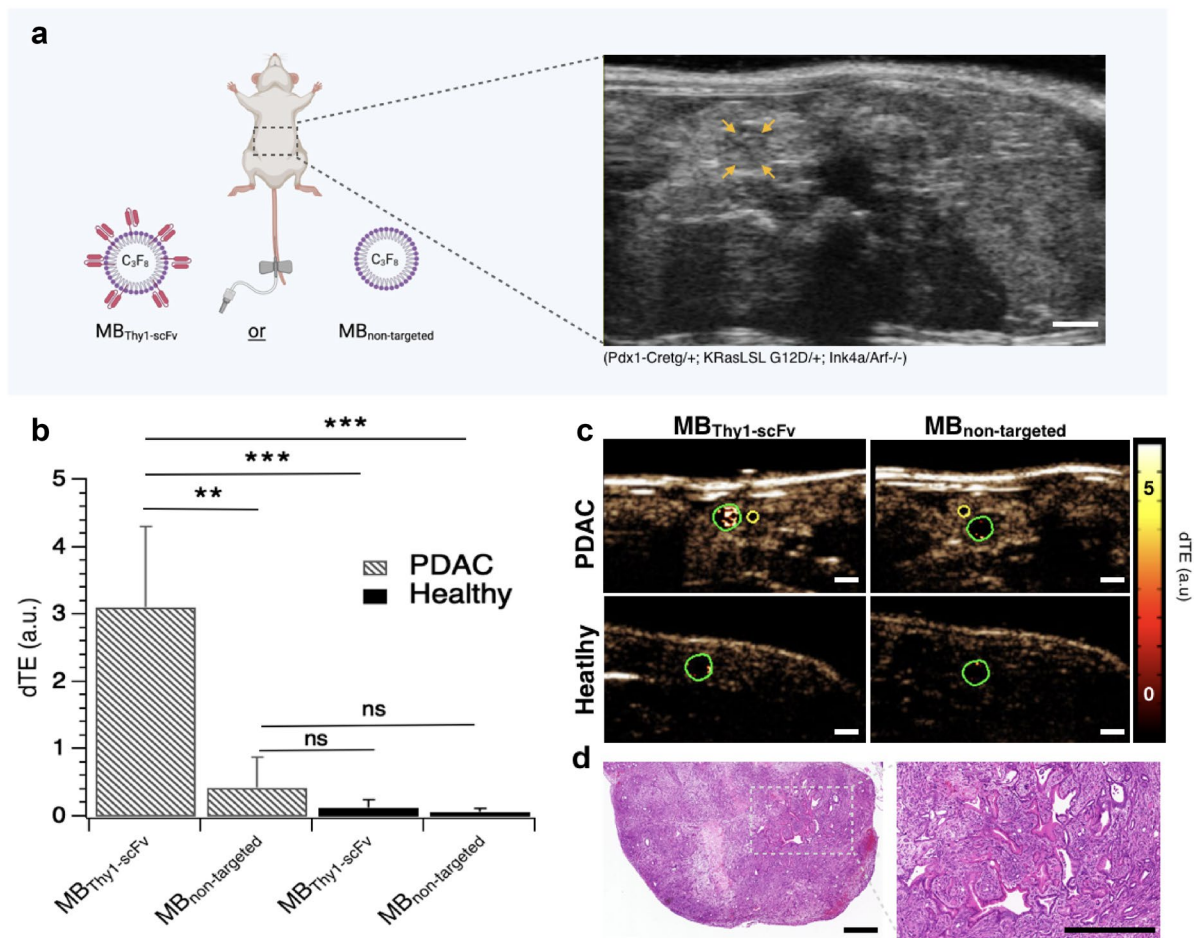


Figure 8. In vivo USMI of Thy1 expression in transgenic mouse model with spontaneous PDAC. **(a)** Schematic illustration of the overall experiment. After localizing the tumor by US B-mode abdominal imaging, MB_{non-targeted} and MB_{Thy1-scFv} were successively tested for Thy1 binding in PDAC transgenic mice. Orange arrows delineate a pancreatic tumor. B-mode images were used as references to draw region of interest (ROIs) in differential targeted enhancement (dTE) images presented in **(c)**. Scale bar 1 mm; **(b)** Quantitative bar graphs of in vivo dTE using targeted and non-targeted contrast agents in PDAC and healthy mice. ** $p < 0.03$; *** $p < 0.02$; **(c)** Representative transverse dTE images showing stronger signal enhancement in PDAC (green ROI) after injection of MB_{Thy1-scFv} and only low signal following injection of MB_{non-targeted}. Background signal was noted in adjacent normal pancreas (a yellow ROI was drawn to quantify imaging signal in adjacent non-PDAC tissue). Scale bar 1 mm; color coded scale is shown for USMI in arbitrary units (a.u.). **(d)** Corresponding hematoxylin-eosin-stained sample confirmed presence of PDAC in transgenic animals. Scale bar 500 μm .

can interfere with the biological applicability of recombinant proteins while generating a strong immune reaction. Therefore, tag removal should be considered, especially if the target protein is intended for pharmaceutical or therapeutic clinical applications, for crystallization, and for structural determination studies¹⁵. Chemical cleavage methods are usually inexpensive²⁹ but most systems rely on endopeptidases to separate the fusion partner from the protein of interest³⁰. Serine proteases such as the activated blood coagulation factor X (Factor Xa), EK, and thrombin, have been used, as well as viral proteases such as tobacco etch virus (TEV) protease and rhinovirus 3C protease. Viral proteases have a more stringent sequence specificity due to their much slower turnover rates (catalytic rate constant (kcat))^{31–35}, however EK has no amino acid specificity requirement on the P' part of the scissile bond (DDDDK↓) (only proline and tryptophan should be avoided on P1' which corresponds to an alanine residue in our study). Consequently, when an affinity tag is joined to the N terminus of the protein of interest, EK is able to regenerate a native N terminus. Moreover, we used a recombinant EK presenting the same affinity tag attached to the protein of interest, i.e., hexa-histidine tag. This allowed to apply the digestion products on the same affinity chromatography for separation. Undigested fusion protein substrate, tagged-protease, cleaved tag, and any endogenous proteins that bound to the affinity resin will be separate from the untagged protein of interest in the unbound effluent. After removal of the fusion partner tag, native Thy1-scFv was recovered (0.22 mg \pm 0.11) and used for further in vitro and in vivo experiments. We proved its binding functionality in vitro using Thy1-expressing cells. It can be noted that our cell binding assay demonstrated a high shift in fluorescence between our scFv and the corresponding commercial full-length antibody. Knowing that Thy1-Ab-APC (Thermo Fisher, CA) was purchased as such, while Thy1-scFv was conjugated to APC following our own protocol where we used biotin-streptavidin chemistry, the number of dyes per molecules could be different depending on the conjugation

chemistry used by the manufacturer. Moreover, full-length antibodies (MW ~ 150 kDa) will likely have more reactive amino acids available than smaller scFvs. For in vivo imaging, we used Thy1-scFv after conjugating to microbubbles by NHS-chemistry. We successfully imaged PDAC neovasculature in transgenic animals. It can be mentioned that a flexible cysteine-tag introduced on the C-terminal end of the scFv could give the possibility for multiple conjugation chemistry while providing site-specific labeling of molecules bearing maleimide functional groups for various applications.

Based on our results, we anticipate that the vector design and basic strategy presented in this study should be applicable to many proteins of biological interest which are currently difficult to purify. Overall, it is generally assumed that for *E. coli*, a 1-L fermentation will generate ~ 150 mg of total cellular proteins. Assuming an average yield between 0.5 and 5% of total proteins, 0.75–7.5 mg of recombinant protein is available in the cells. Based on our results, the protein of interests can be recovered between 30 and 50% in the soluble fraction, hence, our strategy could allow convenient small-scale production of biologically important recombinant proteins to initiate most studies (0.1–3 mg). Scale of production could be expanded by establishing a large-scale fermentation system to produce higher amount of proteins.

Given the low median survival rate (5-year survival rate < 9%) and the low percentage of PDAC patients qualifying for tumor resection (10–20%), the need for early screening methods is globally recognized³⁶. Efforts have been made to develop molecular imaging probes capable of detecting early stage PDAC³⁷. Here, we produced Thy1-scFv conjugated-US contrast agent, MB_{Thy1-scFv} and illustrated its potential for non-invasively enhancing USMI contrast between PDAC and normal pancreatic tissues in mice consistent with related findings^{24,38}. Our probe, able to detect small foci in the pancreas (> 2 mm), constitutes a promising translatable USMI agent. Naturally, the reproducibility and the sensitivity for PDAC molecular imaging will have to be further analyzed. With the success of the first and, to date, only targeted-MB in clinical trials for various cancers, BR55 (kinase insert domain receptor-targeted peptide), a rapid expansion of targeted US contrast agents is expected. Our promising pre-clinical results with MB_{Thy1-scFv} could provide opportunities for improved PDAC prognosis, and the presented targeted US contrast agent strategy, a variable format for other biomarker targeting.

Materials and methods

Ethical approval. The Administrative Panel on Laboratory Animal Care of Stanford University approved all procedures using laboratory animals used in this study, and all experiments were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals (APLAC-33828). This study was carried out following the ARRIVE guidelines.

Expression vector design. The expression vectors pET32b-1XHis-scFv, pET32b-3XHis-scFv, and pET32b-5XHis-scFv were constructed for Thy1-scFv expression. A ligation substrate featuring Thy1-scFv protein was amplified by PCR using a forward primer with NcoI restriction enzyme site and a reverse primer with XhoI restriction enzyme site. The amplified fragment was digested with NcoI and XhoI and ligated into pET-32b(+) prokaryotic expression vector digested with respective restriction enzymes to construct pET32b-1XThy1-scFv with a single inherent hexa-histidine-tag located between the Trx- and S-tags. To introduce more hexa-histidine-tags to construct pET32b-3XHis-scFv and pET32b-5XHis-scFv vectors, we inserted annealed forward and reverse primers coding for 2 and 4 additional hexa-histidine-tags (i.e., 3XHis and 5XHis total) with BglII restriction enzyme site on both the sides as overhangs with 5'-phosphate group. After ligation into pET32b-1XThy1-scFv vector previously digested with BglII restriction enzyme and dephosphorylated using Calf intestine alkaline phosphatase, we generated two additional vectors with 3X and 5X hexa-histidine-tags. All three plasmids contain: a T7 promoter; two fusion partners Trx- and S-tags for enhancing protein folding and solubility; 1, 3 or 5 hexa-histidine tag(s) for purification by immobilized metal affinity chromatography (IMAC); a DDDDK sequence on the N terminus of Thy1-scFv for tag removal using EK cleavage; and the Thy1-scFv gene. Each histidine tag was separated from each other by a few amino acid residues to increase flexible folding. The sequence confirmed, vectors were transformed into SHuffle T7 *E. coli* cells (New England Biolabs, Ipswich, MA) for recombinant protein expression. Oligonucleotides and recombinant protein sequences used in this study for constructing the vectors are listed in Supporting Information (Supplementary Table S1).

Thy1-scFv expression. Bacterial transformation was performed for each expression vector as follows: 50 µL of SHuffle T7 *E. coli* competent cells were transformed with 1 µg of expression plasmid using standard heat-shock procedure; 300 µL SOC growth media was then added in each vial for cell recovery. Cells were grown at 30 °C for 1 h in a shaking incubator (100 r.p.m.), and plated on Lysogeny Broth (LB)-agar medium containing ampicillin (50 µg/mL). After overnight growth, one fresh-picked colony was inoculated in 2 mL LB-ampicillin medium (50 µg/mL) and grown overnight at 30 °C (250 r.p.m.). Bacteria were transferred into 1 L of LB-ampicillin medium and further cultured until the OD_{600nm} reaches 0.4. The culture was induced for protein expression by the addition of isopropyl-β-d-thiogalactoside (IPTG, 1 mM) after diluting with the addition of one-fourth volume of pre-warmed LB-ampicillin medium. Induction was allowed for 4 h at 30 °C (250 r.p.m.). The pellet was then harvested via centrifugation (5000g, 10 min, 4 °C) and stored at – 80 °C.

Fusion protein purification (IMAC1). Cell pellets were resuspended in 20 mL of ice-cold lysis buffer (3 mM monosodium phosphate, 50 mM disodium phosphate, 500 mM NaCl, 5% glycerol (v/v), 5 mM CHAPS, and 20 mM imidazole containing protease inhibitors (Thermo Scientific, Rockford, IL)) and lysed by sonication (60% amplitude, 5 s on/off, 10 cycles, Branson SLPe). The soluble and insoluble fractions were then separated by centrifugation (12,000g, 10 min, 4 °C). Insoluble fractions containing cell debris and possible IBs were washed with 8 mL of the same lysis buffer. The soluble fractions were applied to a 1 mL FF His-trap column (GE

Healthcare Biosciences, PA) in an AKTA FPLC system (GE Healthcare Biosciences) equilibrated with PBS buffer containing 20 mM imidazole to reduce non-specific binding. The tagged-Thy1-scFv was purified using a linear gradient of imidazole (from 20 to 200 mM) in PBS buffer at a flow rate of 1 mL/min. Concentration of proteins were measured by UV spectrometry in each fraction. Purity of the fractions was analyzed on 4–12% gradient SDS-PAGE followed by staining with Coomassie Blue (SimplyBlue SafeStain, Carlsbad, CA) for visualization using a BioRad Gel-Doc system. Fractions containing the protein of interest were pooled and the best expression vector was utilized for further experiments. Western blot analysis was performed using anti-His tag antibody. A 4–12% gradient SDS-PAGE of the insoluble and soluble fractions was electroblotted onto a 0.2 μ m pore size nitrocellulose membrane (Bio Rad, Hercules, CA), blocked in PBS-T (PBS with 0.05% Tween 20) with 5% milk powder for overnight at 4 °C and then treated with anti-His tag antibody (BioLegend, San Diego, CA). After washing, the membrane was incubated with HRP-conjugated anti-mouse IgG antibody. Signals were visualized by the addition of enhanced-chemiluminescence (ECL) substrate and imaging using IVIS in vivo imaging system (Perkin Elmer, Santa Clara, CA).

Fusion protein cleavage by enterokinase. The purified protein was further treated with EK to remove the overhang tags (trx, S-tag and His-tags). Before EK treatment, the protein solution containing PBS-imidazole buffer was exchanged with 10 mM Tris HCL, 300 mM NaCl, pH7.4, using a 30-kDa molecular weight cutoff Vivaspin Protein Concentrator Spin Column (GE Healthcare Lifesciences, Pittsburgh, PA). In addition, the sample was concentrated during the same step. Tag removal was performed using EK enzyme tagged with hexahistidine (Genscript, Piscataway NJ; MW = 22.7 KDa) ranging from 10 to 160 IU per milligram of protein during 1 h to 24 h incubation period at 25 °C. Optimal EK concentration and incubation time were used for further experiments.

Purification of native protein (IMAC2) and characterization for purity. Following EK mediated cleavage, the sample was loaded on the equilibrated 1 mL FF His-trap column. Thereafter, the flow through was collected, followed by the application of a linear gradient of PBS buffer containing imidazole ranging from 20 to 200 mM. The fractions were collected and examined for Thy1-scFv purity and size by Matrix-Assisted Laser Desorption/Ionization time-of-flight (MALDI-TOF) mass spectrometry and SDS-PAGE. Fractions containing the native Thy1-scFv were pooled and further desalted/concentrated using a 10-kDa molecular weight cutoff Vivaspin Protein Concentrator Spin Column (GE Healthcare Lifesciences, Pittsburgh, PA) with PBS buffer, pH 7.4.

Cell culture. Wildtype MILE SVEN 1 (MS1_{WT}) mouse vascular endothelial cells (CRL2279; American Type Culture Collection (ATCC)) and MS1 cells engineered to stably express human Thy1 protein (MS1_{Thy1}) (selected using puromycin antibiotic marker) were cultured under sterile conditions in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and 100 U/mL penicillin and 0.1% streptomycin and maintained in 5% CO₂ at 37 °C and used for Thy1-scFv binding assays as stated below.

Binding of Thy1-scFv to Thy1-expressing MS1 cells. To evaluate the biological functionality of Thy1-scFv to specifically bind Thy1-expressing cells, MS1_{WT} and MS1_{Thy1} cells were incubated with Thy1-scFv-APC conjugate. Dye conjugation was performed by biotinylation of Thy1-scFv followed by streptavidin-APC incubation following the manufacturer's recommendation. Briefly, Thy1-scFv was incubated with 20-fold molar excess of NHS-PEG4-Biotin (Pierce Biotechnology, Rockford, IL) at room temperature for 30 min in PBS. Excess of biotin was removed using a Zeba spin column (Pierce Biotechnology, Rockford, IL) before the addition of streptavidin-APC (Tonbo Biosciences, San Diego, CA) to form Thy1-scFv-APC. Then, MS1_{WT} and MS1_{Thy1} cells were incubated with Thy1-scFv-APC (500 nM) or with a commercial Thy1-Ab-APC (eBioscience Inc, San Diego, CA) for 1 h at 4 °C, washed 3 times with PBS. Fluorescence intensity from the dye-labeled ligands were visualized by flow cytometry (Guava easyCyte; Luminex Corp., Austin, TX) and analyzed using FlowJo software (Becton Dickinson & company).

Synthesis and preparation of MBThy1-scFv as targeted-US contrast agent. Preparation of the US contrast agents, i.e., MBs, is detailed in the Supplementary Materials. Briefly, Thy1-scFv was conjugated to commercial phospholipids by NHS chemistry giving MB_{Thy1-scFv}. A non-targeted control, MB_{non-targeted}, was made with non-functionalized phospholipids.

Transgenic mouse model of PDAC. The transgenic pancreatic cancer mouse model (Pdx1-Cre^{tg/+}; KRas^{LSL G12D/+}; Ink4a/Arf^{-/-}) (n = 3), which spontaneously developed foci within 4–7 weeks after birth, was used³⁹. Correct genotype was validated for all mice. Tumor diameter ranging between 1.3 and 2.2 mm (mean 1.7 ± 0.4 mm) based on US images were used for the study. Healthy age-matched littermates were used as normal control group (C57BL/6, 6 weeks old, Charles River, Wilmington, MA, n = 3).

In vivo US molecular imaging of pancreas. In vivo US imaging of vascular Thy1 expression in transgenic PDAC mice and C57BL/6 mice with normal pancreas was performed using two MB constructs (MB_{Thy1-scFv} and MB_{non-targeted}) by following the protocol reported previously²⁴. In brief, a total of 10⁸ MB_{Thy1-scFv} or MB_{non-targeted} (100 μ L) was utilized for intravenous bolus injection via tail vein. All in vivo imaging studies were performed in contrast mode using a dedicated small animal high resolution US imaging system (Vevo 2100, FUJIFILM VisualSonics, Inc., Toronto, ON, Canada) with a linear transducer (MS250, VisualSonics) placed over the abdomen

of mice, guided by B-mode imaging to detect the target tissue of interest. Contrast mode images were acquired at 18 MHz, and all imaging parameters (focal length, 10 mm; transmit power, 4%; mechanical index, 0.2; dynamic range, 40 dB) were kept constant during all imaging sessions. A total time of 5 min was allowed for MBs to attach their target before binding quantification. To differentiate the acoustic signal owing to MBs attachment to Thy1 and the signal from freely circulating MBs, the previously described destruction-replenishment technique was employed²². The protocol consisted of 3 steps: (1) 200 frames of images capturing blood-vessel bound and unbound MBs within the ROI, (2) a high pressure destructive pulse (1-s continuous high-power destructive pulse of 3.7 MPa, transmit power, 100%; mechanical index, 0.63) to destroy all bound and unbound MBs, and (3) an additional set of 200 frames to measure the signal magnitude from the unbound MBs flowing into the ROI immediately after the destructive pulse. The difference in US imaging signal pre- and post-destruction corresponds to the Thy1 attached contrast agents, MB_{Thy1-scFv} or MB_{non-targeted}. A waiting interval of 20 min was maintained between each MB injection to allow for complete clearance before subsequent imaging. Any remaining attached MBs were destroyed by applying a high-power destruction pulse (see above for acoustic parameters).

Ultrasound molecular imaging data analysis. The molecular imaging signals were quantified post image acquisition with correction for breathing motion artifacts using Vevo 2100 integrated analysis software (VevoCQ; VisualSonics). Data analysis was accomplished by manually drawing ROIs around PDAC tissues, adjacent non-PDAC tissues, as well as in the normal pancreas of control littermates. The magnitude of imaging signal from attached MBs was assessed by subtracting the average imaging signals pre- and post-destruction and expressed as the differential targeted enhancement (dTE) in arbitrary units (a.u.).

Ex vivo analysis of pancreas tissues. PDAC mice were euthanized in accordance with animal care guidelines. The pancreas was excised and fixed in 4% paraformaldehyde (Santa Cruz Biotechnology Inc., CA) at 4 °C for 24 h. Tissues were cryosectioned and stained with hematoxylin eosin before analysis using a Nano-zoomer (Hamamatsu, Japan).

Statistical analysis. Student-t test was applied to determine statistical significance (** $p < 0.03$; *** $p < 0.02$) between groups and data expressed as mean \pm standard deviation (S.D.).

Received: 21 June 2021; Accepted: 16 November 2021

Published online: 29 November 2021

References

- Chang, A. C. Y. & Cohen, S. N. Genome construction between bacterial species in vitro: Replication and expression of staphylococcus plasmid genes in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **71**(4), 1030–1034. <https://doi.org/10.1073/pnas.71.4.1030> (1974).
- Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **27**(3), 297–306. <https://doi.org/10.1016/j.biotechadv.2009.01.008> (2009).
- Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2014.00172> (2014).
- Swartz, J. R. Advances in *Escherichia coli* production of therapeutic proteins. *Curr. Opin. Biotechnol.* **12**(2), 195–201. [https://doi.org/10.1016/S0958-1669\(00\)00199-3](https://doi.org/10.1016/S0958-1669(00)00199-3) (2001).
- Khow, O. & Suntrarachun, S. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pac. J. Trop. Biomed.* **2**(2), 159–162. [https://doi.org/10.1016/S2221-1691\(11\)60213-X](https://doi.org/10.1016/S2221-1691(11)60213-X) (2012).
- Structural Genomics Consortium., Architecture et Fonction des Macromolécules Biologiques., Berkeley Structural Genomics Center *et al.* Protein production and purification. *Nat. Methods* **5**(2), 135–146. <https://doi.org/10.1038/nmeth.f.202> (2008).
- Costa, S., Almeida, A., Castro, A. & Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: The Novel Fh8 system. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2014.00063> (2014).
- Ritz, D. & Beckwith, J. Roles of Thiol-Redox pathways in bacteria. *Annu. Rev. Microbiol.* **55**(1), 21–48. <https://doi.org/10.1146/annurev.micro.55.1.21> (2001).
- Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microb. Cell Fact* **14**(1), 41. <https://doi.org/10.1186/s12934-015-0222-8> (2015).
- Bowden, G. A., Paredes, A. M. & Georgiou, G. Structure and morphology of protein inclusion bodies in *Escherichia coli*. *Nat. Biotechnol.* **9**(8), 725–730. <https://doi.org/10.1038/nbt0891-725> (1991).
- Palmer, I. & Wingfield, P. T. Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. *Curr. Protoc. Protein Sci.* <https://doi.org/10.1002/0471140864.ps0603s70> (2012).
- Singhvi, P., Saneja, A., Srichandan, S. & Panda, A. K. Bacterial inclusion bodies: A treasure trove of bioactive proteins. *Trends Biotechnol.* **38**(5), 474–486. <https://doi.org/10.1016/j.tibtech.2019.12.011> (2020).
- Malhotra, A. Tagging for protein expression. *Methods Enzymol.* **463**, 239–258. [https://doi.org/10.1016/S0076-6879\(09\)63016-0](https://doi.org/10.1016/S0076-6879(09)63016-0) (2009).
- Pina, A. S., Lowe, C. R. & Roque, A. C. A. Challenges and opportunities in the purification of recombinant tagged proteins. *Biotechnol. Adv.* **32**(2), 366–381. <https://doi.org/10.1016/j.biotechadv.2013.12.001> (2014).
- Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.* **12**(7), 1313–1322. <https://doi.org/10.1110/ps.0243403> (2003).
- Sauzay, C., Voutetakis, K., Chatzioannou, A., Chevet, E. & Avril, T. CD90/Thy-1, a cancer-associated cell surface signaling molecule. *Front. Cell Dev. Biol.* <https://doi.org/10.3389/fcell.2019.00066> (2019).
- Croix, B. S. *et al.* Genes expressed in human tumor endothelium. *Science* **289**(5482), 1197–1202. <https://doi.org/10.1126/science.289.5482.1197> (2000).
- Madden, S. L. *et al.* Vascular gene expression in nonneoplastic and malignant brain. *Am. J. Pathol.* **165**(2), 601–608 (2004).
- Chen, X. *et al.* Novel endothelial cell markers in hepatocellular carcinoma. *Mod. Pathol.* **17**(10), 1198–1210. <https://doi.org/10.1038/modpathol.3800167> (2004).

20. Bradley, J. A., Ramirez, G. & Hagood, J. S. Roles and regulation of Thy-1, a context-dependent modulator of cell phenotype. *Bio-Factors* **35**(3), 258–265. <https://doi.org/10.1002/biof.41> (2009).
21. True, L. D. *et al.* CD90/THY1 is overexpressed in prostate cancer-associated fibroblasts and could serve as a cancer biomarker. *Mod. Pathol.* **23**(10), 1346–1356. <https://doi.org/10.1038/modpathol.2010.122> (2010).
22. Foygel, K. *et al.* Detection of pancreatic ductal adenocarcinoma in mice by ultrasound imaging of thymocyte differentiation antigen 1. *Gastroenterology* **145**(4), 885–894.e3. <https://doi.org/10.1053/j.gastro.2013.06.011> (2013).
23. Zhu, J., Thakolwiboon, S., Liu, X., Zhang, M. & Lubman, D. M. Overexpression of CD90 (Thy-1) in pancreatic adenocarcinoma present in the tumor microenvironment. *PLoS One* **9**, 12. <https://doi.org/10.1371/journal.pone.0115507> (2014).
24. Abou-Elkacem, L. *et al.* Thy1-targeted microbubbles for ultrasound molecular imaging of pancreatic ductal adenocarcinoma. *Clin. Cancer Res.* **24**(7), 1574–1585. <https://doi.org/10.1158/1078-0432.CCR-17-2057> (2018).
25. Vera, A., González-Montalbán, N., Aris, A. & Villaverde, A. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol. Bioeng.* **96**(6), 1101–1106. <https://doi.org/10.1002/bit.21218> (2007).
26. Singh, A., Upadhyay, V. & Panda, A. K. Solubilization and refolding of inclusion body proteins. In *Insoluble Proteins Methods in Molecular Biology* Vol 1258 (ed. García-Fruitós, E.) 283–291 (Springer, 2015). https://doi.org/10.1007/978-1-4939-2205-5_15.
27. Upadhyay, V., Singh, A., Jha, D., Singh, A. & Panda, A. K. Recovery of bioactive protein from bacterial inclusion bodies using trifluoroethanol as solubilization agent. *Microb. Cell Fact* **15**(1), 100. <https://doi.org/10.1186/s12934-016-0504-9> (2016).
28. Raines, R. T., McCormick, M., Van Oosbree, T. R. & Mierendorf, R. C. The S-tag fusion system for protein purification. In *Methods in Enzymology*. Vol 326 (ed. Sath, D.) 362–376 (Academic Press, 2000). [https://doi.org/10.1016/S0076-6879\(00\)26065-5](https://doi.org/10.1016/S0076-6879(00)26065-5).
29. Crimmins, D. L., Mische, S. M. & Denslow, N. D. *Chemical Cleavage of Proteins on Membranes*. *Current Protocols in Protein Science* 11.5.1-11.5.13 (Wiley, 2000). <https://doi.org/10.1002/0471140864.ps1105s19>.
30. Waugh, D. S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **80**(2), 283–293. <https://doi.org/10.1016/j.pep.2011.08.005> (2011).
31. Isetti, G. & Maurer, M. C. Employing mutants to study thrombin residues responsible for factor XIII activation peptide recognition: A kinetic study[†]. *Biochemistry* **46**(9), 2444–2452. <https://doi.org/10.1021/bi0622120> (2007).
32. Bromfield, K. M., Quinsey, N. S., Duggan, P. J. & Pike, R. N. Approaches to selective peptidic inhibitors of factor Xa. *Chem. Biol. Drug Des.* **68**(1), 11–19. <https://doi.org/10.1111/j.1747-0285.2006.00404.x> (2006).
33. Gasparian, M. E., Ostapchenko, V. G., Schulga, A. A., Dolgikh, D. A. & Kirpichnikov, M. P. Expression, purification, and characterization of human enteropeptidase catalytic subunit in *Escherichia coli*. *Protein Expr. Purif.* **31**(1), 133–139. [https://doi.org/10.1016/S1046-5928\(03\)00159-1](https://doi.org/10.1016/S1046-5928(03)00159-1) (2003).
34. Kapust, R. B. *et al.* Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng. Des. Sel.* **14**(12), 993–1000. <https://doi.org/10.1093/protein/14.12.993> (2001).
35. Long, A. C., Orr, D. C., Cameron, J. M., Dunn, B. M. & Kay, J. A consensus sequence for substrate hydrolysis by rhino virus 3C proteinase. *FEBS Lett.* **258**(1), 75–78. [https://doi.org/10.1016/0014-5793\(89\)81619-9](https://doi.org/10.1016/0014-5793(89)81619-9) (1989).
36. American Cancer Society. *Cancer Facts et Figures 2020* (American Cancer Society, 2020).
37. Jugniot, N., Bam, R., Meuillet, E. J., Unger, E. C. & Paulmurugan, R. Current status of targeted microbubbles in diagnostic molecular imaging of pancreatic cancer. *Bioeng. Transl. Med.* **6**, 1. <https://doi.org/10.1002/btm2.10183> (2021).
38. Bam, R. *et al.* Toward the clinical development and validation of a Thy1-targeted ultrasound contrast agent for the early detection of pancreatic ductal adenocarcinoma. *Invest. Radiol.* **55**(11), 711–721. <https://doi.org/10.1097/RLL.0000000000000697> (2020).
39. Aguirre, A. J. *et al.* Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes Dev.* **17**(24), 3112–3126. <https://doi.org/10.1101/gad.1158703> (2003).

Acknowledgements

The authors would like to thank Canary Center at Stanford, Department of Radiology for providing the facility and resources for executing this work.

Author contributions

N.J. and R.P. designed the study; N.J. and R.P. performed the experiments; N.J., and R.P., were involved in data analysis, R.B. helped with in vivo imaging; N.J. and R.P., wrote the manuscript. All authors reviewed the manuscript.

Funding

This work was funded by Phase I (R41 CA203090) and Phase II SBIR Grants from NIH (R41 CA203090).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-02445-2>.

Correspondence and requests for materials should be addressed to R.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021