A 33-residue peptide tag increases solubility and stability of *Escherichia coli* produced single-chain antibody fragments

Yang Wang1,7, Wenjie Yuan1,7, Siqi Guo2,3,7, Qiqi Li1, Xiaomei Chen1, Cheng Li1, Qianying Liu4, Lei Sun4, Zhenguo Chen1,4, Zhenghong Yuan1, Cheng Luo2,5, Shijie Chen2,8, Shuping Tong1, Michael Nossal6, Yu-Mei Wen1 & Yong-Xiang Wang1,8

Single-chain variable fragments (scFvs), composed of variable domains of heavy and light chains of an antibody joined by a linker, share antigen binding capacity with their parental antibody. Due to intrinsically low solubility and stability, only two *Escherichia coli*-produced scFvs have been approved for therapy. Here we report that a 33-residue peptide, termed P17 tag, increases the solubility of multiple scFvs produced in *Escherichia coli* SHuffle strain by up to 11.6 fold. Hydrophilic sequence, especially charged residues, but not the predicted α-helical secondary structure of P17 tag, contribute to the solubility enhancement. Notably, the P17 tag elevates the thermostability of scFv as efficiently as intra-domain disulfide bonds. Moreover, a P17-tagged scFv targeting hepatitis B virus surface proteins shows over two-fold higher antigen-binding affinity and virus-neutralizing activity than the untagged version. These data strongly suggest a type I intramolecular chaperone-like activity of the P17 tag. Hence, the P17 tag could benefit the research, production, and application of scFv.

Antibody (Ab) is composed of two identical heterodimers of a heavy (H) chain and a light (L) chain. Each chain contains at least two types of immunoglobulin domains, i.e., one amino (N)-terminal variable domain and one or multiple carboxyl (C)-terminal constant domains. Due to high homogeneity and specificity, and strong affinity to antigens, monoclonal antibodies (mAbs) are widely used for disease diagnosis and therapy. Therapeutic applications of mAb rely heavily on pharmacokinetic properties, including rates of vesicular extravasation, tissue distribution, and clearance. The large size (~150 kDa) of mAb restricts extravasation and distribution, reducing the effective mAb concentration within target tissues.1 Due to long (~3 week) serum half-life, prolonged retention of mAb conjugated with radioisotopes or toxins may lead to unexpected pathogenicity and toxicity. Besides, mAbs may extensively activate lymphocytes via interaction of their crystallizable fragments (Fc) with cell-surface receptors, resulting in adverse effects, e.g., cytokine storm.2 These drawbacks are avoided by Ab fragments, such as antigen binding fragments (Fab) and variable fragments (Fv), which maintain the antigen specificity of their parental mAbs, but have Fc or the entire constant domains of both H and L chains deleted. Due to their smaller sizes, Ab fragments have higher

---

1Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of Basic Medical Sciences, Shanghai Medical College, Fudan University, Shanghai, China. 2Drug Discovery and Design Center, the Center for Chemical Biology, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China. 3School of Pharmacy, Nanchang University, Nanchang, China. 4Institutes of Biomedical Science, Fudan University, Shanghai, China. 5School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China. 6Department of Internal Medicine II/Molecular Biology, University Hospital Freiburg, Freiburg, Germany. 7These authors contributed equally: Yang Wang, Wenjie Yuan, Siqi Guo. 8These authors jointly supervised this work: Shijie Chen, Yong-Xiang Wang. e-mail: shijiechen@simm.ac.cn; yongxiangwang@fudan.edu.cn
to reduce disulfide bonds leading to aggregation14,15. The cytoplasmic environment due to inactivation of the active enzyme (e.g., Erv1p, DsbB or VKOR) plus a disulfide bond plus a peptide12. Short (5–10 aa) peptide linkers to connect VH with VL promote intermolecular VH–VL interactions (called domain-swapping) by sterically disfavoring intramolecular VH–VL interactions, giving rise to dimers (termed diabodies)13. Additional strategies have been developed to improve solubility and stability of bacterially expressed scFv, including directed evolution, rational design, co-expression of chaperones, and fusion with solubility-enhancing tags2. However, despite these efforts, until 2020 only eight scFv fragments, two scFv of them produced in E. coli, have been approved for clinical use by the US Food and Drug Administration29. Intradomain disulfide bonding between two β-sheets of a VH or VL domain is often important to stabilize scFv. Mutations of the involved cysteines increase the propensity for aggregation29,30. The cytoplasmic compartment of common E. coli strains not only lacks components to catalyze the formation of disulfide bonds but instead possesses active glutathione reductase (gor) and thioredoxin reductase (txrB) pathways to reduce disulfide bonds. Hence most cytoplasmically expressed scFv are obtained as non-functional insoluble aggregates. While the periplasmic space, in contrast, provides an oxidative environment and contains thiol-disulfide oxidoreductases such as DsbA and DsbB, recombinant expression yields are mostly very low. Recently, soluble expression of some scFv has been achieved in an engineered E. coli strain (termed SHuffle) which features a more oxidative and disulfide-promoting cytoplasmic environment due to inactivation of the gor and txrB pathways and the constitutive expression of the disulfide-bond isomerase DsbC15,16,24. Also, other E. coli strains co-expressing a redox-active enzyme (e.g., Erv1p, DsbB or VKOR) plus a disulfide-bond isomerase like DsbC or PDH have become available29,30. However, some scFv still exhibit low solubility in these engineered E. coli strains31 and the low stability of scFv remains problematic16.

As reported, folding of select scFv is predominantly mediated by intrinsically strong VH–VL interactions, independent of intradomain disulfide bond15. Their at least partly soluble cytoplasmic expression in E. coli is thus not unexpected. Fusion with some protein tags such as maltose-binding protein (MBP), glutathione-S-transferase (GST), N-utilization substance A (NusA) and thioredoxin (Trx) can further improve the solubility29,32. However, for many scFv with weak VH–VL interactions, this is unlikely the case29. In addition, it remains tricky and laborious to refold insoluble scFv into active forms using detergents and additives, although fusion to MBP may improve expression and refolding yields33. However, it is often mandatory to remove bulky tags like MBP to preserve the scFv’s antigen-binding and pharmacokinetic properties34. In turn, tag removal often destabilizes the scFv and induces non-functional aggregate formation29,32.

The P17 protein from the tail of T7 phage can target phage particles yet also other cargos including small molecules, proteins, siRNA, DNA polyplexes, and liposomes to hepatocytes in vitro and vivo35,36. This is predominantly mediated by the association of a 33-amino acid (aa) P17 protein segment (named P17 peptide or P17 tag hereafter) with cell-surface low density lipoprotein receptor-related protein (LRP)33. We previously generated a scFv of mAb GI218,34, which targets the surface protein of hepatitis B virus (HBV), a virus with pronounced liver tropism. To improve hepatocyte targeting and uptake of GI2-scFv, here we attach the P17 tag to its C-terminus. When expressed in the E. coli SHuffle strain, the GI2-scFv-P17 fusion protein displays much higher solubility than the untagged GI2-scFv. Furthermore, the P17 tag also increases the solubility of three other scFv in SHuffle cells, as revealed by the semiquantitative solubility assay. The current study aims to identify the sequence and structure determinants within the P17 peptide that are responsible for its solubility and stability-enhancing impact on scFv, and to clarify the underlying mechanisms.

**Results**

**P17 tag enhanced solubility of four different scFv in *E. coli* SHuffle T7 strain**

Plasmid pET28a-His-GI2-scFv-HA expresses a GI2-scFv protein with a His tag on the N terminus and a HA tag at the C-terminus18. Into this plasmid, the P17 tag coding sequence was inserted downstream of the HA tag sequence, with the resulting plasmid pET28a-His-GI2-scFv-HA-P17 encoding the same protein but now with the additional P17 tag at the C-terminus. E. coli SHuffle T7 cells transformed with either of the two plasmids were cultured in Luria-Bertani (LB) media supplemented with 0.5 mM IPTG to express recombinant GI2-scFv-HA and GI2-scFv-HA-P17 proteins, and subsequently were suspended in 10 ml lysis buffer per gram (wet weight) of bacteria. After cell rupture by sonication, 10 μl of each lysate containing total bacterial protein was mixed with 40 μl 2× SDS-PAGE loading buffer as the source of total scFv. After centrifugation, 10 μl of the clear lysate supernatant containing soluble bacterial protein was mixed with 40 μl 2× SDS-PAGE loading buffer as the source of soluble scFv. To determine the solubility of scFv defined as the percentage ratio of soluble to total scFv, both fractions of scFv were semi-quantitated by western blotting using a mAb against the common His tag followed by a horseradish peroxidase-conjugated secondary antibody plus a chemiluminescent substrate, and subsequent densitometry of the blotting signals recorded by a charge-coupled device camera (Fig. 1a, upper panel). To avoid signal saturation, both fractions of scFv were diluted serially by 1/5, 1/10, 1/20, 1/40, 1/80, and 1/160 before loading. As shown in Fig. 1a (middle panel), western blotting signals of both fractions of scFv correlated logarithmically with the dilution. However, at high dilutions ranging from 1/160 to 1/20, signal intensities displayed an approximately linear correlation with the dilution (Fig. 1a, bottom panel). Therefore, for quantitation, soluble scFv was diluted in the range from 1/20 to 1/80, while total scFv was diluted by 1/160 due to the higher amount. ScFv solubility was assessed after normalizing the dilution of soluble scFv at 1/160. As shown in the left panel of Fig. 1b, the solubility of GI2-scFv-HA and GI2-scFv-HA-P17 proteins was 14.9% and 45.0%, respectively, indicating the P17 tag increased GI2-scFv solubility.

In line with our previous report18, the GI2-scFv-HA protein was virtually insoluble when expressed in BL21 (DE3) Star cells, which harbor active gor and txrB pathways preventing intradomain disulfide bond formation (Fig. 1b, right panel). Also, the GI2-scFv-HA-P17 protein remained insoluble in this strain (Fig. 1b, right panel), indicating that the solubility-enhancing effect of the P17 tag required an oxidative cytoplasmic environment.

To check whether the P17 tag could enhance the solubility of other scFv, recombinant pET28-derived plasmids were constructed to express three additional scFvs: VRC01-scFv targeting gp120 of HIV-115, ADRI-scFv targeting the small surface protein of HBV16, and MA18/7-scFv targeting pres1 domain of the large HBV surface protein17, either with or without the P17 tag at their C termini. As shown in Fig. 1c, d, the
solubility of these scFvs expressed in SHuffle T7 strain showed some variations, but fusion with the P17 tag increased in all cases solubility from 2.4 to 6.5-fold. The solubility-promoting effect of the P17 tag was most remarkable for MA18/7-scFv, with its untagged version having the lowest solubility (Fig. 1d).

In the scFv constructs described above, a HA tag was located between scFv and P17 tag to facilitate immunological detection. To check whether the HA tag contributed to the solubility-enhancing effect of the P17 tag, we constructed an analogous plasmid, pET28a-His-MA18/7-scFv-P17, which lacks the HA coding sequence. In direct comparison with MA18/7-scFv-HA, the solubility of this new construct, MA18/7-scFv-P17, was even more elevated (by 9.5-fold) (Fig. 1c, d), indicating that the additional presence of the HA tag decreased solubility enhancement by the P17 tag.

![Image](https://example.com/image.png)

**Fig. 1 | P17 tag enhanced solubility of four scFvs in E. coli SHuffle T7 cells.**

- **a** Determination of dilution ranges of total and soluble fractions of G12-scFv-HA protein for linear correlation with western blotting signals. Upper panel: ScFvs from 5 μl total and soluble bacterial protein samples serially diluted with 2× SDS-PAGE loading buffer were detected by western blotting using an anti-His tag. The relative density value (rel. DV) of each full-length (FL) scFv was obtained by setting the density of total scFv diluted by 1/5 at 100%. Middle panel: Over the entire dilution range rel. DV of both total and soluble scFvs correlated with dilutions logarithmically as calculated by the non-linear fitting program of Origin 8.0 software. Bottom panel: In the dilution range from 1/20 to 1/160, rel. DV of both total and soluble scFvs correlated linearly with dilutions as fulfilled by a linear fitting program of Origin 8.0. Linear equations of rel. DV and dilution (D) are also presented.
- **b** Determination of solubility of G12-scFv-HA and G12-scFv-HA-P17 proteins expressed in E. coli SHuffle T7 and BL21(DE3) Star strains. ScFvs solubility was determined as the percentage ratio of soluble versus total scFvs after their dilutions were normalized at 1/160, and the result of one representative experiment is shown. − below the detection limit, T total protein, S soluble protein.
- **c** Determination of solubility of VRC01-scFv-HA, ADRI-scFv-HA, MA18/7-scFv-HA, and their P17-tagged derivatives expressed in SHuffle T7 strain. The result of one representative experiment was shown.
- **d** Fold change of solubility of P17-tagged scFvs to corresponding untagged scFvs and the isoelectric point (pI) of each untagged scFv.

Nature Communications | (2022) 13:4614

https://doi.org/10.1038/s41467-022-32423-9
Both N- and C-terminal fusions of the P17 tag efficiently increased scFv solubility

The impact of a tag on the solubility of a fusion protein can be affected by the tag’s location at the N or C-terminus of a protein of interest\textsuperscript{38}. Therefore, we compared the solubility of MA18/7-scFv with the P17 tag directly linked to either end (Fig. 2a). As shown in Fig. 2b–d, the P17 tag significantly improved solubility at both N- and C-positions by 11.6- and 10.2-fold, respectively. In the additional presence of a HA tag at the C-terminus, the P17 tag increased solubility by 3.5-fold when at the N-terminus and 5.9-fold when at the very C-terminus (Fig. 2d). Therefore, direct attachment of the P17 tag to either terminus of the scFv increased solubility efficiently, whereas an additional HA tag at the scFv C-terminus reduced the P17-mediated solubility enhancement.

A C-terminal P17 tag enhanced GI2-scFv solubility more efficiently than N- or C-terminally fused MBP in SHuffle T7 cells

Multiple fusion tags, including MBP, have been applied to improve the solubility of scFvs in \textit{E. coli}\textsuperscript{27–29}. Indeed, N-terminal but not C-terminal MBP fusion increased the solubility of GI2-scFv by 1.4-fold in SHuffle T7 cells (Fig. 3c) as revealed by the solubility assay (Fig. 3a, b). C-terminal P17 tag fusion increased the solubility of GI2-scFv by 1.7-fold, in line with a higher soluble yield of GI2-scFv-P17 than GI2-scFv-HA and MBP-GI2-scFv (Supplementary Fig. 5b). Hence in SHuffle T7 cells, the P17 tag increased the solubility of GI2-scFv more efficiently than MBP. Glutathione-S-transferase (GST) is another protein tag that is commonly used to improve the solubility of passenger proteins in \textit{E. coli}\textsuperscript{39}, often as N-terminal fusion\textsuperscript{40}. Its effect on the solubility of GI2-scFv was therefore also tested. However, the GST-GI2-scFv-HA protein was even much less soluble in SHuffle T7 cells than the N-terminally untagged version (Fig. 3a–c).

Adding two or three copies of P17 tags did not further improve scFv solubility

The solubility-enhancing ability of the P17 tag may be endowed by recruiting other molecules such as chaperones or perhaps tRNAs\textsuperscript{41–44} to promote protein folding. If so, adding two or three copies of the P17 tag might further elevate the solubility of a test scFv. To test this hypothesis, two or three copies of the tag were fused to the C-terminus of MA18/7-scFv. The results showed that a single P17 tag increased solubility by 8.3-fold, whereas two and three copies enhanced solubility by only 6.6- and 3.8-fold, respectively (Fig. 4).

A predicted α-helix in the P17 tag is not required for solubility enhancement

Structural features of a protein tag can be important for solubility enhancement\textsuperscript{45,46}. The free P17 peptide did not exhibit any characteristic α-helical structure when analyzed by circular dichroism spectroscopy\textsuperscript{47}, but an earlier study of the entire P17 protein had noted a substantial α-helix propensity in the rod-region harboring the P17 tag sequence\textsuperscript{38}. In the absence of direct structural information, we applied up-to-date secondary structure prediction algorithms, JPred\textsuperscript{48} and PEP4\textsuperscript{49}, to the P17 peptide (see Supplementary Fig. S1a for the JPred\textsuperscript{4} results), and the three-dimensional (3D) structure prediction program AlphaFold\textsuperscript{2} to the entire MA18/7-scFv-HA-P17 protein (Supplementary Fig. S1b). Either program predicted a single region within the P17 tag to have some α-helix propensity, namely residues 22–29 (RDEAKRFK, Fig. 5a) near the C-terminus, although the AlphaFold\textsuperscript{2} predicted local Distance Difference Test (plDDT) score in this region was below 70 (Supplementary Fig. S1b), indicating an at most moderate confidence in the predicted α-helix. However, the α-helix might still form in the scFv context or be induced by interaction with bacterial components.

![Figure 2](https://example.com/figure2.png)

**Fig. 2** | P17 tag significantly improved solubility of scFv in SHuffle T7 strain regardless of fusion positions. a Schematic diagrams of His-MA18/7-scFv-HA protein and its N- and C-terminally P17-tagged derivatives with a C-terminal HA tag fusion or not. b Determination of solubility of His-MA18/7-scFv-HA protein and its derivatives expressed in SHuffle T7 cells by western blotting. The result of one representative experiment was shown. T, total protein; S, soluble protein. c Solubility of His-MA18/7-scFv-HA and its derivatives were determined by three independent experiments. Data are presented as mean values ± SD. The statistical significance of differences between two experimental groups was assessed through one-way analysis of variance (ANOVA). d Fold change of solubility of P17-tagged derivatives of His-MA18/7-scFv-HA to untagged version.
notably the abundant tRNA, as in the recently described "chaperna" mechanism where RNA is crucial for chaperoning. Different from the RNA interaction domain used there, we could not detect a stable association of the P17 tag with RNA (Supplementary Fig. S2), disfavoring such a mechanism. To further experimentally assess whether the putative $\alpha$-helix, if formed, was important for solubility enhancement by the P17 tag, we replaced the alanine in the center of the putative helix, A25, by a helix-breaking proline residue (Fig. 5a, variant A25P); JPred4, as well as alphafold2, predicted indeed a loss of the $\alpha$-helix (Supplementary Fig. S1a, b, variant A25P). When attached to the C-terminus of MA18/7-scFv-HA, both the wild-type (wt) and A25P mutant P17 tag increased protein solubility by around fourfold (Fig. 5b–d). Furthermore, a variant lacking the entire sequence forming the putative $\alpha$-helix (variant P17-C15del), was at most slightly less efficient than the wt P17 tag to increase the solubility of MA18/7-scFv (see below, Fig. 6a–d). Therefore, the putative $\alpha$-helix structure, if formed at all, was dispensable for solubility enhancement.

Both N- and C-proximal sequences in the P17 tag contribute to scFv solubility enhancement

The 33-aa P17 tag is highly hydrophilic due to the presence of eight basic aa (R, K), five acidic aa (D, E), and twelve polar aa (e.g., N, Q, S, T), all of which may contribute to the tag’s solubility-enhancing property. As these hydrophilic residues are scattered throughout the sequence (Fig. 6a), we truncated the P17 tag from either terminus (Fig. 6a) to define the relevant sequence element(s).

P17 mutants P17-C8del and P17-C15del lack 8 and 15 aa from the C-terminus to partly and completely delete the sequence forming the predicted $\alpha$-helix. Attached to the C-terminus of MA18/7-scFv-HA, both the wild-type (wt) and A25P mutant P17 tag increased protein solubility by around fourfold (Fig. 5b–d). Furthermore, a variant lacking the entire sequence forming the putative $\alpha$-helix (variant P17-C15del), was at most slightly less efficient than the wt P17 tag to increase the solubility of MA18/7-scFv (see below, Fig. 6a–d). Therefore, the putative $\alpha$-helix structure, if formed at all, was dispensable for solubility enhancement.

The 33-aa P17 tag is highly hydrophilic due to the presence of eight basic aa (R, K), five acidic aa (D, E), and twelve polar aa (e.g., N, Q, S, T), all of which may contribute to the tag’s solubility-enhancing property. As these hydrophilic residues are scattered throughout the sequence (Fig. 6a), we truncated the P17 tag from either terminus (Fig. 6a) to define the relevant sequence element(s).

P17 mutants P17-C8del and P17-C15del lack 8 and 15 aa from the C-terminus to partly and completely delete the sequence forming the predicted $\alpha$-helix. Attached to the C-terminus of MA18/7-scFv-HA, both the wild-type (wt) and A25P mutant P17 tag increased protein solubility by around fourfold (Fig. 5b–d). Furthermore, a variant lacking the entire sequence forming the putative $\alpha$-helix (variant P17-C15del), was at most slightly less efficient than the wt P17 tag to increase the solubility of MA18/7-scFv (see below, Fig. 6a–d). Therefore, the putative $\alpha$-helix structure, if formed at all, was dispensable for solubility enhancement.

Both N- and C-proximal sequences in the P17 tag contribute to scFv solubility enhancement

The 33-aa P17 tag is highly hydrophilic due to the presence of eight basic aa (R, K), five acidic aa (D, E), and twelve polar aa (e.g., N, Q, S, T), all of which may contribute to the tag’s solubility-enhancing property. As these hydrophilic residues are scattered throughout the sequence (Fig. 6a), we truncated the P17 tag from either terminus (Fig. 6a) to define the relevant sequence element(s).

P17 mutants P17-C8del and P17-C15del lack 8 and 15 aa from the C-terminus to partly and completely delete the sequence forming the predicted $\alpha$-helix. Attached to the C-terminus of MA18/7-scFv-HA, both the wild-type (wt) and A25P mutant P17 tag increased protein solubility by around fourfold (Fig. 5b–d). Furthermore, a variant lacking the entire sequence forming the putative $\alpha$-helix (variant P17-C15del), was at most slightly less efficient than the wt P17 tag to increase the solubility of MA18/7-scFv (see below, Fig. 6a–d). Therefore, the putative $\alpha$-helix structure, if formed at all, was dispensable for solubility enhancement.
strongly as the wt tag (Fig. 6g). Taken together, both N- and C-terminal sequences of the P17 tag can independently improve scFv’s solubility in the SHuffle T7 strain.

The solubility-enhancing ability of the P17 tag correlated with the number of charged residues
Polyanionic and polycationic protein tags have been reported to improve the solubility of some recombinant proteins. As 15-aa P17-N18del retained wt-like solubility-enhancing ability, it was used for subsequent point-mutation analysis. Because two anionic residues (D23, E24) and two cationic residues (K26, R27) are clustered in the central region of P17-N18del (Fig. 7a), we investigated their potential roles by site-directed mutants (P17-N18del-KR, P17-N18del-DE, and P17-N18del-DEKR) with D23/E24 and/or K26/R27 replaced alone or in combination by alanine (Fig. 7a) were fused to the C-terminus of MA18/7-scFv-HA protein and its two derivatives with C-terminally fused P17-WT and P17-A25P tags. T total protein, S soluble protein. c Solubility of MA18/7-scFv-HA protein and its derivatives determined by three independent experiments. d Fold change of solubility of MA18/7-scFv-HA-P17 and MA18/7-scFv-HA-P17-A25P proteins compared to MA18/7-scFv-HA protein. Data are presented as mean values ± SD. The statistical significance of differences between the two experimental groups was assessed by one-way ANOVA.

The P17 tag increases scFv stability
Protein tags can enhance solubility by assisting in the correct folding of the fusion partners. This reduces exposure of hydrophobic regions to solvent and inhibit formation of aggregates. Protein stability decreases with increasing temperature and can be reflected by the melting temperature (Tm) at which the concentrations of folded and unfolded proteins are equal. Generally, a higher Tm indicates higher stability and vice versa. As reported previously, the Tm of a test protein can be determined by a thermal shift assay. In this assay, fluorescent dye (SYPRO Orange) binds to hydrophobic regions exposed during heat denaturation, yielding more 570-nm-wavelength quanta. As determined by this assay, treatment of G12-scFv-HA with 5 mM DTT to reduce disulfide bonds led to ~3 °C decrease of Tm (Fig. 8a), in line with the presence of such disulfides and their importance of intradomain disulfide bonds to maintain scFv stability in previous studies.

Next, to check the effect of P17 tag on scFv stability, recombinant G12-scFv-HA and G12-scFv-HA-P17 proteins were firstly purified by IMAC (Supplementary Fig. S3a). Quantitation of both proteins by SDS-PAGE and Coomassie-Blue staining using BSA as standard revealed soluble yields of 3 mg and 8 mg per liter culture (Supplementary Fig. S3b), corroborating the solubility enhancement by the P17 tag. Five μM solutions of both proteins were mixed with equal concentrations of SYPRO Orange and subjected to the thermal shift assay. As shown in Fig. 8b, G12-scFv-HA-P17 displayed a 2.3 °C higher Tm than G12-scFv-HA, indicating that the P17 tag stabilized scFv by inhibiting exposure of hydrophobic regions. As a control, incubation in trans with in vitro synthesized P17 peptide failed to increase the Tm of G12-scFv-HA even in 10-fold excess (Fig. 8a), in line with the presence of such disulfides and their importance of intradomain disulfide bonds to maintain scFv stability in previous studies.

P17-tagged G12-scFv has higher antigen binding affinity and specificity and HBV infection-neutralizing activity than the untagged protein
Due to its increased stability, the scFv-P17 fusion protein may possess stronger antigen-binding affinity than unfused scFv, unless the tag would sterically hinder antigen binding. To compare the binding affinity of G12-scFv-HA and G12-scFv-HA-P17 to recombinant HBV small surface protein produced in CHO cells, the equilibrium dissociation constants (Kd) were determined by biolayer interferometry. A 2.2-fold lower Kd (0.79 × 10⁻⁹ M versus 1.77 × 10⁻⁹ M, Fig. 9a) for G12-scFv-HA-P17 than for G12-scFv-HA was obtained. These results were further corroborated by enzyme-linked immunosorbent assay (ELISA), whereby the pooled sera of chronic hepatitis B patients and healthy donors, confirmed to be positive and negative for hepatitis B surface
antigen (HBsAg) by a commercial ELISA kit (Fig. 9b), were reacted with the microtiter wells coated with 5 μg/ml solutions of G12-scFv-HA and G12-scFv-HA-P17, respectively. ELISA A450nm value of G12-scFv-HA-P17 reacting with HBsAg positive sera was significantly (P < 0.0001), approximately twofold higher than G12-scFv-HA (Fig. 9b). These data collectively indicate stronger antigen binding affinity of G12-scFv-HA-P17 than G12-scFv-HA, and also suggest that the P17 tag did not interfere with scFv-antigen binding. Generally, stronger antigen binding affinity of scFv suggests higher specificity. Indeed, ELISA A450nm value of G12-scFv-HA-P17 incubating with HBsAg negative sera was significantly (P < 0.05) lower than G12-scFv-HA, thus indicating the higher specificity (Fig. 9b). Importantly, both G12-scFv-HA and G12-scFv-HA-P17 proteins efficiently inhibited HBV infection of a susceptible cell line as judged by the dose-dependent suppression of HBV RNA and secreted hepatitis B e antigen (Fig. 9c). However, G12-scFv-HA-P17 protein displayed higher neutralizing activity than G12-scFv-HA, as indicated by of 50% inhibitory concentrations at 2.91 nM versus 6.74 nM (Fig. 9c, right panel).

P17 tag did not lead to oligomerization of G12-scFv

Some protein tags like GST mediate oligomerization of fusion partners by spontaneous self-assembly into dimer or higher oligomers. Recombinant full-length P17 protein assembles into trimer, as natively present in T7 phage tail fibers. In contrast, the 33-aa P17 peptide existed as a monomer in solution as determined by circular dichroism and analytical ultracentrifugation. To assess the potential impact of the P17 tag on G12-scFv quaternary structure, recombinant G12-scFv-HA and G12-scFv-HA-P17 proteins were analyzed by size-exclusion chromatography (SEC). As shown in Fig. 10a (left panel), the Superdex 200 Increase 10/300 column separated protein markers ranging from 0.137 kDa to 670 kDa very well, giving the linear calibration curve and equation of average distribution constants (Kav) of each protein marker and the logarithm of the respective molecular weight (Log(MW), Fig. 10a, right panel). Under identical experimental conditions, both the G12-scFv-HA and the G12-scFv-HA-P17 protein displayed two peaks (Fig. 10a, left panel), representing monomeric and dimeric forms of each protein based on the calibration curve and equation. To confirm
this interpretation, both protein preparations were cross-linked using 3% (v/v) formaldehyde, and the products were analyzed by SDS-PAGE and subsequent western blotting using the anti-His-tag mAb (Fig. 10b). This revealed dimers as the major higher-order products. The ratios of dimer to monomer were essentially the same for G12-scFv-HA and G12-scFv-HA-P17, both by the intensity of the western blot signals and the areas of the SEC elution peaks (Fig. 10a, b). Taken together, the P17 tag neither resulted in oligomerization of G12-scFv nor did it alter the equilibrium between monomeric and dimeric forms, rather in line with a type I than an assembly-mediating type II intramolecular chaperone mechanism\(^5\).

### Discussion

As one format of antibody fragments, scFv are applied in the fields of basic research, diagnostics, and therapeutics. Amongst different expression hosts\(^5\),\(^6\),\(^7\) *E. coli* offers rapid growth at low cost\(^8\). However, scFv are often expressed as insoluble aggregates because common *E. coli* strains with their reducing cytoplasmic environment fail to...
Support formation of intramolecular disulfide bond. Strategies to improve scFv solubility in *E. coli* include expression in the oxidative periplasmic space, reduced culture temperature, co-expression of chaperone(s), using engineered *E. coli* variants providing a more oxidizing cytoplasm, and fusion of solubility-promoting tags to the scFv.

Here, we characterized the 33-aa P17 peptide tag, and further shortened variants, as a small-size fusion tag which markedly increases the solubility, soluble yield, and stability of scFvs in *E. coli*.

**Solubility-enhancing effect of P17 tag on multiple scFvs**

Analogous to other protein tags, solubility enhancement of scFvs by the P17 tag depends on both the expression host and properties of the specific scFv. Firstly, in line with intradomain disulfide-bond formation promoting scFv solubility in *E. coli*, G12-scFv was partly soluble in the SHuffle strain but not in the BL21(DE3) strain with its reducing cytoplasm (Fig. 1b). Only in SHuffle cells did the P17 tag further increase scFv solubility (Fig. 1d). Secondly, while the P17 tag did not improve solubility of a truncated duck hepatitis B virus polymerase, notorious for poor solubility in *E. coli* (Supplementary Fig. S4), the tag enhanced solubility of four different scFvs (Fig. 1d). Part of Ab diversity is due to many subgroups of heavy and light chains. Heavy chains are encoded by a single gene cluster (IGH), light chains by the κ cluster (IGK) and the λ cluster (IGL). Sequence analysis of the four scFvs using online International ImmuGeneTics Information System (http://www.imgt.org/) revealed that their variable regions are encoded by different subgroups of heavy (IGHV) and light chain variable gene clusters (IGKV, and IGLV). Accordingly, the variable regions of MA18/7-scFv are encoded by IGHV3 and IGKV6; those of G12-scFv by IGHV4 and IGKV3; those of VRC01-scFv by IGHV1 and IGKV1; and those of ADRI-scFv by IGHV3 and IGLV3. Although these scFvs vary considerably in sequence and their intrinsic solubility properties in *E. coli*, attaching the P17 tag improved the solubility of all by at least 2.4-fold and up to 11.6-fold (Figs. 1 and 2). Hence, the P17 tag can improve the solubility of multiple scFvs in the SHuffle strain.

**Properties of the P17 tag and implications**

Their small sizes of around 30 kDa convey to scFvs efficient tissue penetration and renal glomerular filtration. Oligomerization and the corresponding increase in size would substantially dampen these properties. However, the P17 peptide as such exists as monomer in solution and also as a tag it neither induced oligomerization of scFv nor did it affect the intrinsic scFv monomer-dimer equilibrium (Fig. 10).
With about 4 kDa in size the P17 tag is much smaller than most other currently used protein tags. Bulkier tags impair size-related pharmacokinetic scFv properties and also increase the metabolic burden of the expression hosts. Furthermore, bulky tags such as MBP may sterically interfere with the scFv-antigen interaction since MBP-G12-scFv fusion protein displayed significantly lower binding affinity to recombinant HBV small surface protein than untagged G12-scFv-HA (Supplementary Fig. S5). Hence larger tags often need to be proteolytically removed, as shown by virtually identical band intensities of equal amounts of untreated vs. enterokinase-treated protein after centrifugation at 150,000 × g for 30 min (Supplementary Fig. S3c). These data also suggest that the P17 tag facilitates the adoption of a stably folded scFv conformation, as practically supported by the much fewer insoluble aggregates observed after 1.5 years of cryo-preservation of P17-tagged scFvs. Hence to our knowledge, P17 tag is the first reported peptide tag to increase the thermostability of scFvs.

As reported, P17 peptide targeted multiple cargos including proteins, siRNA, and DNA-liposome complex to hepatocytes by interaction with the receptor LRP on cell surface, leading to endocytosis of cargos by hepatocytes. Indeed, P17-tagged G12-scFv was taken up by primary mouse hepatocytes more efficiently than untagged G2-scFv, as observed by immunofluorescence confocal microscopy (Supplementary Fig. S6). In this regard, besides increasing solubility and production of scFv, P17 tag can specifically augment scFv’s effect in hepatocytes by increasing intracellular concentration. In our previous study, G2-scFv targeting hepatitis B surface antigen could be endocytosed by hepatocytes and potently inhibited secretion of HBV virions. Hence a stronger inhibition by P17-tagged G12-scFv than untagged version can be envisaged due to stronger antigen-binding affinity as well as more efficient uptake.

**Mechanisms by which P17 tag increases the solubility of scFv in SHuffle strain and scFv stability**

Many current protein tags and peptide tags, with molecular weights of 4.5–55 kDa and 0.15–1.6 kDa, respectively, have been developed for enhancing the solubility of proteins of interest. As reported, the P17 tag not only promotes expression solubility in E. coli SHuffle strain but also increased thermostability and antigen binding affinity of G12-scFv. In this way, the P17 tag could benefit long-term storage stability of scFvs, as practically supported by the much fewer insoluble aggregates observed after 1.5 years of cryo-preservation of P17-tagged scFvs. Hence to our knowledge, P17 tag is the first reported peptide tag to increase the thermostability of scFvs.
ability than P17-N18del instead (Fig. 7d). At physiological pH, the P17 tag, VRC01-scFv, and G12-scFv with predicted pIs of 9.82, 8.84, and 8.35 have positive charges, but MA18/7-scFv and ADRI-scFv with pIs of 5.52 and 5.53 have negative charges (Fig. 1d). Disregarding charge accessibility, the P17 tag increases net positive charges for VRC01-scFv, and G12-scFv, but reduces overall net charges of MA18/7-scFv and ADRI-scFv. Nonetheless, the P17 tag had a weaker solubility-enhancing effect on VRC01-scFv and G12-scFv than on MA18/7-scFv and ADRI-scFv (Fig. 1d), arguing against the supercharging model. Forcing P17-tagged scFv into micelle-like structures is also unlikely since SEC only revealed monomeric and dimeric forms of G12-scFv-HA-PI7 (Fig. 10a). The same holds for chaperone recruitment by the P17 tag because two or three copies of P17 tags did not increase but rather weaken the solubility enhancement of a single tag copy (Fig. 4). Though it is not excluded that P17 tag residues 22 to 29 adopt an α-helical secondary structure in P17-tagged scFvs, such an α-helix is dispensable for solubility enhancement, as evidenced by point mutation and deletion mutation analysis (Figs. 5 and 6d). We also found no positive evidence for an interaction of the P17 tag with RNA, which could contribute to chaperoning.

Higher thermostability, as well as stronger antigen affinity and virus neutralization activity of P17-tagged versus untagged G12-scFv-HA, suggest that the P17 tag promoted correct folding of scFv via an intramolecular chaperone-like activity. Type I intramolecular chaperones are currently defined. Type I intramolecular chaperones assist tertiary structure formation at an early stage of protein folding by acting as the folding nucleus or by preventing misfolding and aggregation, while type II intramolecular chaperones are not directly involved in protein folding but play key roles in the formation of quaternary structure. P17 tag, monomeric on its own, did not alter the quaternary conformation of scFv (Fig. 10). Moreover, excess P17 peptide added in trans post purification did not increase thermostability of G12-scFv (Fig. 5b). All these data suggested that P17 tag functions as a HA tag inserted between scFv and P17 tag, DNA coding for HA-P17 fused with a single tag copy (Fig. 4). Though it is not informed consent from the donors.

Ruijin Hospital Ethics Committee (EC approval No. 2016-124) and got informed consent from the donors. The research use of these sera has been approved by the Ruijin Hospital Ethics Committee (EC approval No. 2016-124) and got informed consent from the donors.

**Plasmid constructs**

pET28a-His-G12-scFv-HA, pET28a-His-MA18/7-scFv-HA, and pET28a-His-VRCo1-scFv-HA plasmids expressing G12-scFv, MA18/7-scFv, and VRC01-scFv with N and C termini attached respectively by hexahistidine and HA tags have been described. pET28a-His-ADRI-scFv-HA was constructed by replacing the G12-scFv open reading frame (ORF) of pET28a-His-G12-scFv-HA with a customized ADRI-scFv ORF via digestion with NdeI and BamHI restriction endonucleases. To construct pET28a-His-G12-scFv-P17 and pET28a-His-MA18/7-scFv-P17 plasmids expressing G12-scFv-P17 and MA18/7-scFv-P17 proteins with the P17 tag directly fused to their termini, the complete G12-scFv-P17 and MA18/7-scFv-P17 ORFs were separately amplified by two runs of PCR. Using pET28a-His-G12-scFv-HA and pET28a-His-MA18/7-scFv-HA as the templates, the first-run PCR products were amplified separately by two pairs of primers: G12-VL-NdeI-fw plus P17-rev1, 5'-aaacctctggtctctgatcgcgtccacctctggtctcatcgccatgccatggacatcgagctgacccagag-3', and MA18/7-VL-NdeI-fw plus P17-rev1, 5'-ttggacattgactgactggatgccataggctcaagctggcccctctgtggctg-3', and then inserted into pET28a vector.

**Methods**

**Ethical statement**

The sera of CHB patients were obtained from Drs. Demin Yu and Xinxin Zhang from Ruijin Hospital of Shanghai JiaoTong University School of Medicine. The research use of these sera has been approved by the Ruijin Hospital Ethics Committee (EC approval No. 2016-124) and got informed consent from the donors.
To construct pET28a-His-G12-scfv-HA-PI7-N8del, pET28a-His-G12-scfv-HA-PI7-N14del, and pET28a-His-G12-scfv-HA-PI7-N18del plasmids expressing G12-scfv-HA fusion proteins attached by deletion mutants of PI7 tag with N-terminal 8 aa and 15 aa deleted, G12-scfv-HA-PI7-C5del and G12-scfv-HA-PI7-C15del genes were amplified separately by PCR using pET28a-His-G12-scfv-HA-PI7 as the template and two primers of PI7-Ndel-flwd plus PI7-C5del-XhoI-rev: 5'-ggagatccggatccggagtcaagggtttccgagacgaagccgccgccttcaagaatacggctggt-3' and PI7-Ndelflwd plus PI7-C15del-XhoI-rev: 5'-ggagatccggatccggagtcaagggtttccgagacgaagccgccgccttcaagaatacggctggt-3' and inserted into pET28a vector by digestion with Ndel and XhoI endonucleases.

To construct pET28a-His-G12-scfv-HA-PI7-N8del, pET28a-His-G12-scfv-HA-PI7-N14del, and pET28a-His-G12-scfv-HA-PI7-N18del plasmids expressing G12-scfv-HA fusion proteins attached by deletion mutants of PI7 tag with N-terminal 8 aa, 14 aa, and 18 aa deleted, G12-scfv-HA-PI7-N8del, G12-scfv-HA-PI7-N14del, and G12-scfv-HA-PI7-N18del genes were amplified separately by PCR using pET28a-His-G12-scfv-HA-PI7 as the template and three pairs of primers: PI7-N8del-BamHI-flwd: 5'-gcaagcggagccggagctagccaggggtttccgagacgagccgccgc-3' plus pET28a-DraIII-rev, PI7-N14del-BamHI-flwd: 5'-gcaagcggagccggagctagccaggggtttccgagacgagccgccgc-3' plus pET28a-DraIII-rev, and PI7-N18del-BamHI-flwd: 5'-gcaagcggagccggagctagccaggggtttccgagacgagccgccgc-3' plus pET28a-DraIII-rev, cleaved by BamHI and DraIII endonucleases, and then ligated with pET28a-His-G12-scfv-HA PI7 vector digested by BamHI and DraIII enzymes.

To construct pET28a-His-G12-scfv-HA-PI7-N18del-DE, pET28a-His-G12-scfv-HA-PI7-N18del-DE, cleaved by NdeI and BamHI digestion to yield the total scFv protein. After centrifugation of bacterial lysate at 12,000×g at 4 °C for 20 min, 10 μl clear supernatants representing soluble fraction were mixed with 40 μl 2× SDS-PAGE loading buffer for western blotting analysis of total scFv protein. Soluble scFv protein in the supernatant was purified by immobilized metal ion affinity chromatography (IMAC) with an Ni-nitrilotriacetic acid (Ni-NTA) superflow column (Qiagen), followed by size-exclusion chromatography on a Superdex 200 Increase 10/300 column (GE Healthcare). ScFvs were finally dialyzed against PBS and concentrated by Amicon centrifugal ultrafiltration devices (Millipore) with a molecular cutoff of 3 kDa.

Western blotting

Proteins mixed with 2× SDS-PAGE loading buffer were separated by electrophoresis on a SDS-12.5% PAGE, and electro-transferred onto a polyvinylidene fluoride membrane in the Towbin buffer (25 mM Tris base, 192 mM glycine, pH 8.3, 0.1% SDS, and 20% methanol) at 100 V for 90 min. After blocking the membrane with TBST buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 0.1% Tween 20) containing 5% skim milk, scFv was detected by sequential incubations with 1000-fold diluted mouse monoclonal Ab against hexahistidine tag (ProteinTech, catalog no. 66005-1-lg) and 5000-fold diluted horseradish peroxidase-conjugated goat anti-mouse secondary antibody (Jackson Immunoresearch Laboratories), followed by reaction with chemiluminescent substrates (PerkinElmer).

SEC

Two hundred μl protein markers (catalog No: 69385, Sigma) or test scFvs diluted in PBS were loaded onto a Superdex 200 Increase 10/300 column using AKTA Puri system (GE Healthcare). Proteins eluted by PBS at a flow rate of 0.5 ml/min were monitored via ultraviolet absorbance at 280-nm wavelength. According to the user manual of the column, the average distribution constant (Kd) of test protein was determined according to the equation Kd = (Vv - Vg)/(Vv - V0) using the parameters: elution volume (Vv) of test protein, void volume (V0) and bed volume (Vc) of the column. The calibration curve of SEC was plotted as Kd vs. molecular weight of the proteins using the linear fitting program of Origin 8.0 software.
Thermal shift assay
Thermal shift assays were performed as previously described\(^1\) with minor modifications. SYPRO Orange dye was firstly mixed with 5 µM final concentration of scFv according to user manual (ThermoFisher Scientific). An optical 96-well reaction plate (Applied Biosystems) loaded with 20 µl reaction mixture per well was mounted onto a real-time PCR instrument (QuantStudio 6 Flex, Applied Biosystems). The mixture was heated up by 0.05 °C/s from 30 to 90 °C, and the fluorescence signal in 570-nm wavelength yielded by the dye was recorded concomitantly. Melting temperature (Tm) for each scFv was determined by fitting fluorescence signal and temperature to the Boltzmann equation using Protein Thermal Shift software V3.1 (ThermoFisher Scientific).

ELISA
The assay was done as previously described\(^6\). Microwell plates coated respectively with 5 µg/ml G12-scFv-HA, G12-scFv-HA-P17, milk, or 10 µg/ml MBP-G12-scFv solutions were incubated with PBS diluted sera of healthy donors (GemCell, catalog No. 100-SI2) and CHB patients at 37 °C for 1 h, and subsequently reacted with peroxidase-conjugated anti-S (Shanghai Kehua Bio-Engineering Company, KHB). After extensive washing, the plates were developed using 3, 3’, 5’, 5’-tetramethylbenzidine (KHB) as substrate with subsequent stopping by sulphuric acid. Absorbance values at 450 nm (A450 nm) were recorded by an ELISA plate reader (Biorad).

Production of HBV, neutralization assay, and northern blotting
These assays were performed as previously reported\(^6\). Briefly, HBV particles secreted by HepAD38 cells were precipitated with 7.5% PEG8,000 solution and quantitated by real-time quantitative PCR using HBV-specific primers. HBV at a nominal multiplicity of infection of 100 vge/cell was incubated with various concentrations of scFv at room temperature for 30 min, and then was used to inoculate 1.5 × 10\(^4\) HepG2-NTCP cells per well grown on a collagen-coated 24-well plate in the presence of 4% PEG8,000. After 16 h, cells were extensively washed by 300 µl PBS per well for five times and cultured with DMEM containing 2.5% DMSO until harvest. At day 6 post HBV infection, hepatitis B e antigen (HBeAg) secreted by HBV-infected cells in the culture supernatants was determined by enzyme-linked immunosorbent assay (ELISA) following the user manual. Total RNA was extracted from HBV-infected cells using TRIZol reagents (ThermoFisher) according to the manufacturer’s protocol. HBV RNA was detected by northern blotting using HBV-specific 32P-labeled DNA probe\(^6\). Densitometry of hybridization signals was done using MultiGauge V2.2 (Fuji Film).

The HepG2-NTCP cell line was previously established\(^8\) and here authenticated by morphology, western blotting of NTCP protein, or PCR with NTCP-specific primers. HepAD38 cell line was provided by Dr. Christoph Seeger (Fox Chase Cancer Center, Philadelphia, USA) and authenticated by morphology and PCR with HBV-specific primers.

Immunofluorescence microscopy
The assay was conducted as previously reported\(^8\). Primary mouse hepatocytes grown on collagen-coated coverslips were fixed with 4% paraformaldehyde and permeabilized with 0.25% Triton X-100. After blocking with 3% bovine serum albumin (fraction V, Sigma) for 10 min, cells were incubated with the respective primary antibodies followed by Alexa Fluor 488 or Cy3 conjugated secondary antibodies. Nuclear DNA was stained with 4’, 6-diamidino-2-phenylindole (Sigma). Cells were visualized by a laser confocal scanning microscope (Leica TCS SP8). Image processing was done by Las AF Lite 2.6 software (Leica).

Statistics and reproducibility
Results acquired by three independent experiments were presented as average values ± standard deviations. Statistical significance of differences between two experimental groups was assessed one-way analysis of variance as implemented in Origin 8.0 software. Fifty percent inhibitory concentrations of scFvs were also calculated by Origin 8.0 software.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Source data are provided with this paper.

References
1. Ryman, J. T. & Meibohm, B. Pharmacokinetics of monoclonal antibodies. CPT Pharmacomet. Syst. Pharm. 6, 576–588 (2017).
2. Gil, D. & Schrum, A. G. Strategies to stabilize compact folding and minimize aggregation of antibody-based fragments. Adv. Biosci. Biotechnol. 4, 73–84 (2013).
3. Li, Z. et al. Influence of molecular size on tissue distribution of antibody fragments. Mabs 8, 113–119 (2016).
4. Morell, A., Terry, W. D. & Waldmann, T. A. Metabolic properties of IgG subclasses in man. J. Clin. Invest. 49, 673–680 (1970).
5. Chatenoud, L. & Bluestone, J. A. CD3-specific antibodies: a portal to the treatment of autoimmunity. Nat. Rev. Immunol. 7, 622–632 (2007).
6. Goldenberg, D. M. & Sharkey, R. M. Novel radiolabeled antibody conjugates. Oncogene 26, 3734–3744 (2007).
7. Kenanova, V. & Wu, A. M. Tailoring antibodies for radiounclide delivery. Expert Opin. Drug Deliv. 3, 53–70 (2006).
8. Bang, T. H. & Seong, B. L. Solubility, stability, and avidity of recombinant antibody fragments expressed in microorganisms. Front. Microbiol. 11, 1527 (2020).
9. Worn, A. & Pluckthun, A. Stability engineering of antibody single-V Fv fragments. J. Mol. Biol. 305, 989–1010 (2001).
10. Glocshuber, R., Malia, M., Pfitzinger, I. & Pluckthun, A. A comparison of strategies to stabilize immunoglobulin Fv-fragments. Biochemistry 29, 1362–1367 (1990).
11. Bird, R. E. et al. Single-chain antigen-binding proteins. Science 242, 423–426 (1988).
12. Young, N. M., MacKenzie, C. R., Narang, S. A., Oomen, R. P. & Baenziger, J. E. Thermal stabilization of a single-chain Fv antibody fragment by introduction of a disulfide bond. FEBS Lett. 377, 135–139 (1995).
13. Holliger, P., Prospero, T. & Winter, G. “Diabodies”: small bivalent and bispecific antibody fragments. Proc. Natl Acad. Sci. USA 90, 6444–6448 (1993).
14. Montoliu-Gaya, L., Martinez, J. C. & Villegas, S. Understanding the contribution of disulfide bridges to the folding and misfolding of an anti-Abeta scFv. Protein Sci. 26, 1138–1149 (2017).
15. Glocshuber, R., Schmidt, T. & Pluckthun, A. The disulfide bonds in antibody variable domains: effects on stability, folding in vitro, and functional expression in Escherichia coli. Biochemistry 31, 1270–1279 (1992).
16. Stewart, E. J., Aslund, F. & Beckwith, J. Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. EMBO J. 17, 5543–5550 (1998).
17. Bardwell, J. C. et al. A pathway for disulfide bond formation in vivo. Proc. Natl Acad. Sci. USA 90, 1038–1042 (1993).
18. Li, C. et al. An E. coli-produced single-chain variable fragment (scFv) targeting hepatitis B virus surface protein potently inhibited virion secretion. Antivir. Res. 162, 118–129 (2019).
19. Lobstein, J. et al. Shuffling, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Micro. Cell Fact. 11, 56 (2012).
20. Yusakul, G. et al. Bacterial expression of a single-chain variable fragment (scFv) antibody against ganoderic acid A: a cost-effective approach.

https://doi.org/10.1038/s41467-022-32423-9
approach for quantitative analysis using the scFv-based enzyme-linked immunosorbent assay. *Biol. Pharm. Bull.* **40**, 1767–1774 (2017).

21. Robinson, M. P. et al. Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nat. Commun.* **6**, 8072 (2015).

22. Hatahet, F., Nguyen, V. D., Salo, K. E., & Ruddock, L. W. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*. *Micro. Cell Fact.* **9**, 67 (2010).

23. Hatahet, F. & Ruddock, L. W. Topological plasticity of enzymes involved in disulfide bond formation allows catalysis in either the periplasm or the cytoplasm. *J. Mol. Biol.* **425**, 3268–3276 (2013).

24. Nguyen, V. D. et al. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of *E. coli*. *Micro. Cell Fact.* **10**, 1 (2011).

25. Gaciarz, A. et al. Systematic screening of soluble expression of antibody fragments in the cytoplasm of *E. coli*. *Micro. Cell Fact.* **15**, 22 (2016).

26. Gaciarz, A. & Ruddock, L. W. Complementarity determining regions and frameworks contribute to the disulfide bond independent folding of intrinsically stable scFv. *PloS ONE* **12**, e0189964 (2017).

27. Bach, H. et al. *Escherichia coli* maltose-binding protein as a molecular chaperone for recombinant intracellular cytoplasmic single-chain antibodies. *J. Mol. Biol.* **312**, 79–93 (2001).

28. Sarker, A., Rathore, A. S. & Gupta, R. D. Evaluation of scFv protein recovery from *E. coli* by in vitro refolding and mild solubilization process. *Micro. Cell Fact.* **18**, 5 (2019).

29. Sun, W. et al. A combined strategy improves the solubility of aggregation-prone single chain variable fragment antibodies. *Protein Expr. Purif.* **83**, 21–29 (2012).

30. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17**, 353–358 (2006).

31. Wong, S. C. et al. Hepatocyte targeting of nucleic acid complexes and liposommes by a T7 phase p7 peptide. *Mol. Pharm.* **3**, 386–397 (2006).

32. Sokoloff, A. V. et al. A new peptide ligand that targets particles and heterologous proteins to hepatocytes in vivo. *Mol. Ther.* **8**, 867–872 (2003).

33. Ludtke, J. J. et al. Pep tide-mediated targeting of hepatocytes via low density lipoprotein receptor-related protein (LRP). *Drug Deliv.* **16**, 268–273 (2009).

34. Wang, W. et al. A human monoclonal antibody against small envelope protein of hepatitis B virus with potent neutralization effect. *MAbs* **8**, 468–477 (2016).

35. Wu, X. et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856–861 (2010).

36. Cerino, A., Bremer, C. M., Glebe, D. & Mondelli, M. U. A human monoclonal antibody against hepatitis B surface antigen with potent neutralizing activity. *PloS ONE* **10**, e0125704 (2015).

37. Kuttner, G. et al. Characterization of neutralizing anti-pre-S1 and anti-pre-S2 (HBV) monoclonal antibodies and their fragments. *Mol. Immunol.* **36**, 669–683 (1999).

38. Waugh, D. S. Making the most of affinity tags. *Trends Biotechnol.* **23**, 316–320 (2005).

39. Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4**, 32 (2004).

40. Malhotra, A. Chapter 16 tagging for protein expression. In Methods in Enzymology (eds Burgess, R. R. & Deutscher, M. P.) *463*, 239–258 (Academic Press, 2009).

41. Kwon, S. B., Yu, J. E., Park, C., Lee, J. & Seong, B. L. Nucleic acid-dependent structural transition of the intrinsically disordered N-terminal appended domain of human lysyl-tRNA synthetase. *Int. J. Mol. Sci.* **19**, 3016 (2018).

42. Lim, J. et al. RNA-dependent assembly of chimeric antigen nanoparticles as an efficient H5N1 pre-pandemic vaccine platform. *Nanomedicine* **37**, 102438 (2021).

43. Hwang, B. J. et al. RNA-assisted self-assembly of monomeric antigens into virus-like particles as a recombinant vaccine platform. *Biomaterials* **269**, 120660 (2021).

44. Lee, J. et al. RNA-dependent chaperone (chaperna) as an engineered pro-region for the folding of recombinant microbial transglutaminase. *Biotechnol. Bioeng.* **116**, 490–502 (2019).

45. Costa, S., Almeida, A., Castro, A. & Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel FlhN system. *Front. Microbiol.* **5**, 63 (2014).

46. Paraksevopoulou, V. & Falcone, F. H. Polyionic tags as enhancers of protein solubility in recombinant protein expression. *Microorganisms* **6**, 47 (2018).

47. Steven, A. C. et al. Molecular substructure of a viral receptor recognition protein. The gp17 tail-fiber of bacteriophage T7. *J. Mol. Biol.* **200**, 351–365 (1988).

48. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).

49. Singh, H., Singh, S. & Singh Raghava, G. P. Peptide secondary structure prediction using evolutionary information. Preprint at https://www.biorxiv.org/content/10.1101/858791v1.full (2019).

50. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

51. Li, S. C., Goto, N. K., Williams, K. A. & Deber, C. M. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc. Natl Acad. Sci. USA* **93**, 6676–6681 (1996).

52. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).

53. Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. Protein Sci.* **79**, 29 28 21 29 28 15 (2014).

54. Kaplan, W. et al. Conformational stability of pOEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag. *Protein Sci.* **6**, 399–406 (1997).

55. Chen, Y. J. & Inouye, M. The intramolecular chaperone-mediated protein folding. *Curr. Opin. Struct. Biol.* **18**, 765–770 (2008).

56. Satheeshkumar, P. K. Expression of single chain variable fragment (scFv) molecules in plants: a comprehensive update. *Mol. Biotechnol.* **62**, 151–167 (2020).

57. Gupta, S. K. & Shukla, P. Microbial platform technology for recombinant antibody fragment production: a review. *Crit. Rev. Microbiol.* **43**, 31–42 (2017).

58. Young, C. L., Britton, Z. T. & Robinson, A. S. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* **7**, 620–634 (2012).

59. Beck, J. & Nassal, M. A Tyr residue in the reverse transcriptase domain can mimic the protein-priming Tyr residue in the terminal protein domain of a hepadnavirus P protein. *J. Virol.* **85**, 7742–7753 (2011).

60. Schroeder, H. W. Jr. & Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, 541–552 (2010).

61. Martin, C. D. et al. A simple vector system to improve performance and utilisation of recombinant antibodies. *BMC Biotechnol.* **6**, 46 (2006).

62. Raju, M., Santhoshkumar, P. & Krishna Sharma, K. Alpha-crystallin-derived peptides as therapeutic chaperones. *Biochim Biophys. Acta* **1860**, 246–251 (2016).
63. Suss, O. & Reichmann, D. Protein plasticity underlines activation and function of ATP-independent chaperones. *Front. Mol. Biosci.* **2**, 43 (2015).

64. Yabuta, Y., Subbion, E., Oiry, C. & Shinde, U. Folding pathway mediated by an intramolecular chaperone. A functional peptide chaperone designed using sequence databases. *J. Biol. Chem.* **278**, 15246–15251 (2003).

65. Mitra, R. et al. Mechanism of the small ATP-independent chaperone Spy is substrate specific. *Nat. Commun.* **12**, 851 (2021).

66. Goto, Y. & Hamaguchi, K. The role of the intrachain disulfide bond in the conformation and stability of the constant fragment of the immunoglobulin light chain. *J. Biochem.* **86**, 1433–1441 (1979).

67. Carey, J., Lindman, S., Bauer, M. & Linse, S. Protein reconstitution and three-dimensional domain swapping: benefits and constraints of covalency. *Protein Sci.* **16**, 2317–2333 (2007).

68. Wang, Y. X., Wen, Y. M. & Nassal, M. Carbonyl J acid derivatives block protein priming of hepadnaviral P protein and DNA-dependent DNA synthesis activity of hepadnaviral nucleocapsids. *J. Virol.* **86**, 10079–10092 (2012).

69. Wang, Y. X. et al. Interferon-inducible MX2 is a host restriction factor of hepatitis B virus replication. *J. Hepatol.* **72**, 865–876 (2020).

**Acknowledgements**

This study was in part supported by the National Natural Science Foundation of China (32170153, 31370195, and 81672017 to Y.X.W.), Outstanding Young Scholar Project (C850005 to Y.X.W.) of Shanghai Municipal Commission of Health and Family Planning, Open Research Fund of Key Laboratory of Tumor Microbiology, Fujian Medical University (to Y.X.W.), a fund from Shanghai Medical College and Zhongshan Hospital Immunotherapy Technology Transfer Center (to Y.X.W.), and grants from Science and Technology Commission of Shanghai Municipality (21ZK1474700 to S.C., 18431907100 and 19XD1404700 to C.L.).

Y.X.W. is indebted to the Alexander von Humboldt Foundation for a revisit fellowship that allowed short-term collaboration work at Dr. Michael Nassal’s lab. S.T. acknowledges the financial support from the National Science and Technology Major Project of China (2017ZX102020203-003-002). Z.Y. acknowledges the financial support from CAMS Innovation Fund for Medical Sciences (2019-I2M-5-040).

**Author contributions**

Y.W., W.Y., S.G., Q-Q.L., X.C., C.Li, and Q-Y.L. performed the experiments and collected data; S.C., Z.C., and L.S. supervised the study and revised the manuscript; M.N., Y.M.W., S.T., C.Luo, and Z.Y. revised the manuscript; S.T. shared the experimental facilities; Y.X.W. conceived, designed, and supervised the study; wrote and revised the manuscript.

**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/s41467-022-32423-9.

**Correspondence** and requests for materials should be addressed to Shijie Chen or Yong-Xiang Wang.

**Peer review information** *Nature Communications* thanks Baik L Seong, Shihua Wang and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at http://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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022