



OPEN Design of an alternate antibody fragment format that can be produced in the cytoplasm of *Escherichia coli*

Aatir A. Tungekar & Lloyd W. Ruddock ✉

With increased accessibility and tissue penetration, smaller antibody formats such as antibody fragments (Fab) and single chain variable fragments (scFv) show potential as effective and low-cost choices to full-length antibodies. These formats derived from the modular architecture of antibodies could prove to be game changers for certain therapeutic and diagnostic applications. Microbial hosts have shown tremendous promise as production hosts for antibody fragment formats. However, low target protein yields coupled with the complexity of protein folding result in production limitations. Here, we report an alternative antibody fragment format 'Fab_H3' designed to overcome some key bottlenecks associated with the folding and production of Fabs. The Fab_H3 molecule is based on the Fab format with the constant domains replaced by engineered immunoglobulin G1 (IgG₁) C_H3 domains capable of heterodimerization based on the electrostatic steering approach. We show that this alternative antibody fragment format can be efficiently produced in the cytoplasm of *E. coli* using the catalyzed disulfide-bond formation system (CyDisCo) in a natively folded state with higher soluble yields than its Fab counterpart and a comparable binding affinity against the target antigen.

Among biotherapeutics, the monoclonal antibodies (Mabs) segment holds a dominant market position, however this trend appears to be gradually changing with the development of novel antibody formats and new medicines¹. Full-length monoclonal antibodies have a long circulating half-life which might render them unsuitable for certain therapeutic and diagnostic applications². As a potential solution, antibody fragments (Fabs) have emerged as key players in the biopharmaceutical industry. They offer certain advantages such as improved and deep tumor penetration, binding to specific epitopes inaccessible to Mabs, monovalent antigen binding with potentially reduced immunogenicity, higher stability than smaller antibody fragment formats and faster clearance^{3–5}. A multitude of applications of antibody-based molecules as therapeutic and diagnostic agents have intensified their demand and subsequently the need for their production in high yields.

Many approved biosimilar and 'me-too' type products have been produced in microbial expression systems of which *E. coli* has remained a host of choice⁶. As Fab and scFv antibody fragments are non-glycosylated and relatively small, their production in *E. coli* can be simpler and more economically viable than the conventional mammalian cell culture systems. Antibody fragments are disulfide-bonded proteins which are traditionally recombinantly produced in the oxidizing periplasm or as inclusion bodies in the reducing cytoplasm of *E. coli*. However, both approaches pose critical bottlenecks in the efficient production of these proteins of interest. Periplasmic expression of Fabs often results in low protein yields, mainly due to inefficient protein translocation and the small volume of the periplasm, while cytoplasmic expression faces limitations in terms of protein degradation of the non-native state and aggregation to inclusion bodies. Of the seven marketed Fabs, two are produced in *E. coli* in the form of inclusion bodies^{3,7}. Fab folding is a complex process, involving disulfide-bond formation, cis–trans prolyl isomerization and controlled oligomerization, and hence refolding of inclusions bodies is often inefficient and costly⁸.

Fab fragments are composed of two covalently linked polypeptide chains, namely the heavy (HC) and light (LC) chains which respectively contain two constant domains C_H1 and C_L and two antigen-binding variable domains V_H and V_L (Fig. 1A). The C_H1 domain is an intrinsically disordered protein in isolation and remains in an unfolded state irrespective of the formation of disulfide bonds⁹. One of the critical rate-limiting steps in the folding of antibody fragments is the association of the constant domains as the C_H1 domain adopts the typical

Protein and Structural Biology Research Unit, Faculty of Biochemistry and Molecular Medicine, University of Oulu, 90220 Oulu, Finland. ✉email: lloyd.ruddock@oulu.fi

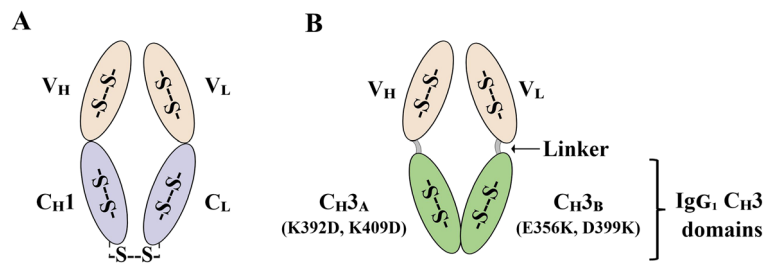


Figure 1. Schematic illustration of (A) Antibody fragment (Fab) and (B) Antibody fragment format with C_H3 domains (Fab_{H3}). Reversing symmetric charge complementarity at the C_H3 domain interface allows the suppression of homodimerization due to unfavourable repulsive charge interactions thereby allowing the formation of a Fab like heterodimer.

immunoglobulin (Ig) fold only on interaction with the C_L domain^{9,10}. In addition, low HC stability coupled with formation of covalent and non-covalently linked LC dimers warrants the need for balancing the synthesis of the two chains^{11,12}. Designing and screening constructs with varying translational strengths of the LC and HC for optimal Fab expression can be laborious, expensive and may not necessarily result in success. There have been several approaches to drive efficient heterodimerization in a Fab molecule, for example, Ojima-Kato et al. reported the fusion of leucine-zipper peptide pairs to the constant domains (Zipbody)¹³. However, this limits the therapeutic application of Fabs due to the requirement for tag cleavage steps and subsequent immunogenicity concerns.

Limitations associated with efficient Fab production raises the need for alternative antibody fragments which retain the beneficial properties of Fabs while overcoming their production bottlenecks. Here, we communicate an alternate antibody fragment format called ‘Fab_{H3}’ in which the constant domains of a Fab molecule, namely REGN10987 (PDB ID: 6XDG) against the receptor binding domain (RBD) of SARS-CoV-2, are replaced by IgG₁ heavy chain constant 3 (C_H3) domains (Fig. 1B). This removes the limitations associated with the C_H1 domain in Fabs. The rationale behind using IgG₁ C_H3 domains is their high solubility and stability coupled with their remarkable dimerization propensity^{14,15}. As heterodimerization is a critical attribute of Fabs, we modified the IgG₁ C_H3 domains using the previously established ‘Charge-to-charge swap design’ which allows efficient non-covalent heterodimerization of the C_H3 domains and allows the suppression of homodimerization though with a reduction in thermostability¹⁶. We report that the Fab_{H3} format can be produced solubly with comparatively higher yields than the wild type Fab molecule in the cytoplasm of *E. coli* in different strains and culture media. In addition, we show that the REGN10987-based Fab_{H3} format produced using the CyDisCo system is a natively folded heterodimer with a comparable affinity to its Fab counterpart against the SARS-CoV-2 RBD.

Results

Fab_{H3} design. A number of crystallizable fragment (Fc)-engineering approaches for the development of bispecific protein-based therapies based on C_H3 heterodimerization have been reported in the past few decades^{17–20}. For example, Wozniak-Knopp G. et al. reported the replacement of C_H1 and C_L domains with covalently linked C_H3 domains using the ‘knobs-into-holes’ (KiH) technology for a full-length trastuzumab molecule produced in HEK293-6E cells²¹. However, their resultant domain-exchanged monoclonal antibody showed low solubility and a four-fold weaker binding affinity to its target as compared to the parent trastuzumab. As well as the KiH technology to enable C_H3 heterodimerization, the other widely used method is the asymmetric charge polarity approach. We chose to adopt this to test the Fab_{H3} format as it is reported to suppress homodimerization more efficiently than the ‘knobs-into-holes’ without the introduction of an additional disulfide bond^{16,22} as this may increase the oxidative folding related stress on the production system as well as increase the misfolding and aggregation propensity²³.

It has been shown that the IgG₁ C_H3 domain interface consists of charged residues that interact through favorable electrostatic interactions surrounding a central hydrophobic core^{16,24}. As the hydrophobic core plays an important role in protein folding and stability²⁵, the electrostatic steering approach involves reversing the charge polarity on the rim of the core at the interface of the C_H3 domains to favor heterodimerization over homodimerization¹⁶. In the Fab_{H3} construct we have designed, mutations K392D, K409D were introduced into the C_H3 domain fused to the V_H domain and E356K, D399K were introduced into the C_H3 domain fused to the V_L domain (nomenclature based on¹⁶; PDB ID: 1L6X). To ensure sufficient flexibility and folding of the variable domains and to minimize steric hindrance during C_H3 domain dimerization, the V_H and V_L domains of the target Fab were fused to the C_H3 domains via flexible linkers composed of glycine and serine residues of varying length.

Soluble expression of wild-type REGN10987 Fab and Fab_{H3}. As the interfaces between the V_L/V_H and C_H3 domains are non-native, there is the potential for steric or other repulsions between them. While in silico modelling based on published structures suggested a linker as short as four amino acids could work, we wanted to experimentally validate the optimal inter-domain linker length. To do this, we compared the soluble expression of wild type REGN10987 Fab against Fab_{H3} constructs with varying linker lengths between the variable domains and C_H3 domains from 4 to 17 amino acids in length (amino acid sequences in Supplementary Table S1).

Since all the proteins of interest are disulfide-bonded proteins, their soluble expression was tested in the presence of CyDisCo components in the cytoplasm of *E. coli*. CyDisCo is a catalyzed cytoplasmic disulfide formation system allowing the production of disulfide-bonded recombinant proteins in *E. coli*^{26,27}. The sulfhydryl oxidase Erv1p (*S. cerevisiae*) catalyzes the oxidation of thiols on the protein to form disulfides and the isomerase PDI (*H. sapiens*) brings about the isomerization of non-native disulfides thus allowing the protein to be recombinantly produced in a natively folded disulfide-bonded state in the *E. coli* cytoplasm²⁸.

SDS-PAGE analysis of immobilized metal affinity chromatography (IMAC) based purified proteins showed that the Fab_{H3} is efficiently produced in the form of a heterodimer with the two chains being present in a 1:1 ratio (Fig. 2). The Fab_{H3} format was produced in comparatively higher soluble yields than the REGN10987 wild-type Fab without any further process optimization (Fig. 2). The linker length did not influence the soluble yields of the Fab_{H3} format or the ratio between the chains. The linker length also did not influence the thermal stability of Fab_{H3} (as determined by Nano Differential Scanning Fluorimetry) as all the variants displayed a similar melting temperature (T_m) (Supplementary Fig. S1). As the linker length did not influence the solubility or the thermal stability of the Fab_{H3} format, all further analysis was performed on the Fab_{H3} containing GS-G₄ as the linker. We chose GS-G₄ as the linker length as long linkers may give too much structural flexibility which can be non-optimal for downstream applications and too short linkers may result in steric or other inter-domain repulsions for Fab_{H3} molecules based on other Fabs.

To confirm that the increase in soluble yields arises from the format design i.e., removal of limitations associated with the C_{H1} domain, and not the expression conditions, we compared the CyDisCo-based soluble expression of Fab_{H3} (GS-G₄ linker) to the Fab counterpart between an *E. coli* B and K strain (BL21 (DE3) and MG1655 respectively) in rich autoinduction media. Our results demonstrate that whereas soluble yields of the Fab_{H3} format are comparable between the two *E. coli* strains tested, they are circa two-fold higher than those of the Fab counterpart in both strains (Table 1). We also examined the CyDisCo-based soluble production of Fab_{H3} and Fab in *E. coli* BL21 (DE3) in chemically defined minimal autoinduction media. The wild type REGN10987 Fab could not be produced solubly while the Fab_{H3} format was produced in a soluble form, with comparatively lower yields as compared to the rich media (Table 1). These findings suggest that the higher soluble yields observed for the Fab_{H3} format are a result of the format design potentially overcoming folding and production constraints of the Fab format regardless of the expression conditions.

Biophysical characterization of the Fab_{H3} format. Production of disulfide-bonded therapeutic proteins in the cytoplasm of microbial hosts such as *E. coli* is often challenging due to its limited post-translational modification capabilities which often leads to protein degradation and aggregation²⁹. Although employing the CyDisCo system in the *E. coli* cytoplasm allowed the production of the REGN10987 Fab and Fab_{H3} in a soluble form, with the latter in comparatively higher yields, developing an antibody format where quantity meets quality is of paramount importance. We evaluated the folding state of the antibody formats produced using a set of orthogonal analytical characterization tools prior to their functional characterization.

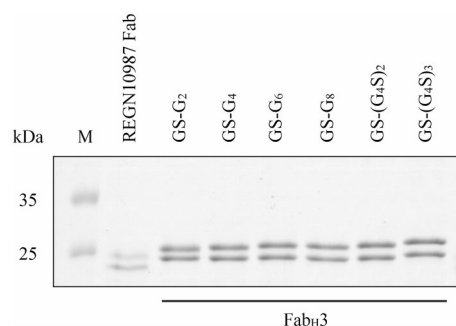


Figure 2. SDS-PAGE analysis of IMAC-purified REGN10987 Fab and Fab_{H3} produced solubly using CyDisCo in the cytoplasm of *E. coli*. Fab_{H3} variants with different linker lengths (G: Glycine, S: Serine) were compared to the wild type Fab to determine the influence on soluble protein yields (M: Protein Marker). See Supplementary Fig. S2 for the uncropped gel image.

Culture media	Strain	REGN10987 Fab (mg/L)	REGN10987 Fab _{H3} (mg/L)
Rich autoinduction media	BL21 (DE3)	64 ± 4	143 ± 4
	MG1655	64 ± 3	128 ± 1
Chemically defined minimal autoinduction media	BL21 (DE3)	–	46 ± 3

Table 1. Yields of REGN10987-Fab and Fab_{H3} from 20 mL cultures in shake flasks. The protein yields are expressed in mg/L of culture media. See Supplementary Figs. S5 and S6 for uncropped SDS-PAGE gel images used for protein quantification.

The secondary structure of the purified REGN10987 Fab and Fab_{H3} (GS-G₄ linker) was examined using circular dichroism (CD). As both the proteins of interest are composed of domains that possess the characteristic immunoglobulin (Ig)-fold, the CD spectra of these proteins were expected to exhibit shapes typical for proteins with a high content of β structure³⁰. CD analysis indicated that both the proteins exhibited far-ultraviolet (UV) CD spectra with a minima around 217 nm which is consistent with having a characteristic Ig-fold (Fig. 3A,B). The differences in the spectra can be explained as spectra of different Fab molecules have been shown to differ in terms of amplitude, shape and interception of the baseline based on aromatic side chain and/or disulfide bond chromophores^{31–33}.

Further analysis by electrospray ionisation mass spectrometry (ESI-MS) of the purified REGN10987 Fab and Fab_{H3} confirmed that the proteins had the expected molecular weights (Table 2) consistent with having all cysteines in disulfide bonds. To further evaluate the presence of any free cysteines, the proteins were treated with N-ethyl maleimide (NEM) under denaturing conditions prior to mass spectrometric analysis. No increase in mass corresponding to NEM binding (+ 125 Da) was observed for the proteins of interest, thereby confirming the absence of free thiols (Table 2). A combination of CD and ESI-MS data suggests that the proteins produced using CyDisCo in the cytoplasm of *E. coli* are natively folded.

The solution behaviour of purified REGN10987-Fab and Fab_{H3} was analyzed using SEC-MALS to determine the presence of native dimers as well as the absence of any degradants and high-molecular weight oligomeric species. The molecular weight determination of the two proteins under non-denaturing conditions confirmed the presence of both in a dimeric state with uniform molar mass points calculated across the elution peak to within 2% or less indicating monodisperse species (Supplementary Fig. S4A). The results obtained also demonstrate the absence of any monomeric fragments or oligomeric aggregates (Supplementary Fig. S4B).

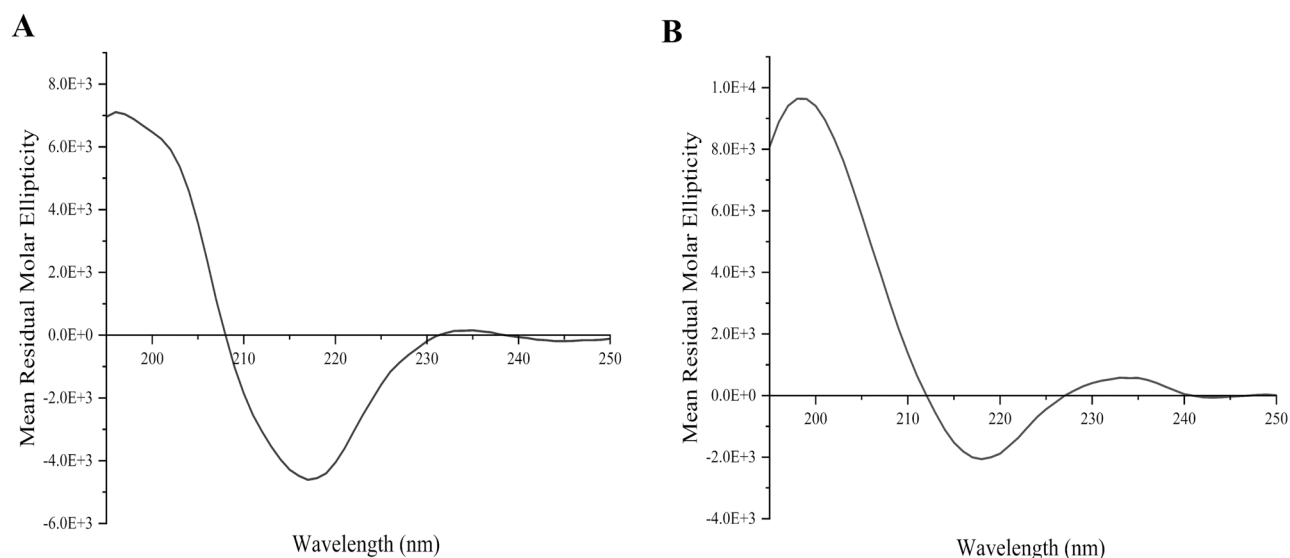


Figure 3. Far-ultraviolet Circular Dichroism (CD) spectra of purified proteins (A) REGN10987 Fab (B) REGN10987-based Fab_{H3}. Both proteins show spectra consistent with their expected β -rich structure of immunoglobulin domains i.e., Ig-fold.

Protein of Interest	Sample preparation	No. of cysteines	M_{theor} (Da) (oxidized)	M_{exp} (Da)	Δ mass ($M_{\text{theor}} - M_{\text{exp}}$)
REGN10987 Fab	- NEM	10	47,892	47,894	- 2
	+ NEM	10	47,892	47,892	0
REGN10987-based Fab _{H3}	- NEM	LC: 4	24,042	24,042	0
		HC: 4	26,618	26,619	- 1
	+ NEM	LC: 4	24,042	24,043	- 1
		HC: 4	26,618	26,618	0

Table 2. Molecular weight analysis of purified REGN10987 Fab and Fab_{H3} by electrospray ionization mass spectrometry (ESI-MS) in the presence and absence of NEM under denaturing conditions. Δ mass accounts for the mass difference between the theoretical molecular weight with disulfides formed (M_{theor}) and the experimental molecular weight (M_{exp}). None of the samples analyzed showed NEM binding (+ 125 Da) which suggests the absence of free thiols. The light (LC) and heavy (HC) chains of the Fab_{H3} molecule were detected as single chains instead of as a dimer since there is no covalent linkage between the chains.

The wild type IgG₁ C_H3 domain is widely known to be a highly thermostable immunoglobulin domain^{34,35}. However, heterodimerization is essential for a functional Fab-analogue and introducing mutations in the C_H3 domain to allow heterodimerization in bispecific monoclonal antibodies has previously been shown to result in a reduction in the thermal stability^{16,17,21,24}. We examined the thermal stability of the Fab_H3 format produced in the cytoplasm of *E. coli* using Nano Differential Scanning Fluorimetry (NanoDSF) and it was found to be in accordance with that previously reported for the electrostatic steering based bispecific scFv-Fc fusion protein molecule produced in a mammalian expression system¹⁶. A co-operative unfolding was observed for the Fab_H3 format in contrast to the non-cooperative folding observed for its Fab counterpart (Fig. 4) in accordance with previously reported unfolding characteristics of λ-Lc containing Fabs³⁶. Biphasic Fab unfolding involves the loss of antigen binding activity prior to the secondary structure breakdown³⁷. As observed from the thermal unfolding curves, onset of unfolding of the REGN10987 Fab occurs comparatively much earlier than the melting temperature (Fig. 4). Even though the Fab_H3 format was found to possess a comparatively lower final T_m than the Fab, it is still within a range that allows viable therapeutic protein production as the thermal stability is in accordance with Amgen's patented technology for generation of bispecific antibodies by electrostatic steering^{16,38}.

In vitro functional characterization of the Fab_H3 format. One of the critical quality attributes to be investigated when developing an alternate antibody format is the interaction of the proposed molecule against the target antigen. The binding ability of an antibody is a function of the Complementarity Determining Regions (CDRs) located in the variable domains which in turn depends on the stability and association of the domain interfaces³⁹.

We have previously shown that the wild type REGN10987 Fab produced using the CyDisCo system is functional and binds to the target antigen SARS-CoV-2 RBD⁴⁰. After ensuring that the REGN10987-based Fab_H3 produced in the cytoplasm of *E. coli* was natively folded, we empirically investigated whether the molecule developed was functionally active and compared its binding affinity with that of the REGN10987 Fab using Bi-layer Interferometry (BLI). The binding affinity between the Fab or Fab_H3 against the target antigen was assessed based on the ratio of the dissociation (k_d) and association (k_a) rate constants i.e., as a function of the equilibrium dissociation constant (K_D). Five concentrations each of the Fab and Fab_H3 were tested in parallel for their binding to SARS-CoV-2 RBD and the K_D values were determined using a global fit analysis based on a 1:1 interaction model.

Both Fab and Fab_H3 displayed binding to the target antigen in a concentration dependent manner (Fig. 5A,B). The Fab_H3 was found to exhibit a comparable binding affinity ($K_D = 20 \pm 2.6$ nM) to its counterpart REGN10987 Fab ($K_D = 43 \pm 4.4$ nM) against SARS-CoV-2 RBD. The K_D values observed for the REGN10987 Fab are in accordance with previous reports^{40,41} thereby validating our finding. These results suggest that the replacement of constant domains in a Fab molecule and the C_H3 domain-based heterodimerization does not interfere with the folding and function of the variable domains such that interactions with its antigen are maintained.

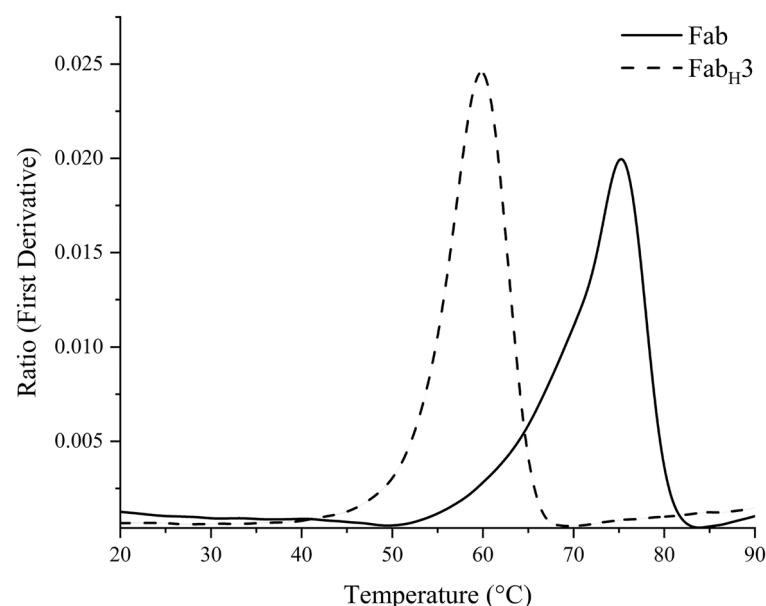


Figure 4. Thermal stability analysis of REGN10987 Fab and REGN10987-based Fab_H3 using NanoDSF. The onset points for REGN10987 Fab and Fab_H3 were found to be 50.0 °C and 42.5 °C respectively, while the final T_m values were found to be 75.1 °C and 59.9 °C respectively.

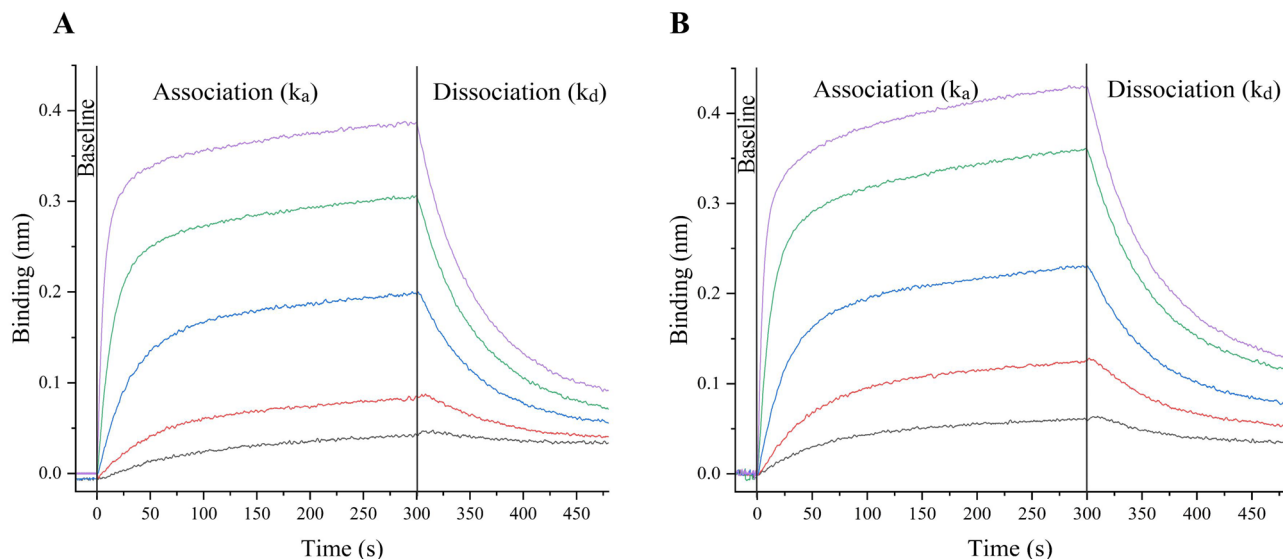


Figure 5. Bi-layer Interferometry (BLI)-based analysis of the binding interaction between (A) REGN10987 Fab, (B) REGN10987-based Fab_{H3} and the receptor binding domain (RBD) of SARS-CoV-2. Both the antibody formats tested show concentration-dependent binding to the target antigen (Grey: 3 nM, Red: 9 nM, Blue: 27 nM, Green: 81 nM, Purple: 243 nM). The steady-state analysis curve for the antibody-antigen interaction as a function of protein concentration is shown in Supplementary Fig. S3.

Discussion

Advances in synthetic biology leading to cellular engineering of microbial hosts have resulted in an increased acceptance of these cell factories as an alternative for producing antibody-based therapeutics⁴². While monoclonal antibodies have dominated treatment strategies against life-threatening diseases, advances in antibody engineering have led to an array of around 100 different antibody formats (see^{43–45} for comprehensive reviews on bispecific antibody formats and applications). There is no consensus on ‘one ideal format’ that is suitable for most of the desired applications, but instead the development of new formats offers a valuable source to overcome therapeutic and production bottlenecks of the current formats.

Fab antibody fragments are one of the earliest small antibody molecules studied and have demonstrated an important role in diagnostic and therapeutic interventions⁴⁶. With the increasing demand and future potential of Fabs or their fusion proteins, there is a constant need for the ‘developability’ of this format to be produced efficiently. Building on the advances made in the development of bispecifics, the Fab_{H3} format that we introduce here has the potential to overcome certain critical limitations of Fab production in *E. coli*. The previously proven IgG₁ C_{H3} domain heterodimerization strategy used here allows the production of heterodimeric molecules thereby preventing the light chain homodimerization limitations in wild type Fab molecules^{11,47}. The limitations associated with the low intrinsic stability and degradation tendency of the C_{H1} domain in Fabs have been known to result in low-yield production of Fabs in *E. coli*^{3,10,48}. These bottlenecks are potentially overcome with the strong dimerization propensity, high solubility, and stability of the IgG₁ C_{H3} domains used in the proposed Fab_{H3} format. This supposition is validated by the higher soluble production levels of the Fab_{H3} format than the corresponding Fab counterpart. In addition, the Fab_{H3} format is biologically active and shows a comparable binding affinity to the Fab format against the target antigen. Although the wild type IgG₁ C_{H3} domain has been reported to exhibit redox heterogeneity when produced using CyDisCo in the cytoplasm of *E. coli*¹⁴, the suppression of homodimerization potentially slows down the rate of folding which allows the formation of disulfide bonds and thereby native folding. This is not the first report to use IgG C_{H3} homodimers or heterodimers as scaffolds for the production of antibody formats, for e.g., scDb-C_{H3} (KiH), Di-Diabody, Minibody, etc.^{49–51}. However, to the best of our knowledge, this is the first report of an alternative Fab format based on C_{H3} domain heterodimerization, one that can be produced more efficiently using the CyDisCo system in the cytoplasm of *E. coli*.

Based on the ‘proof-of-concept’ reported here, this format design can be a potential solution to Fabs which face production bottlenecks. However, the general applicability of this format has to be devised empirically as it depends on the interface interactions between the variable and C_{H3} domains, and the solubility of the variable domain partners. Potential future work can involve engineering of the Fab_{H3} format to improve the thermal stability based on the previously reported ZW heterodimer variants⁵². Furthermore, protein engineering to expand the applicability of the Fab_{H3} format building on Fab-fusion proteins for a myriad of therapeutic and diagnostic applications can also be undertaken^{53,54}. Mutations in the IgG₁ C_{H3} domain have been shown to promote Fcγ receptor interaction in the absence of glycosylation⁵⁵. Being able to be produced in a microbial host, these mutations can potentially be applied to the Fab_{H3} format to increase half-life and bring about effector functions for certain therapeutic applications⁵⁶.

Methods

Cloning. All the expression vectors reported here were constructed using standard molecular biology techniques (see Supplementary Table S2 for vectors used in this study). A polycistronic gene for the REGN10987 wild type Fab was synthesized codon optimized (GenScript Biotech Corp.) for *E. coli* expression. The gene for the heavy chain with a C-terminal hexahistidine tag was placed downstream of the gene for the light chain and cloned using *Xba1-EcoRI* based restriction digestion and ligation in a modified pET23-based vector with the T7 promoter replaced by a pTac promoter to generate vector pAAT50^{26,40}. Genes for the heavy and light chains of REGN10987-based Fab_{H3} variants with different linker lengths were synthesized codon optimized (GenScript Biotech Corp.) for *E. coli* expression. These genes were cloned in the same backbone vector using *Nde1-EcoRI* based restriction digestion and ligation. To generate a polycistronic vector with a similar placement of genes as the wild type Fab, the gene for the V_H domain fused to an IgG₁ C_{H3} domain (K392D, K409D) with a C-terminal hexahistidine tag was placed downstream of the gene for the V_L domain fused to an IgG₁ C_{H3} domain (E356K, D399K). The gene for the Fab_{H3} heavy chain was cloned into the vector containing the Fab_{H3} light chain using *Xba1/Spe1-Xho1* based restriction digestion and ligation to generate vectors pAAT202–pAAT207. Purified plasmid vectors were obtained using the E.Z.N.A Plasmid DNA Mini Kit I (Omega Bio-Tek Inc.) and purification of DNA from agarose gels was performed using the Gene/PCR DNA Fragments Extraction Kit (GeneAid Biotech), both according to manufacturers' guidelines. All the gene inserts in the constructed vectors were fully sequenced prior to expression tests to avoid any errors in the cloned genes.

Protein expression. Initial tests employed to screen the Fab_{H3} variants with multiple linker lengths and compare the soluble expression yields were performed in 24-deep well plates (DWPs). The polycistronic plasmid containing the genes of interest and the polycistronic plasmid containing the CyDisCo components Erv1p and PDI (pMJS205)²⁶ were used to cotransform chemically competent *E. coli* BL21 (DE3) and incubated overnight at 37 °C on Lysogeny Broth (LB) agar plates containing the appropriate antibiotics (35 µg/mL chloramphenicol for pLysS derivatives and 100 µg/mL ampicillin for pET23 derivatives). The transformed colonies obtained after overnight growth were used to inoculate 2 mL of LB media supplemented with 4 g/L glucose and suitable antibiotics per well in a 24-DWP covered with an oxygen permeable AirOTop (Thomson) membranes. These starter cultures were allowed to grow at 30 °C, 250 rpm (2.5 cm radius of gyration) for about 6–8 h. Expression cultures containing 3 mL autoclaved terrific broth autoinduction media (Formedium), supplemented with 0.8% glycerol and suitable antibiotics per well were seeded with the starter cultures in a 1:100 ratio. The expression cultures were grown at 30 °C, 250 rpm in 24 deep well plates covered with oxygen permeable AirOTop (Thomson) membranes to ensure efficient oxygenation for approximately 23–24 h. Soluble fraction of the cell lysate was used to perform the further purification and SDS-PAGE analysis steps to determine and compare soluble protein expression yields.

For yield quantification and comparison, pAAT50 or pAAT203 along with pMJS205 were used to cotransform chemically competent *E. coli* BL21(DE3) or MG1655 cells and grown on LB agar plates with appropriate antibiotics. Expression cultures containing 20 mL autoclaved terrific broth autoinduction media (Formedium) supplemented with 0.8% glycerol, or chemically defined minimal autoinduction media prepared in accordance with³⁷ were seeded with starter cultures in a 1:100 ratio. The expression cultures were covered with oxygen permeable AirOTop (Thomson) membranes and grown in 250 mL shake flasks at 30 °C, 250 rpm for 23–24 h for rich media and ~40 h for chemically defined media. Final optical density values of the cultures were measured at 600 nm and were found to be in the range of 24.0–26.5 and 18.5–20.0 for *E. coli* BL21 (DE3) and MG1655 respectively in rich media, and 15.0–17.0 for *E. coli* BL21 (DE3) in chemically defined media. The protein yields were calculated from IMAC-purified fractions of the soluble lysates in triplicate via densitometry analysis using purified REGN10987-Fab or Fab_{H3} as the concentration standards. Densitometric analysis to determine target protein yields was carried out using the ImageJ software.

On completion of the expression screening at the small scale, *E. coli* BL21 (DE3) cells transformed with the vectors for REGN10987 wild type Fab (pAAT50) or Fab_{H3} with a GS-G₄ linker (pAAT203) along with the vector for CyDisCo components (pMJS205) were grown in 1 L flasks containing 100 mL of the terrific broth autoinduction media (Formedium). These culture flasks were covered with oxygen permeable AirOTop (Thomson) membrane filters to ensure efficient oxygenation and incubated at 30 °C, 250 rpm for 23–24 h.

Protein purification. The small-scale cultures in 24 DWPs or shake flasks were harvested using centrifugation at 6500×g at 4 °C and the cell pellets were resuspended in 3 mL or 20 mL of lysis buffer respectively (50 mM sodium phosphate pH 7.4, 20 µg/mL DNase, 0.1 mg/mL egg white lysozyme), incubated for 15 min at room temperature and frozen at –20 °C. The cells were lysed by freeze thawing and as the proteins of interest contain a hexahistidine tag, they were purified with standard immobilized metal affinity chromatography (IMAC) using HisPur Cobalt Superflow Agarose (Thermo Scientific) resin under native conditions following clearance of the cell lysate by centrifugation (4000 rpm, 20 min, 4 °C). The protocol followed for Cobalt-IMAC based purification has been described in¹⁴.

For the large-scale cultures in 1 L flasks, the cells were harvested, and the soluble fraction of the lysate was processed for purification as described in⁴⁰. REGN10987 wild type Fab and Fab_{H3} (GS-G₄ linker) were purified using a combination of two chromatography steps: Nickel-based Immobilized Metal Affinity Chromatography (IMAC) and Anion Exchange Chromatography (AnEx). The detailed protocol for both these chromatography steps has been described in⁴⁰. The purified protein obtained after the AnEx step was buffer exchanged into 20 mM phosphate, 150 mM NaCl, pH 6.5 for REGN10987 Fab and pH 6.0 for the Fab_{H3}. Aliquots of the purified protein fractions were flash-frozen using liquid nitrogen and stored at –20 °C until further analysis. The method

of production and purification of mutated-truncated DsbC (mtDsbC) tagged SARS-CoV-2 Receptor Binding Domain (RBD) has been described in detail in⁴⁰.

Circular dichroism. Far-ultraviolet circular dichroism (CD) spectra of the purified proteins were measured using a Chirascan-Plus spectrophotometer. The scans were performed in duplicates as an average of 3 scans at 22 °C using a cuvette with a path length of 0.1 cm, spectral bandwidth and step size of 1 nm, and a scan speed of 1 nm/s. A wavelength range of 195–250 nm was used for the measurement and a high-tension (HT) voltage value of below 800 V was ensured for the scans. The final concentration of the purified protein sample (in 20 mM phosphate, 150 mM NaCl, pH 6.0/6.5) used for the analysis was 0.1 mg/mL diluted in ultrapure water. The final spectrum was obtained as an average of the scans and the blank subtracted.

Nano differential scanning fluorimetry (NanoDSF). Thermal stability of the proteins was assessed by NanoDSF using the Prometheus NT.48 system (NanoTemper Technologies, Germany) in triplicates. The purified proteins of interest were buffer exchanged into 20 mM sodium phosphate, pH 7.4 using a microcon-10 kDa centrifugal filter (Merck, USA) and loaded into standard capillaries (10 µL) of NanoDSF grade. These loaded samples were then subjected to programmed thermal ramping (1 °C/minute) from 20 °C to 90 °C. A dual UV detector was used to measure the fluorescence signal resulting from the thermal denaturation of the protein at 330 nm and 350 nm. The first derivative of the ratio between the two signals was used to determine the inflection point/s and the melting temperature (T_m) was calculated using the PR.ThermControl Software (NanoTemper Technologies).

Size exclusion chromatography-multi-angle light scattering (SEC-MALS). SEC-MALS tests were performed on a Shimadzu HPLC system with an in-line UV detector, an RI detector (Optilab, Wyatt Technology) and a MALS detector (miniDAWN, Wyatt Technology). The separation of purified proteins (1.5 mg/mL) was carried out on a Superdex™ 200 Increase 10/300 GL (Cytiva) column at a flow rate of 0.5 mL/min using 20 mM phosphate, 150 mM NaCl, pH 6.0, filtered with 0.1 µm membrane as the mobile phase. Data processing and analysis was carried out using ASTRA software v. 7.3.2.19 (Wyatt Technology).

Mass spectrometry. Molecular weight analysis of the purified proteins was performed by Electrospray ionization mass spectrometry in combination with liquid chromatography (LC-ESI-MS) using a Q-Exactive Plus Mass Spectrometer (Waters, MA, USA). 0.5 mg/mL of the purified protein of interest was first subjected to denaturation using 5 M guanidine hydrochloride. Post-denaturation, NEM-trapped samples were treated with 10 mM NEM, incubated at room temperature for 10 min and the alkylation reaction was then quenched with 0.5% trifluoroacetic acid (TFA) prior to analysis. For the non-NEM trapped samples, 0.5% TFA was directly added to the denatured protein samples prior to analysis. The experimental molecular weight (M_{exp}) of the protein was obtained by mass spectrometry analysis and the theoretical molecular weight (M_{theor}) was calculated using ExPasy ProtParam tool⁵⁸ based on the amino acid sequence of the proteins.

Biolayer interferometry. Prior to in vitro interaction analysis, the purified antigen mtDsbC-SARS-CoV-2 wild type was subjected to biotinylation using the protocol described in⁴⁰. The in vitro interaction of the biotinylated antigen with wild type REGN10987 Fab and Fab_{H3} was analysed using an Octet RED384 instrument (ForteBio, USA). The assays were performed with continuous agitation at 1000 rpm, 30 °C. All the measurements were performed using 1X Phosphate Buffer Saline (PBS) Kinetics Buffer (ForteBio) in 96-well plates. After obtaining an initial baseline with the running buffer, 5 µg/mL of the biotinylated antigen was immobilized on Streptavidin (SA) Dip and Read™ Biosensors (ForteBio) for 600 s. Five different concentrations each of the purified REGN10987 Fab and Fab_{H3} were tested for their binding to the antigen. Data analysis was performed on Octet Data Analysis High Throughput (HT) software 11.0.

Data availability

The data presented in this study is contained within the article and supplementary material. The datasets are available from the corresponding author on reasonable request.

Received: 23 May 2023; Accepted: 28 August 2023

Published online: 30 August 2023

References

- Rettenbacher, L. A. *et al.* Microbial protein cell factories fight back?. *Trends Biotechnol.* **40**, 576–590 (2022).
- Alibakhshi, A. *et al.* Targeted cancer therapy through antibody fragments-decorated nanomedicines. *J. Control. Release* **268**, 323–334 (2017).
- Chen, H. *et al.* Strategies and applications of antigen-binding fragment (Fab) production in *Escherichia coli*. *Pharm. Front.* **03**, e39–e49 (2021).
- Kang, T. H. & Seong, B. L. Solubility, stability, and avidity of recombinant antibody fragments expressed in microorganisms. *Front. Microbiol.* **11**, 1–10 (2020).
- Nelson, A. L. Antibody fragments: Hope and hype. *MAbs* **2**, 77–83 (2010).
- Walsh, G. & Walsh, E. Biopharmaceutical benchmarks 2022. *Nat. Biotechnol.* **40**, 1722–1760 (2022).
- Sandomenico, A., Sivaccumar, J. P. & Ruvo, M. Evolution of *Escherichia coli* expression system in producing antibody recombinant fragments. *Int. J. Mol. Sci.* **21**, 1–39 (2020).
- Gani, K., Bhambure, R., Deulgaonkar, P., Mehta, D. & Kamble, M. Understanding unfolding and refolding of the antibody fragment (Fab). I. In-vitro study. *Biochem. Eng. J.* **164**, 107764 (2020).

9. Lee, Y. K., Brewer, J. W., Hellman, R. & Hendershot, L. M. BiP and immunoglobulin light chain cooperate to control the folding of heavy chain and ensure the fidelity of immunoglobulin assembly. *Mol. Biol. Cell* **10**, 2209–2219 (1999).
10. Feige, M. J. *et al.* An unfolded CH1 domain controls the assembly and secretion of IgG antibodies. *Mol. Cell* **34**, 569–579 (2009).
11. Humphreys, D. P. *et al.* A plasmid system for optimization of Fab¹ production in *Escherichia coli*: Importance of balance of heavy chain and light chain synthesis. *Protein Expr. Purif.* **26**, 309–320 (2002).
12. Mehta, D., Chirmade, T., Tungekar, A. A., Gani, K. & Bhambure, R. Cloning and expression of antibody fragment (Fab) I: Effect of expression construct and induction strategies on light and heavy chain gene expression. *Biochem. Eng. J.* **176**, 108189 (2021).
13. Ojima-Kato, T. *et al.* ‘Zipbody’ leucine zipper-fused Fab in *E. coli* in vitro and in vivo expression systems. *Protein Eng. Des. Sel.* **29**, 149–157 (2016).
14. Tungekar, A. A. & Ruddock, L. W. Efficient production of Fc fusion proteins in the cytoplasm of *Escherichia coli*: Dissecting and mitigating redox heterogeneity. *Int. J. Mol. Sci.* **23**, 14740 (2022).
15. Thies, M. J. W. *et al.* Folding and oxidation of the antibody domain CH3. *J. Mol. Biol.* **319**, 1267–1277 (2002).
16. Gunasekaran, K. *et al.* Enhancing antibody Fc heterodimer formation through electrostatic steering effects: Applications to bispecific molecules and monovalent IgG. *J. Biol. Chem.* **285**, 19637–19646 (2010).
17. Ridgway, J. B. B., Presta, L. G. & Carter, P. ‘Knobs-into-holes’ engineering of antibody C(H)3 domains for heavy chain heterodimerization. *Protein Eng.* **9**, 617–621 (1996).
18. Moore, G. L. *et al.* A novel bispecific antibody format enables simultaneous bivalent and monovalent co-engagement of distinct target antigens. *MAbs* **3**, 546–557 (2011).
19. Davis, J. H. *et al.* SEEDbodies: Fusion proteins based on strand-exchange engineered domain (SEED) CH3 heterodimers in an Fc analogue platform for asymmetric binders or immunofusions and bispecific antibodies. *Protein Eng. Des. Sel.* **23**, 195–202 (2010).
20. Choi, H. J., Kim, Y. J., Lee, S. & Kim, Y. S. A heterodimeric Fc-based bispecific antibody simultaneously targeting vegfr-2 and met exhibits potent antitumor activity. *Mol. Cancer Ther.* **12**, 2748–2759 (2013).
21. Wozniak-Knopp, G. *et al.* An antibody with Fab-constant domains exchanged for a pair of CH3 domains. *PLoS ONE* **13**, 1–19 (2018).
22. Merchant, A. M. *et al.* An efficient route to human bispecific IgG. *Nat. Biotechnol.* **16**, 677–681 (1998).
23. Müller, A. *et al.* Nonnative disulfide bond formation activates the σ 32-dependent heat shock response in *Escherichia coli*. *J. Bacteriol.* **195**, 2807–2816 (2013).
24. Dall’Acqua, W., Simon, A. L., Mulkerrin, M. G. & Carter, P. Contribution of domain interface residues to the stability of antibody C(H)3 domain homodimers. *Biochemistry* **37**, 9266–9273 (1998).
25. Matthews, B. W. Studies on protein stability with T4 lysozyme. *Adv. Protein Chem.* **46**, 249–278 (1995).
26. Gaciarz, A. *et al.* Systematic screening of soluble expression of antibody fragments in the cytoplasm of *E. coli*. *Microb. Cell Fact.* **15**, 22 (2016).
27. Hatahet, F., Nguyen, V. D., Salo, K. E. H. & Ruddock, L. W. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*. *Microb. Cell Fact.* **9**, 67 (2010).
28. 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*. *Microb. Cell Fact.* **10**, 1 (2011).
29. Bhatwa, A. *et al.* Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. *Front. Bioeng. Biotechnol.* **9**, 1–18 (2021).
30. Chiu, M. L., Goulet, D. R., Teplyakov, A. & Gilliland, G. L. Antibody structure and function: The basis for engineering therapeutics. *Antibodies* **8**, 55 (2019).
31. Tetin, S. Y., Prendergast, F. G. & Venyaminov, S. Y. Accuracy of protein secondary structure determination from circular dichroism spectra based on immunoglobulin examples. *Anal. Biochem.* **321**, 183–187 (2003).
32. Tetin, S. Y. & Linthicum, D. S. Circular dichroism spectroscopy of monoclonal antibodies that bind a superpotent guanidinium sweetener ligand. *Biochemistry* **35**, 1258–1264 (1996).
33. Tetin, S. Y., Denzin, L. K., Weidner, K. M., Voss, E. W. & Mantulin, W. W. Comparative circular dichroism studies of an anti-fluorescein monoclonal antibody (Mab 4-4-20) and its derivatives. *Biochemistry* **31**, 12029–12034 (1992).
34. Akazawa-Ogawa, Y., Nagai, H. & Hagihara, Y. Heat denaturation of the antibody, a multi-domain protein. *Biophys. Rev.* **10**, 255–258 (2018).
35. Atwell, S., Ridgway, J. B. B., Wells, J. A. & Carter, P. Stable heterodimers from remodeling the domain interface of a homodimer using a phage display library. *J. Mol. Biol.* **270**, 26–35 (1997).
36. Toughiri, R. *et al.* Comparing domain interactions within antibody Fabs with kappa and lambda light chains. *MAbs* **8**, 1276–1285 (2016).
37. Lilie, H. Folding of the Fab fragment within the intact antibody. *FEBS Lett.* **417**, 239–242 (1997).
38. Gunasekaran, K., Wittekind, M., Yan, W. & Pentony, M. Method for making antibody Fc-heterodimeric molecules using electrostatic steering effects. *WO/2009/089004* (Amgen Inc., 2009).
39. Gaciarz, A. & Ruddock, L. W. Complementarity determining regions and frameworks contribute to the disulfide bond independent folding of intrinsically stable scFv. *PLoS ONE* **12**, 1–20 (2017).
40. Tungekar, A. A. & Ruddock, L. W. Production of neutralizing antibody fragment variants in the cytoplasm of *E. coli* for rapid screening: SARS-CoV-2 a case study. *Sci. Rep.* **13**, 4408 (2023).
41. Hansen, J. *et al.* Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **369**, 1010–1014 (2020).
42. Rathore, A. S. & Batra, J. Antibody production in microbial hosts. *BioPharm Int.* **29**, 18–23 (2016).
43. Brinkmann, U. & Kontermann, R. E. The making of bispecific antibodies. *MAbs* **9**, 182–212 (2017).
44. Ha, J. H., Kim, J. E. & Kim, Y. S. Immunoglobulin Fc heterodimer platform technology: From design to applications in therapeutic antibodies and proteins. *Front. Immunol.* **7**, 1–16 (2016).
45. Kontermann, R. E. Dual targeting strategies with bispecific antibodies © 2012 Landes Bioscience. *MAbs* **4**, 182–197 (2012).
46. Rader, C. Overview on concepts and applications of Fab antibody fragments. *Curr. Protoc. Protein Sci.* **55**, 1–14 (2009).
47. Eleniste, P. P., Hofstetter, H. & Hofstetter, O. Expression and characterization of an enantioselective antigen-binding fragment directed against α -amino acids. *Protein Expr. Purif.* **91**, 20–29 (2013).
48. Demarest, S. J. *et al.* Engineering stability into *Escherichia coli* secreted Fabs leads to increased functional expression. *Protein Eng. Des. Sel.* **19**, 325–336 (2006).
49. Alt, M., Müller, R. & Kontermann, R. E. Novel tetravalent and bispecific IgG-like antibody molecules combining single-chain diabodies with the immunoglobulin γ 1 Fc or CH3 region. *FEBS Lett.* **454**, 90–94 (1999).
50. Lu, D. *et al.* Di-diabody: A novel tetravalent bispecific antibody molecule by design. *J. Immunol. Methods* **279**, 219–232 (2003).
51. Hu, S. Z. *et al.* Minibody: A novel engineered anti-carcinoembryonic antigen antibody fragment (single-chain Fv-CH3) which exhibits rapid, high-level targeting of xenografts. *Cancer Res.* **56**, 3055–3061 (1996).
52. Von Kreudenstein, T. S. *et al.* Improving biophysical properties of a bispecific antibody scaffold to aid developability: Quality by molecular design. *MAbs* **5**, 646–654 (2013).
53. Kang, H. J., Kim, H. J., Jung, M. S., Han, J. K. & Cha, S. H. Optimal expression of a Fab-effector fusion protein in *Escherichia coli* by removing the cysteine residues responsible for an interchain disulfide bond of a Fab molecule. *Immunol. Lett.* **184**, 34–42 (2017).

54. Hölzer, W., Petersen, F., Strittmatter, W., Matzku, S. & Von Hoegen, I. A fusion protein of IL-8 and a Fab antibody fragment binds to IL-8 receptors and induces neutrophil activation. *Cytokine* **8**, 214–221 (1996).
55. Robinson, M. P. *et al.* Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nat. Commun.* **6**, 9072 (2015).
56. Jung, S. T. *et al.* Aglycosylated IgG variants expressed in bacteria that selectively bind FcγRI potentiate tumor cell killing by monocyte-dendritic cells. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 604–609 (2010).
57. Li, Z., Kessler, W., Van Den Heuvel, J. & Rinas, U. Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems. *Appl. Microbiol. Biotechnol.* **91**, 1203–1213 (2011).
58. Gasteiger, E. *et al.* ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).

Acknowledgements

This work was funded by the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 813979 (SECRETTERS). The use of the facilities of the Biocenter Oulu core facilities, a member of Biocenter Finland, is gratefully acknowledged. Special thanks to Dr Ulrich Bergmann and Dr Hongmin Tu for their help with mass spectrometry and biolayer interferometry analysis respectively.

Author contributions

Experimental investigation and analysis A.A.T., Supervision: L.W.R., Conceptualization L.W.R., Manuscript writing and editing: A.A.T., Manuscript review and editing: L.W.R., Funding: L.W.R. All authors have read and approved the manuscript.

Competing interests

A patent for the CyDisCo system is held by the University of Oulu: Method for producing natively folded proteins in a prokaryotic host (Patent number 9238817; date of patent 19 January 2016). Inventor: Lloyd W. Ruddock. The other author declares no competing interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-41525-3>.

Correspondence and requests for materials should be addressed to L.W.R.

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) 2023