frontiers | Research Topics

Phage display: Technique and applications

Edited by
Jian Huang, Yoichi Takakusagi and Beibei Ru

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-83251-103-9
DOI 10.3389/978-2-83251-103-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Phage display: Technique and applications

Topic editors

Jian Huang — University of Electronic Science and Technology of China, China
Yoichi Takakusagi — Institute of Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology, Japan
Beibei Ru — National Institutes of Health (NIH), United States

Citation

Huang, J., Takakusagi, Y., Ru, B., eds. (2023). *Phage display: Technique and applications*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-103-9

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 06 December 2022
DOI 10.3389/fmicb.2022.1097661

OPEN ACCESS

EDITED AND REVIEWED BY
Sangryeol Ryu,
Seoul National University, South Korea

*CORRESPONDENCE
Jian Huang
hj@uestc.edu.cn

SPECIALTY SECTION
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

RECEIVED 14 November 2022
ACCEPTED 21 November 2022
PUBLISHED 06 December 2022

CITATION
Huang J, Takakusagi Y and Ru B (2022) Editorial: Phage display: Technique and applications.
Front. Microbiol. 13:1097661.
doi: 10.3389/fmicb.2022.1097661

Editorial: Phage display: Technique and applications

Jian Huang[1]*, Yoichi Takakusagi[2] and Beibei Ru[3]

[1]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, [2]Institute for Quantum Life Science, National Institutes for Quantum Science and Technology, Chiba, Japan, [3]Cancer Data Science Lab, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States

KEYWORDS

phage display, next generation sequencing, artificial intelligence (AI), peptide, antibody, phage library

Editorial on the Research Topic
Phage display: Technique and applications

Phages are the most abundant organisms on the earth and have a great impact on the earth's ecosystems. Some phages have also been engineered into good expression vectors for coupling genotypes and phenotypes, which can be utilized to harness the power of evolution by artificial selection. Phage display is a powerful and versatile technology for constructing a library of peptides or antibodies displayed on phage virions, and then screening them for peptides or antibodies with desired properties. In 2018, one-half of the Nobel Prize in chemistry was awarded jointly to George P. Smith and Sir Gregory P. Winter for the "phage display of peptides and antibodies" (Smith, 2019; Winter, 2019). Phage display has been widely used to map epitopes, identify drug targets (Takakusagi et al., 2020), and develop therapeutics, diagnostics, and vaccines.

In the present Research Topic, we have organized 11 articles, including seven original research papers, two method papers, and two reviews, to share the authors' understanding of the phage display technique and its applications from various angles.

Next-generation sequencing (NGS) has been employed in most, if not all, fields of biology and medicine. Phage display is no exception, either for antibodies or peptides. Using the Roche 454 NGS platform, Lisowska et al. identified over 13,000 peptides targeting ubiquitin and over 10,000 peptides against ubiquitin-like modifier NEDD8 from the Ph.D.-12 Phage Display Peptide Library. Two peptides were further confirmed to inhibit both E3 ubiquitin ligases, MDM2 and CHIP, and their binding modes were solved by the NMR analysis. Zhao et al. made a phage panning scheme that separately covered selections against phosphotyrosine (pTyr) and sulfated tyrosine (sTyr) peptides, followed by NGS. After mining the NGS data and validating with experiments, they successfully identified some sTyr superbinders. Maruthachalam et al. designed three synthetic Fab libraries based on a modified trastuzumab framework. Using the Ion Torrent NGS platform, they found 12 Fabs against Notch-1, and 2 Fabs showed strict specificity for Notch-1 with very high affinities.

Artificial intelligence (AI) is also increasingly applied in phage display studies. He et al. used NGS with phage display and a lot of peptides were obtained from the Ph.D.-12 Phage Display Peptide Library panning against PD-L1. Based on the NGS data obtained, they further adopted different sequence features and various machine-learning methods to train models for predicting PD-L1 binders. Finally, PDL1Binder, an ensemble computational model, was implemented as a web server. Liu et al. studied the sequence features of phage proteins and non-phage proteins. They proposed a feature selection method based on ensemble learning. Their findings might help to find out new phages from metagenomes, providing candidate genes or vectors for new phage library construction.

In addition to NGS and AI, relevant scholars have been trying to improve the phage display technique and explore its application from different aspects. In the review by André et al., the state-of-the-art *in vivo* phage display methodologies were summarized and discussed, especially the promising emerging selection strategy for improving antibody targeting and drug delivery properties. Allen et al. reviewed how bioconjugation and the incorporation of non-canonical amino acids had expanded the chemical diversity of peptides and proteins displayed by M13 phage virions for a variety of purposes. Ma et al. developed and compared three different methods to increase the signal-to-background ratio of ELISA assay during biopanning, which might help probe the weak protein–protein interactions using phage display. Thavorasak et al. applied phage display to reveal the enhancing epitope on the spike protein of the porcine epidemic diarrhea virus (PEDV), which would be helpful and useful in designing a safe and effective PEDV vaccine devoid of the enhancing epitope. Glab-ampai et al. tested the effect of a phage display-derived superantibody originally targeting a conformational epitope on RNA-dependent RNA polymerase of the hepatitis C virus. Their results showed that the superantibody could also inhibit the replication of many RNA viruses, such as DENV, ZIKV, JEV, EV71, CVA16, PEDV, and SARS-CoV-2 (Wuhan wild-type and the variants of concern) in a dose-dependent manner. Chen et al. developed a method called Pi-mqPCR, which was short for phage display mediated immuno-multiplex quantitative PCR. They applied it to monitor and distinguish the differences in the immune response to antigenic domains of multiple SARS-CoV-2 variants simultaneously.

In conclusion, the phage display technique is very flexible and evolves continuously. It has been applied more and more frequently with NGS and AI, which makes phage display more productive and efficient. In our opinion, phage display is still one of the most useful experimental techniques, which greatly helps biomarker discovery, vaccine design, and peptide and antibody drug development (Ning et al., 2021).

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was supported by a grant from the National Natural Science Foundation of China (62071099).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ning, L., Abagna, H. B., Jiang, Q., Liu, S., and Huang, J. (2021). Development and application of therapeutic antibodies against COVID-19. *Int. J. Biol. Sci.* 17, 1486–1496. doi: 10.7150/ijbs.59149

Smith, G. P. (2019). Phage display: simple evolution in a petri dish. *Angew. Chem. Int. Ed.* 58, 14428–14437. doi: 10.1002/anie.201908308

Takakusagi, Y., Takakusagi, K., Sakaguchi, K., and Sugawara, F. (2020). Phage display technology for target determination of small-molecule therapeutics: an update. *Expert Opin. Drug Discov.* 15, 1199–1211. doi: 10.1080/17460441.2020.1790523

Winter, G. (2019). Harnessing evolution to make medicines. *Angew. Chem. Int. Ed.* 58, 14438–14445. doi: 10.1002/anie.201909343

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fmicb.2022.899578

Development in Detection Methods for the Expression of Surface-Displayed Proteins

Chenglong Ma[1,2†], Chunyang Jiang[2†], Dongping Zhao[2], Shuhao Li[2], Ronggui Li[1] and Lei Li[2*]*

[1]College of Life Sciences, Qingdao University, Qingdao, China, [2]School of Basic Medicine, Qingdao Medical College, Qingdao University, Qingdao, China

Directed evolution is a widely-used engineering strategy for improving the stabilities or biochemical functions of proteins by repeated rounds of mutation and selection. A protein of interest is selected as the template and expressed on a molecular display platform such as a bacteriophage for engineering. Initially, the surface-displayed protein template needs to be checked against the desired target *via* ELISA to examine whether the functions of the displayed template remain intact. The ELISA signal is subject to the protein-target binding affinity. A low-affinity results in a weak ELISA signal which makes it difficult to determine whether the weak signal is because of low affinity or because of poor expression of the protein. Using a methyllysine-binding chromodomain protein Cbx1 that weakly binds to the histone H3K9me3 peptide, we developed and compared three different approaches to increase the signal-to-background ratio of ELISA measurements. We observed that the specific peptide-binding signal was enhanced by increasing the Cbx1 phage concentration on the ELISA plate. The introduction of previously known gain-of-function mutations to the Cbx1 protein significantly increased the ELISA signals. Moreover, we demonstrated that the H3K9me3-specific binding signal was enhanced by fusing Cbx1 with a high-affinity phosphotyrosine-binding protein and by coating the ELISA plate with a mixture of H3K9me3 and phosphotyrosine peptides. This approach also worked with binding to a lower affinity momomethyllysine peptide H3K9me1. These approaches may help improve ELISA experiments when dealing with low-affinity ligand-protein interactions.

Keywords: ELISA, molecular display platform, phage display, SH2, chromodomain

OPEN ACCESS

Edited by:
Jian Huang,
University of Electronic Science and Technology of China, China

Reviewed by:
Jianghai Liu,
Chengdu Medical College, China
Feifei Song,
University of North Carolina at Chapel Hill, United States

***Correspondence:**
Ronggui Li
lrg@qdu.edu.cn
Lei Li
leili@qdu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

Received: *18 March 2022*
Accepted: *07 April 2022*
Published: *25 April 2022*

Citation:
Ma C, Jiang C, Zhao D, Li S, Li R and Li L (2022) Development in Detection Methods for the Expression of Surface-Displayed Proteins. Front. Microbiol. 13:899578. doi: 10.3389/fmicb.2022.899578

INTRODUCTION

Protein engineering through directed evolution is an effective way to obtain proteins with enhanced properties [e.g., increased binding affinity (Nowak et al., 2020) and catalytic activity (Bilal et al., 2018)], which has significantly impacted biological research and biotechnology. Generally, the protein of interest is selected as the template and is engineered by applying random or site-directed mutagenesis techniques to generate a library of protein variants. This library is then screened to identify the mutations that confer the desired property. Several

molecular display platforms have been developed, on which the engineered variants are tethered to the surface of phage (Smith, 1985), bacteria (Francisco et al., 1993), or yeast (Boder and Wittrup, 1997) cells. In the cell surface display, each cell is transformed with a single vector encoding a protein variant fused to a cell-surface anchor protein and therefore, the variant is accessible to the extracellular space. The surface-displayed proteins include antibody derivatives [e.g., scFv (Tadayoni et al., 2021) and Fab (Smith and Kaiser, 2016)] and some proteins that bind to specific targets. For example, ADP-ribose binding macro domain protein Af1521 was engineered with a 1,000-fold increased affinity towards ADP-ribose (Nowak et al., 2020), the Fyn SH2 domain was engineered with over 100-fold increased affinity towards phosphotyrosine (pY; Kaneko et al., 2012; Li et al., 2021), and H3K9me3-binding protein Cbx1 was engineered to specifically recognize H3K9me3 with enhanced affinity (Hard et al., 2018; Albanese et al., 2020).

Protein engineering through directed evolution starts with the premise that the protein template is functional and well displayed on the surface. That is necessary; otherwise, the variants in the library constructed based on the template may be poorly expressed, and those variants with the desired property may not be successfully selected from the library. For instance, before the Fyn SH2 domain was considered a template to find variants with potentially higher affinity to phosphotyrosine, it should be ensured that this SH2 domain is displayed on the platform surface with the original phosphotyrosine-binding property. When conducting directed evolution experiments by the phage display technology, the expression and function of the protein presented on the phage surface need to be routinely checked against the desired target *via* phage ELISA, in which phage-containing supernatants were commonly used. However, the phage ELISA signal is subject to a few factors, such as buffer composition and the protein-target binding affinity. A low binding affinity may result in a weak ELISA signal, making it difficult to determine whether the weak ELISA signal is caused by poor protein expression on the phage or by low-affinity binding of the variant to the target. This study investigated phage-displayed proteins with low affinities to their targets and figured out how to check whether the proteins are well expressed on the phage surface. Here, we expressed as study cases the phage-displayed human chromodomain protein Cbx1, which recognizes Histone 3 lysine 9 (H3K9) mono-, di- and tri-methylation (Liu et al., 2010), as well as human Fyn SH2 domain, which recognizes tyrosine phosphorylation (Kaneko et al., 2012; Li et al., 2021).

MATERIALS AND METHODS

Bacterial Strains and Plasmids

The *E. coli* XL1-blue was purchased from New England Biolabs, and *E. coli* CJ236 was from Takara Biotechnology. The phagemid pFN-OM6 was obtained from Shanghai AsiaUnited Antibody Medical Co., Ltd. The Fyn SH2 phagemid contained N-terminal Flag-tagged Fyn SH2 domain (Kaneko et al., 2012; Li et al., 2021), and the Cbx1 phagemid included N-terminal Flag-tagged Cbx1 protein (UniProt ID: P83916, 1–185). The Fyn SH2-Cbx1 fusion phagemid was constructed by connecting N-terminal Flag-tagged Fyn SH2, a 3x GGGGS linker and the Cbx1 protein.

Abbreviations: ELISA, Enzyme-Linked ImmunoSorbent Assay; SH2, Src Homology 2.

Reagents

Tryptone, Yeast Extract were purchased from Thermo Fisher Scientific, OXOID. Agar powder and skim milk powder was purchased from the Solarbio company. Glycerol, Sodium chloride (NaCl), Sodium hydroxide (NaOH), and other reagents of analytical reagent grade were purchased from Sinopharm Chemical Reagent (Shanghai, China); T4 Polynucleotide Kinase, T4 DNA Polymerase and T4 DNA Ligase were purchased from New England Biolabs (NEB).

Media for Bacterial Growth

2 × Yeast Extract Tryptone (YT) Media

Five grams of NaCl, 10 g Yeast Extract and 16 g Tryptone were dissolved in 900 ml ddH_2O. The pH of 2YT mediun was adjusted to 7.5 with about 1 ml 4 M NaOH and the volume of medium was made up to 1 l with ddH_2O. Then the medium was sterilized in an autoclave and cooled to room temperature for use.

Luria-Bertani (LB) Medium Plate

Ten grams per litre of Tryptone, 5 g/l Yeast Extract, 10 g/l NaCl, 15 g/l agar powder, adjusted to a pH of 7.5 with 4 M NaOH were sterilized in an autoclave. The medium was allowed to cool to 50°C, after which carbenicillin (final concentration 50 μg/ml) was added. The medium was poured in to the plates in the super clean bench. Then the LB medium plates were cooled to room temperature for use.

Phage Preparation and Precipitation of Phage Particles

From a fresh LB/carb plate, a single colony of *E. coli* XL1-blue [New England Biolabs (NEB)] harbouring the phagemid pFN-OM6 was picked into 2 ml of 2YT media supplemented with Carb (carbenicillin, 50 μg/ml). The colony was incubated with shaking at 200 rpm at 37°C for 6 h, and M13KO7 helper phage was added with a final concentration of 1×10^{10} pfu/mL. After the incubation with shaking at 200 rpm at 37°C for 45 min, the culture was transferred to a 250 ml flask that contained 50 ml 2YT media supplemented with Carb (carbenicillin, 50 μg/ml) and Kana (kanamycin, 25 μg/ml). The flask was incubated on a shaking platform with a speed of 200 rpm at 37°C for 16 to 18 h. The culture was centrifuged for 10 min at 16,000 g at 4°C, and the amplified phages were solved in the supernatant.

The phage supernatant was transferred to a new tube containing 1/5 volume of PEG8000/NaCl and incubated for 30 min at an ice bath to concentrate bacteriophages. After 10-min centrifugation at 16,000 g at 4°C, the supernatant was discarded. The phage pellets were resuspended in 1 ml of 1×PBS

and transferred to a 1.5-mL microcentrifuge tube. After 10-min centrifugation at 16,000 g at 4°C, the supernatant was transferred to a new 1.5-mL microcentrifuge tube.

Phage ELISA

In a 96-well microplate (Costar, No. 42592), 6 pmol streptavidin (Solarbio, S9171) was coated per well in 50 µl 1×PBS at 4°C overnight. The 3% (m/v) skim milk was prepared by dissolving skim milk powder (Solarbio, No. D8340) in PBS and could be regard as blank control. Next, the solution in the well was discarded, and 200 µl/well 3% (m/v) skim milk was added for blocking at room temperature for 1 h. After removing the solution, the wells were washed three times by the PT buffer (PBS supplemented with 0.05% Tween 20, Solarbio, T8220). Then 50 µl biotinylated peptides (24 pmol/well) were added for immobilization for one hour at room temperature. For the group of blank control, 50 µl 3% (m/v) skim milk were added rather than biotinylated peptides. After washing by the PT buffer three times, the phage solution was added for binding for one hour at room temperature. Unbound phages were washed away six times by the PT buffer. 50 µl anti-M13/HRP antibodies (Sino Biological, No. 11973) were added and incubated for 30 min at room temperature. The wells were washed six times by the PT buffer and two times by the PBS buffer. 50 µl freshly prepared TMB substrates (Innoreagents, TMB-S-003) were added to develop, and 50 µl 1.0 M H_3PO_4 were added to stop the reaction. The signals were read spectrophotometrically at 450 nm in a plate reader.

Cbx1 Mutants Construction Based on Kunkel Method

The dut$^-$/ung$^-$ *E. coli* CJ236 (Takara Biotechnology, Dalian, China) was applied to produce the uracil-containing ssDNA (dU-ssDNA), as it lacks dUTPase and uracil-N glycosylase activities, and the dut$^+$/ung$^+$ *E. coli* XL1-blue (New England Biolabs) was applied to proliferate newly synthesized strands and digest the uracil-containing parental template. Primer 1 (TATCTTCTAAAGTGGGCAGGTTTCTCAGATGAG) and primer 2 (CCAGAAGAAAATCTGTTCTGCCCTGACCTTATT) were used in a combinatorial mutation to construct Cbx1 double mutant (K43A/D59F). Primer 3 (GAGGAAGAGGAA TATGAAGTGGAAG AAGTTCTTGATCGGCGA) and primer 4 (CCAGAAGAAAATCTGTCTTGCCCTGACCTTATT) were used to construct Cbx1 triple mutant (V22E/K25E/D59S).

Production of dU-ssDNA

The dU-ssDNA of the wild-type human Cbx1 as the template in the Kunkel reaction was made as described before (Tonikian et al., 2007). The phagemid Cbx1-pFN-OM6 was transformed into the CJ236 cells. The following day, a single clone was inoculated in 2 ml 2×YT medium containing carbenicillin (50 µg/ml) at 37°C, shaking at 200 rpm for 5 h. M13K07 helper phage (New England Biolabs, Ipswich, MA, United States) was then added into the medium with a final concentration of 1×10^{10} pfu/mL, and the cells were incubated for another 1 h. The whole culture was transferred into 100 ml 2×YT/Carb (50 µg/ml)/Kana (25 µg/ml)/Uridine (0.25 µg/ml) medium, incubated at 37°C and shaken at 200 rpm for 18 to 20 h. The culture was then centrifuged at 16,000 g for 10 min to remove the cells. Phage particles in the supernatant were precipitated with PEG8000/NaCl and were resuspended in 1 ml phosphate-buffered saline (PBS). The phage dU-ssDNA was extracted by the Spin M13 kit (Omega Bio-Tek, Norcross, GA, United States), quantified by Nanodrop1000 spectrophotometer, and analyzed by agarose gel electrophoresis.

In vitro CCC-dsDNA Synthesis

A classic three-step procedure of the Kunkel method was employed to synthesize heteroduplex covalently closed, circular, double-stranded DNA (CCC-dsDNA; Tonikian et al., 2007). Briefly, 20 pmol mutagenic oligonucleotides were phosphorylated in a 20 µl reaction with T4 polynucleotide kinase at 37°C for 1 h. Then, the product was annealed with 2 pmol dU-ssDNA by using cooling programs. Finally, the CCC-dsDNA was synthesized by a fill-in reaction with T4 DNA polymerase and T4 DNA ligase at 22°C overnight.

RESULTS

The Phage ELISA Signal for Cbx1 and H3K9me3 Interaction Is as Weak as the Background Signal

To investigate differences between low affinity and high-affinity interactions in phage-ELISA experiments, we chose two protein domains, the chromodomain protein Cbx1 and the tyrosine kinase Fyn SH2 domain, representing low- or high-affinity interactions. Human Cbx1 protein binds to the histone 3 lysine 9 trimethylated peptide (H3K9me3: KQTARKme3STGGKA) with relatively low affinity (Li et al., 2021), measured by fluorescence polarization, compared with the interaction between the Fyn SH2 domain and the MidT-pY^{324} peptide (EPQpYEEIPIYL; Kaneko et al., 2012). The dissociation constant value (K_d = 3.2 µM) of the former is ten times larger than that (K_d = 0.327 µM) for the latter (Li et al., 2021). Additionally, Cbx1 does not bind to the corresponding nonmethylated H3K9 peptide, and the Fyn SH2 domain fails to bind to the nonphosphorylated MidT-Y^{324} peptide (Kaneko et al., 2012; Li et al., 2021).

We fused flag-tagged Cbx1 and Fyn SH2 domain to the capsid protein P3 of the phagemid pFN-OM6, respectively, and checked their ligand binding by measuring their interactions with the corresponding biotinylated targets using supernatant phage ELISA. The streptavidin-coated wells were used to measure the protein and target interactions, taking the biotinylated unmodified targets as the negative control and milk as the background. The anti-FLAG antibody-coated wells were applied as the positive control to check the expression of the engineered P3. We considered an ELISA signal as positive when the signal intensity is averagely triple as strong as the background intensity, and they are statistically different from each other with a value of p <0.05. **Figure 1** showed that positive controls for both

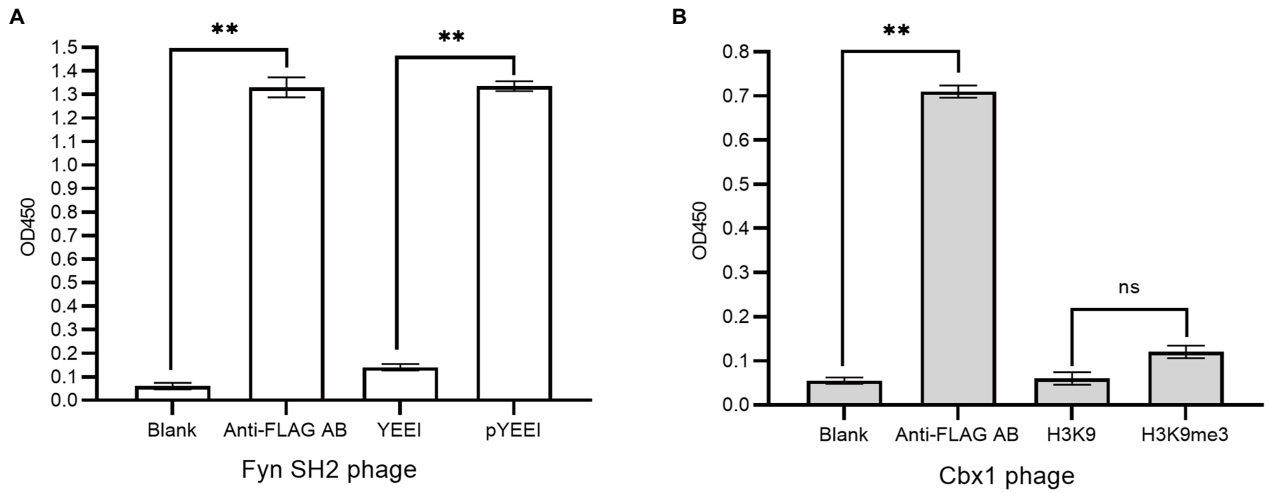


FIGURE 1 | Supernatant phage ELISA showed the expression level and binding to the pY (MidT-pY324) peptide of the Fyn SH2 domain **(A)** and the H3K9me3 peptide of Cbx1 **(B)**, respectively. ELISA signals (OD450) are plotted as a function of the supernatant phages added to the streptavidin-coated wells conjugated with biotinylated modified peptides (pY or H3K9me3), unmodified counterparts (Y or H3K9) as negative controls and skim milk as the blank control, respectively. The anti-FLAG antibody (AB)-coated wells were taken as positive controls. The experiments were repeated three times and a statistical significance in comparison was calculated using unpaired two-sample t-test: ns: not significant; *$p < 0.05$; **$p < 0.01$. The same statistical analysis was performed in the following figures.

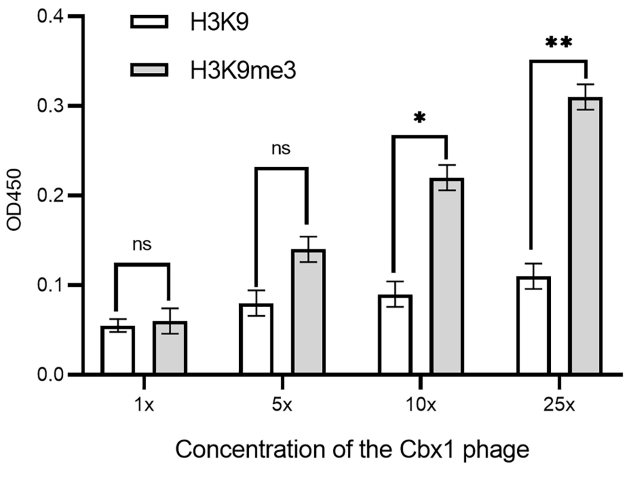


FIGURE 2 | ELISA signals were plotted as a function of the Cbx1-containing phage with different phage concentrations added to the streptavidin-coated wells conjugated with biotinylated H3K9me3 or H3K9 peptides as the negative control. 1x concentration represents a solution with a phage concentration similar to the supernatant phage. The experiments were repeated three times.

phages had stronger signals (1.37 or 0.71) than the background signals (0.05), indicating that the engineered P3 was well expressed. Fyn SH2 domain had a strong ELISA signal (1.31) towards the pY peptide but a weak signal (0.14) towards the negative control, indicating that Fyn SH2 was functionally displayed (**Figure 1A**). By contrast, the ELISA signal (0.11) of Cbx1 and the H3K9me3 peptide is close to the negative control (0.06; value of $p = 0.057$), showing the uncertainty about whether Cbx1 was well or poorly expressed. This low ELISA signal of the Cbx1-H3K9me3 complex might be due to either their low affinity or the incorrect design of the CBX1-P3 fusion gene. We figured it out by performing several different approaches, each with its advantages and disadvantages.

Approach One: High Phage Concentration Increases the ELISA Signal

As there is a relatively low concentration of phage particles in the supernatants, it is easy to elute the bound phages if the phage-displayed protein has a low affinity to the target. These may result in a low ELISA signal for the Cbx1 and H3K9me3 binding. One possible solution is the increase in the number of bound phages. To this end, we concentrated the phage particles in the supernatants and repeated the phage ELISA. **Figure 2** shows that the ELISA signal for the Cbx1 and H3K9me3 complex is proportional to the phage concentrations. The ELISA signal (0.31) related to the highest phage concentration is five times as strong as that (0.06) of the supernatant phage and triple that (0.11) of its binding to the H3K9 peptide (as negative control). These observations indicate that Cbx1 is functionally displayed on the phage surface. Moreover, they suggest that phage concentration increases the phage amount and more bound phages remain after elution, increasing the ELISA signal. Nevertheless, the ELISA signal for the negative control increased from 0.05 to 0.11 with the concentrations, possibly caused by the increase of nonspecific binding. This approach may be unsuitable for testing the protein with much low binding affinity to its target.

Approach Two: Gain-of-Function Protein Mutations Boost the ELISA Signal

If a protein-target interaction has a low affinity which results in a low phage ELISA signal, the affinity can be increased based on the published variants with a higher affinity. For

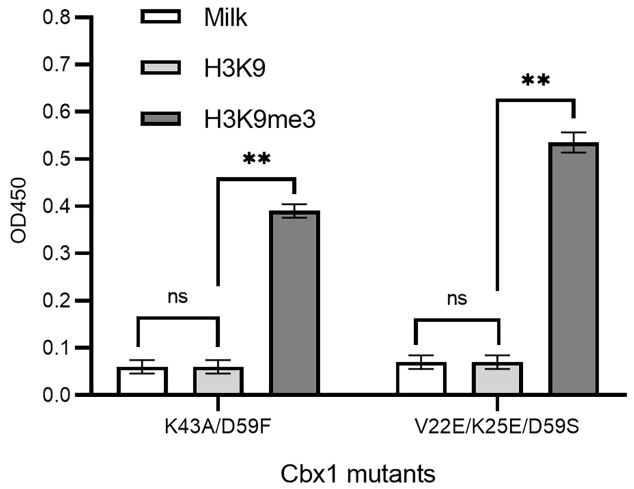


FIGURE 3 | Supernatant phage ELISA showed the expression level and binding to the H3K9me3 peptide of the Cbx1 mutants. The milk and the H3K9 peptide were considered background and the negative control, respectively. The experiments were repeated three times. The experiments were repeated three times.

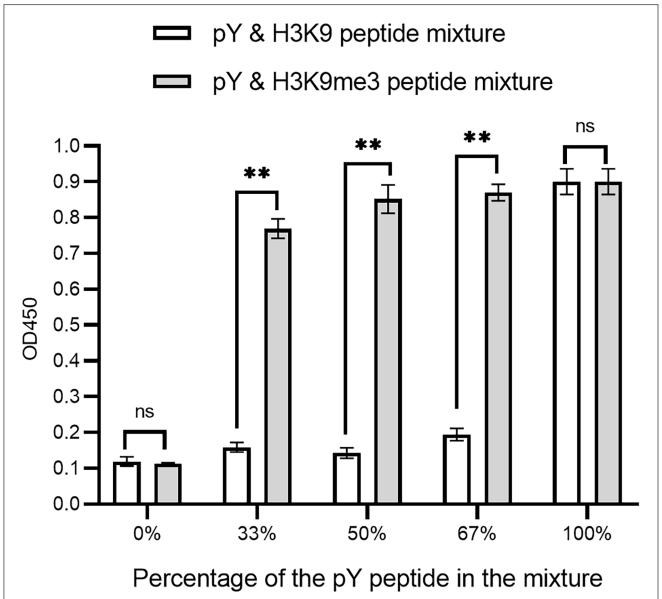


FIGURE 4 | Supernatant ELISA signals were plotted as a function of the Cbx1-and-FynSH2-containing phage binding to the mixtures of pY and H3K9me3 peptides with different ratios. The proportion of pY peptide in the mixtures ranged from 0 to 100%. The mixtures of pY and H3K9 peptides were taken as the negative control. The experiments were repeated three times.

instance, it was documented that the Cbx1 double mutant (K43A/D59F; Albanese et al., 2020) had a four-fold higher affinity, and the triple mutant (V22E/K25E/D59S; Hard et al., 2018) had ten-fold high affinity compared to the wild type Cbx1. These two mutants may have higher ELISA signals than the wild type. To test it, we mutated the fused Cbx1 to both variants using the Kunkel method, respectively, and measured their binding to H3K9me3 and H3K9 peptides by supernatant phage ELISA. **Figure 3** showed that both mutants had around seven-fold stronger ELISA signals (0.38 or 0.53) for binding to the H3K9me3 peptide than those (~0.06) for binding to the H3K9 peptide. It indicates that the fused Cbx1 is well expressed. The advantage of this method is that even though the wild-type protein has a low affinity, it can be mutated as the variants with high affinity, which can be employed to check the expression of the variants on the surface. Therefore, the variants could be used as the library template. Its disadvantage is that such mutant may not have been reported.

Approach Three: Combinatorial Binding Increases the ELISA Signals

The affinity clamping technology has been developed by which a low-affinity peptide-binding domain is combined with a second domain to enhance affinity (Hard et al., 2018) dramatically. For example, a fibronectin type III domain (FN3) was attached to Grb2 SH2 so that a target peptide can be "clamped" with the improved affinity (Yasui et al., 2014). Inspired by this technology, we investigated whether Cbx1 is functionally displayed by joining it with the Fyn SH2 domain. We reasoned that binding affinity for the domain combination to the mixture of pY and H3K9me3 peptides is stronger than that of the mixture of pY and unmodified peptides. Accordingly, we fused both Fyn SH2 and Cbx1 to the capsid protein P3 and examined their binding to the immobilized mixture of pY and H3K9me3 peptides in different ratios, compared with the mixture of pY and the unmodified peptides in the same ratios. **Figure 4** showed that supernatant phage ELISA signals (0.90) were solid and similar between the Cbx1-and-FynSH2-containing phage and the FynSH2-containing phage binding to the pY peptides, whereas the signals (0.11) were weak for both phages binding to the H3K9me3 peptides. It indicates that the Fyn SH2 domain is displayed on the phage surface. As the proportion of pY peptide in the modified peptide mixture decreased from 2/3 to 1/2 to 1/3, the ELISA signals were similar and high (0.77 ~ 0.87). By contrast, the signals significantly decreased from 0.91 to 0.12 with the proportion of the pY peptide in the mixture of pY and unmodified peptides (**Figure 4**). The significant differences of ELISA signals between two distinct peptide mixtures reveal that Cbx1 is functionally displayed on the phage surface. In the modified peptide mixture, pY and H3K9me3 peptides bind to the Cbx1-and-FynSH2-containing phage and the combinatorial binding increases the ELISA signals. By contrast, only the pY peptide contributes to the binding in another mixture, and the related ELISA signals are relatively low. In summary, it is feasible to examine the expression of a low-affinity protein through the assistance of another protein-target interaction.

Estimation of These Approaches for Measuring Protein Expression With Relatively Low Affinity

We have introduced three approaches to determining whether Cbx1 can be well expressed on the phage surface through

the ELISA test by testing its binding to the H3K9me3 peptide. We wondered if these methods could be applied to detect specific binding to a peptide with lower affinity than the Cbx1 and H3K9me3 complex. The binding of Cbx1 to the monomethyllysine peptide H3K9me1 has the K_d value (12.5 μM) four times larger than that (K_d = 3.2 μM) for the Cbx1 and H3K9me3 complex. We supposed that we only knew the Cbx1-H3K9me1 complex and used it to test the expression of the displayed Cbx1. We performed the binding of Cbx1 and the H3K9me3 peptide (**Figure 2**) and the H3K9me3 peptide (**Figure 5A**) simultaneously using phage ELISA. **Figure 5A** showed that high phage concentration failed to distinguish between the binding of Cbx1 to the H3K9me1 peptide and the H3K9 peptide, although the former had slightly stronger ELISA signals compared to the latter. Additionally, the approaches of protein mutation (**Figure 5B**) and combinatorial binding with the Fyn SH2 domain (**Figure 5C**) could increase the ELISA signals and show a significant difference compared with the negative control. These results suggest that the phage concentration approach may not be helpful to measure the proteins with much low binding affinity to their targets, whereas these proteins could be gauged using the mutation approach or the combinatorial binding approach.

DISCUSSION

Protein engineering can modulate the affinity of a protein to its target. Besides antibody engineering, natural modification-binding proteins are generally engineered to bind to pan-modified targets with enhanced affinity through surface display technologies. For example, the protein Af1521 was tailored to enrich the ADP-ribosylated proteome (Nowak et al., 2020), and the Fyn SH2 domain was engineered to enrich tyrosine phosphoproteome (Liu et al., 2019; Veggiani et al., 2019). As the natural protein has been identified to interact with the target with a certain affinity, their interaction can be used to measure whether the protein fused with the surface protein (e.g., phage P3, P8 or P9) is well and functionally expressed. Some of the protein-target complexes (e.g., the Fyn SH2 and MidT-pY324 complex) have relatively high affinity (e.g., K_d < 1 μM) so that the interactions are easily detected through supernatant phage ELISA. By contrast,

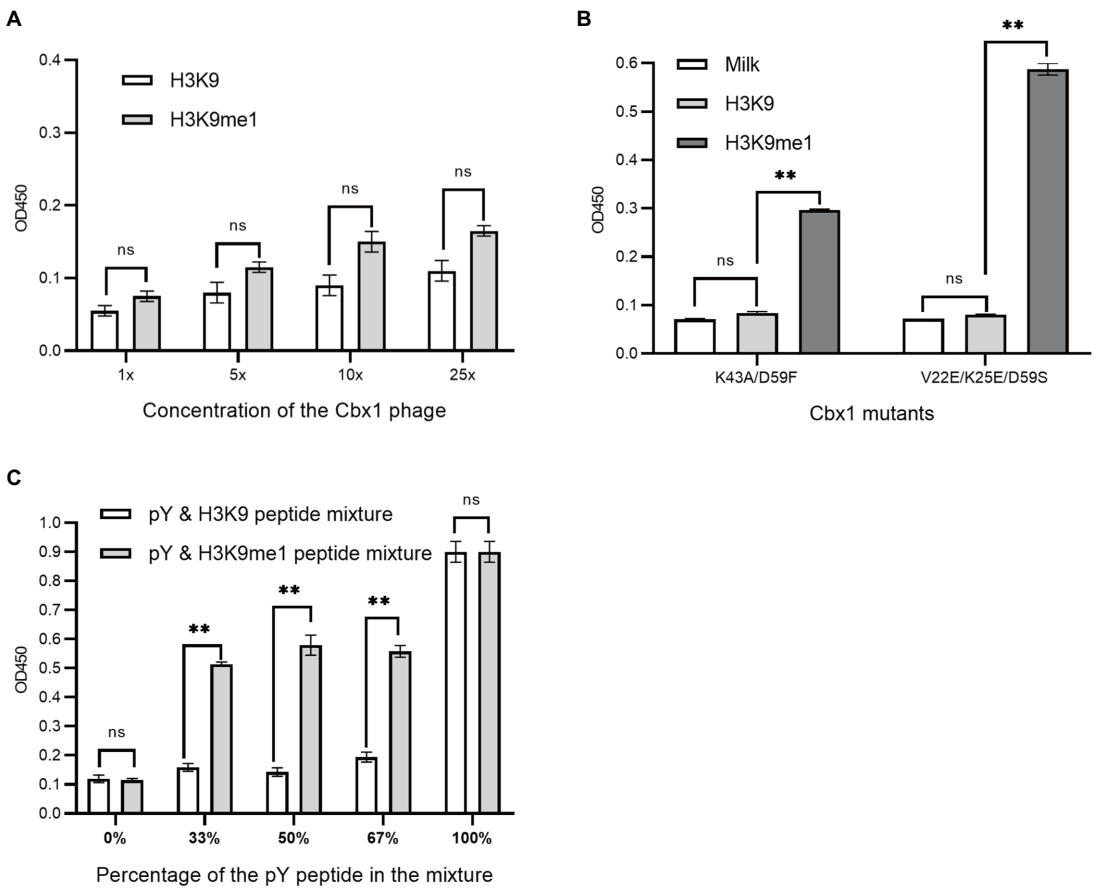


FIGURE 5 | Phage ELISA showed the binding to the H3K9me1 peptide of the Cbx1 protein with different phage concentrations **(A)** or supernatant Cbx1-mutant-containing phages **(B)** or the binding of Cbx1-and-FynSH2-containing phage **(C)** to the mixtures of pY and H3K9me1 peptides with different ratios. The mixtures of pY and H3K9 peptides were considered as negative controls. The experiments were repeated three times.

others (e.g., the Cbx1 and H3K9me3 complex) have a poor affinity (e.g., $K_d > 1\,\mu M$), and their weak interactions are facilely separated during elution and thus difficult to be identified using the simple supernatant phage ELISA approach. To solve this problem, we developed different approaches and compared them using, as the study case, the phage-displayed human Cbx1, which interacts with lysine methylation (Liu et al., 2010).

A straightforward approach is the increase of phage concentration. It indeed increases the ELISA signal but amplifies the background signal too. Therefore, this method may be suitable for detecting interactions with a medium affinity but not a weak affinity. Second, if the protein of interest has been reported to have the variant with enhanced affinity, the variant can be tethered to the surface protein, and its binding to the target can be tested using supernatant phage ELISA or concentrated phage ELISA. If the above two methods fail, we suggested fusing a small protein (or a peptide) with a known high affinity to its target to the phage-displayed protein and testing their binding to the mixture of their targets using ELISA. The rationale is that the affinity for the combined proteins binding to the mixture of both targets is stronger than that to the mixture of the target of the small protein and one non-target molecule because the former mixture has a greater valency to the latter mixture. Indeed, the Cbx1-and-FynSH2-containing phages showed higher ELISA signals when binding to the pY and H3K9me3 peptide mixture than binding to the pY and unmodified peptide mixture. The limitation of this method is that the length of the combined proteins should be considered as large proteins may be inefficiently displayed. The Fyn SH2 domain comprises approximately 100 amino acids, affecting its practical application. The Fyn SH2 domain and pY complex may be replaced by the complex of a short peptide and a protein with a similar affinity. For instance, the Fyn SH3 domain and the class I peptide (VSLARRPLPPLP) have a similar affinity [$K_d = 0.18\,\mu M$ (Ben-David et al., 2019)] to that ($K_d = 0.327\,\mu M$) of the Fyn SH2 domain and MidT-pY324 peptide. Therefore, the fused SH2 domain may be replaced by the class I peptide and, accordingly, the pY peptide could be substituted by the Fyn SH3 domain.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

LL and RL supervised the whole project. CM, DZ, CJ, and SL designed the experiments. CM created the figures and tables. LL and CM wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by funds from the National Natural Science Foundation of China (Grant Nos. 31770821 and 32071430 to LL).

REFERENCES

Albanese, K. I., Krone, M. W., Petell, C. J., Parker, M. M., Strahl, B. D., Brustad, E. M., et al. (2020). Engineered reader proteins for enhanced detection of methylated lysine on histones. *ACS Chem. Biol.* 15, 103–111. doi: 10.1021/acschembio.9b00651

Ben-David, M., Huang, H., Sun, M. G. F., Corbi-Verge, C., Petsalaki, E., Liu, K., et al. (2019). Allosteric modulation of binding specificity by alternative packing of protein cores. *J. Mol. Biol.* 431, 336–350. doi: 10.1016/j.jmb.2018.11.018

Bilal, M., Iqbal, H. M. N., Guo, S., Hu, H., Wang, W., and Zhang, X. (2018). State-of-the-art protein engineering approaches using biological macromolecules: a review from immobilization to implementation view point. *Int. J. Biol. Macromol.* 108, 893–901. doi: 10.1016/j.ijbiomac.2017.10.182

Boder, E. T., and Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* 15, 553–557. doi: 10.1038/nbt0697-553

Francisco, J. A., Campbell, R., Iverson, B. L., and Georgiou, G. (1993). Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. U. S. A.* 90, 10444–10448. doi: 10.1073/pnas.90.22.10444

Hard, R., Li, N., He, W., Ross, B., Mo, G. C. H., Peng, Q., et al. (2018). Deciphering and engineering chromodomain-methyllysine peptide recognition. *Sci. Adv.* 4:eaau1447. doi: 10.1126/sciadv.aau1447

Kaneko, T., et al. (2012). Superbinder SH2 domains act as antagonists of cell signaling. *Sci. Signal.* 5:ra68.

Li, S., Zou, Y., Zhao, D., Yin, Y., Song, J., He, N., et al. (2021). Revisiting the phosphotyrosine binding pocket of Fyn SH2 domain led to the identification of novel SH2 superbinders. *Protein Sci.* 30, 558–570. doi: 10.1002/pro.4012

Liu, H., Galka, M., Iberg, A., Wang, Z., Li, L., Voss, C., et al. (2010). Systematic identification of methyllysine-driven interactions for histone and nonhistone targets. *J. Proteome Res.* 9, 5827–5836. doi: 10.1021/pr100597b

Liu, H., Huang, H., Voss, C., Kaneko, T., Qin, W. T., Sidhu, S., et al. (2019). Surface loops in a single SH2 domain are capable of encoding the spectrum of specificity of the SH2 family. *Mol. Cell. Proteomics* 18, 372–382. doi: 10.1074/mcp.RA118.001123

Nowak, K., Rosenthal, F., Karlberg, T., Bütepage, M., Thorsell, A. G., Dreier, B., et al. (2020). Engineering Af1521 improves ADP-ribose binding and identification of ADP-ribosylated proteins. *Nat. Commun.* 11:5199. doi: 10.1038/s41467-020-18981-w

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317. doi: 10.1126/science.4001944

Smith, A. G., and Kaiser, P. K. (2016). Therapeutic monoclonal antibodies and fragments: Ranibizumab. *Dev. Ophthalmol.* 55, 246–251. doi: 10.1159/000431200

Tadayoni, R., Sararols, L., Weissgerber, G., Verma, R., Clemens, A., and Holz, F. G. (2021). Brolucizumab: a newly developed anti-VEGF molecule for the treatment of Neovascular age-related macular degeneration. *Ophthalmologica* 244, 93–101. doi: 10.1159/000513048

Tonikian, R., Zhang, Y., Boone, C., and Sidhu, S. S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat. Protoc.* 2, 1368–1386. doi: 10.1038/nprot.2007.151

Veggiani, G., Huang, H., Yates, B. P., Tong, J., Kaneko, T., Joshi, R., et al. (2019). Engineered SH2 domains with tailored specificities and enhanced affinities for phosphoproteome analysis. *Protein Sci.* 28, 403–413. doi: 10.1002/pro.3551

Yasui, N., Findlay, G. M., Gish, G. D., Hsiung, M. S., Huang, J., Tucholska, M., et al. (2014). Directed network wiring identifies a key protein interaction in embryonic stem cell differentiation. *Mol. Cell* 54, 1034–1041. doi: 10.1016/j.molcel.2014.05.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 June 2022
doi: 10.3389/fmicb.2022.901558

Engineering of Src Homology 2 Domain Leading to Sulfotyrosine Recognition With a High Affinity by Integrating a Distinctive Selection Theme and Next-Generation Sequencing

Dongping Zhao[1,2†], Chan Li[1†], Haoqiang Jiang[1], Yuqing Yin[2], Changjing Zhou[3], Haiming Huang[2], Yunkun Qi[4] and Lei Li[1]**

[1] School of Basic Medicine, Qingdao University, Qingdao, China, [2] Department of Drug Discovery, Noventi Biopharmaceuticals Co., Ltd., Shanghai, China, [3] Sino Genomics Technology Co., Ltd., Qingdao, China, [4] School of Pharmacy, Qingdao University, Qingdao, China

Tyrosine sulfation plays a vital role in various biochemical reactions. Although sulfated tyrosine (sTyr) has a similar structure to phosphotyrosine (pTyr), the number of available sTyr sites is significantly less than that of pTyr sites, mainly because of the lack of effective sTyr probes. A few sTyr binders were identified on the basis of structural similarity by engineering the pTyr-binding pocket of an Src Homology 2 (SH2) domain through phage selections against sTyr peptides. Nevertheless, they still interact with pTyr peptides with comparable affinity. This study aims to identify sTyr superbinders using the SH2 domain as a template. We created a distinctive phage selection scheme that separately covered selections against sTyr and pTyr peptides, followed by next-generation sequencing (NGS). After selections, phage pools showed strong enzyme-linked immunosorbent assay (ELISA) signal intensities for both modified peptides, indicating that the variants evolved with a high affinity for these peptides, which causes difficulty in identifying sTyr-specific binders. In contrast, NGS data from selected pools showed significant differences, suggesting the enrichment of sTyr-specific variants during selections. Accordingly, we obtained the sTyr features based on NGS data analysis and prioritized a few potential sTyr binders. The variant SH2-4 showed a stronger affinity for sTyr than pTyr and was superior to previous sTyr binders as measured by the Biolayer Interferometry assay. In summary, we described the strategy of integrating NGS data mining with a novel selection scheme to identify sTyr superbinders.

Keywords: phage display, SH2 domain, next-generation sequencing, tyrosine sulfation, directed evolution

OPEN ACCESS

Edited by:
Jian Huang,
University of Electronic Science and Technology of China, China

Reviewed by:
Wensheng Qin,
Lakehead University, Canada
Chenfang Dong,
Zhejiang University, China
Yue Chen,
University of Minnesota Twin Cities, United States

***Correspondence:**
Yunkun Qi
qiyunkun@163.com
Lei Li
lileime@hotmail.com

† These authors share first authorship

Specialty section:
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

Received: *22 March 2022*
Accepted: *25 April 2022*
Published: *03 June 2022*

Citation:
Zhao D, Li C, Jiang H, Yin Y, Zhou C, Huang H, Qi Y and Li L (2022) Engineering of Src Homology 2 Domain Leading to Sulfotyrosine Recognition With a High Affinity by Integrating a Distinctive Selection Theme and Next-Generation Sequencing. Front. Microbiol. 13:901558. doi: 10.3389/fmicb.2022.901558

INTRODUCTION

Protein tyrosine O-sulfation is a common type of post-translational modifications (PTMs) that occur in transmembrane and secreted proteins. Tyrosine sulfation, catalyzed by two tyrosyl protein sulfotransferase enzymes TPST-1 and TPST-2, plays a vital role in several cell signal pathways and biochemical reactions (Moore, 2003, 2009; Leung et al., 2016; Naider and Anglister, 2018).

In AIDS, the interaction between HIV and CCR5 is affected by tyrosine sulfation of CCR5 as the mutation of sulfated tyrosine (sTyr) to phenylalanine significantly reduces HIV infection rates (Choe et al., 1996).

Up to 1% of tyrosine residues in animals are predicted to be sulfated (Huttner, 1988). Due to the lack of biomolecular probes to effectively enrich sulfated tyrosine, several sTyr-containing proteins or sTyr sites have been reported. Available anti-sTyr antibodies are specific for sTyr over phosphorylated tyrosine (pTyr), but their binding affinities are moderate (IC50 = 1 μM or 1.25 mM) (Hoffhines et al., 2006; Kehoe et al., 2006). In contrast, although sTyr and pTyr have a similar structure (**Supplementary Figure 1**), anti-pTyr antibodies have an excellent affinity for pTyr, and tens of thousands of pTyr sites have been documented (Hornbeck et al., 2019). Recently, the Src Homology 2 (SH2) domain (about 100 amino acids) has been engineered as an SH2 superbinder (i.e., SH2-trm; Kaneko et al., 2012; **Figure 1B**) by phage display to recognize phosphotyrosine (pTyr) through the evolution of its pTyr-binding pocket and has been shown to be superior to anti-pTyr antibodies (Bian et al., 2016). Using structural similarities between phosphorylation and sulfation (**Supplementary Figure 1**), the pTyr-binding pocket of SH2-trm was engineered to identify sTyr (Ju et al., 2016; Lawrie et al., 2021). A couple of SH2 variants (e.g., SH2-58.6, **Figure 1B**) were reported to bind to sTyr, but they also recognize pTyr with a moderate affinity (Ju et al., 2016; Lawrie et al., 2021; Li et al., 2021). All of the variants described above are derived from the human Fyn or Src SH2 domain, both belonging to the Src family and with high homology (e.g., 66% sequence identity) (Liu et al., 2006).

In this study, we engineered the sTyr binder SH2-58.6 to identify SH2 variants with improved affinity and specificity for sTyr. Our strategy is illustrated in **Figure 1A**. There are three differences between this study and previous studies. First, previous sTyr binders were developed by randomizing amino acid residues at four or five positions in the pTyr-binding pocket (Ju et al., 2016; Lawrie et al., 2021). In contrast, we increased the number of evolvable positions to 11 because more variable residues might result in variants of higher affinity and specificity. Second, the reported sTyr binders evolved through positive selections against sTyr peptides or integrated with negative selections against the pTyr counterpart. As sTyr binders interact with both sTyr and pTyr peptides, we doubt the effectiveness of these selection strategies. Accordingly, we separately developed a distinctive phage selection theme, which included positive selections against both sTyr and pTyr peptides. Variants that were enriched during selections on sTyr but depleted during selections against pTyr were considered to have a high affinity and specificity for sTyr. Third, in previous studies, sTyr binders were identified using random selections of clones from the phage pool, so that variants with high specificity might be omitted. We performed next-generation sequencing (NGS) of the rounds of selections and performed NGS data analysis to prioritize variants. Finally, we identified the SH2 variant SH2-4, which has improved affinity and specificity for sTyr and is superior to the documented sTyr binders. In summary, we described the strategy for integrating the novel selection scheme and NGS and model construction to identify sTyr superbinders. We expect that they will be used to enrich sTyr peptides and boost the study of tyrosine sulfation on a proteomic scale.

RESULTS

Construction of the SH2 Domain Library and Its Quality Control by Next-Generation Sequencing

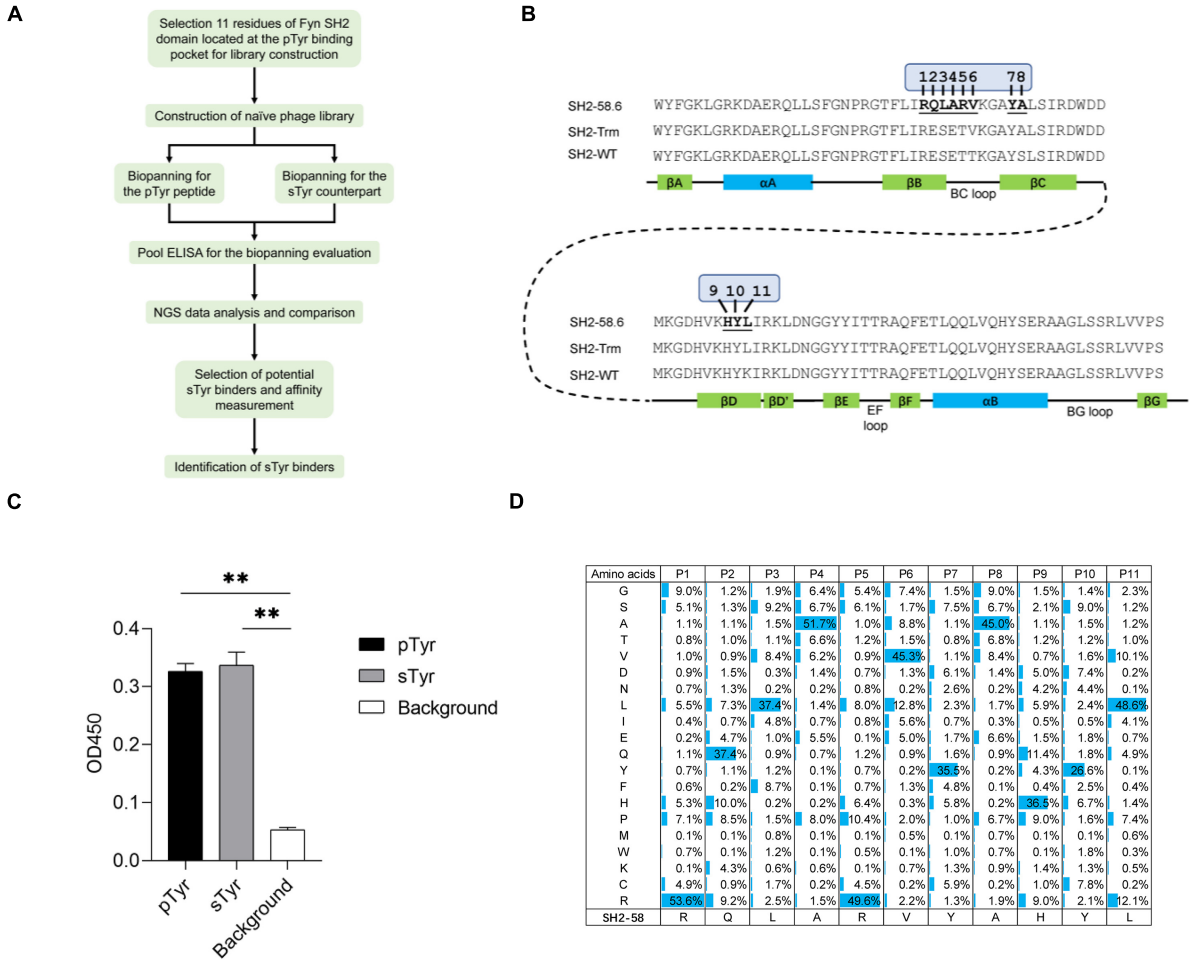
Two SH2 variants (SH2-58.6 and SH2-60.1) were documented as specific for sTyr over pTyr (Ju et al., 2016). We selected the variant SH2-58.6 as a template for constructing the SH2 domain library. We generated Fyn SH2-58.6 (**Figure 1B**) in the pFN-OM6 phagemid and found that this variant could bind to pTyr and sTyr peptides (**Figure 1C**). Next, we randomized 11 positions on the pTyr-binding pocket, including those selected for recognizing sTyr binders (Ju et al., 2016; Lawrie et al., 2021), to identify SH2 variants with improved affinity and specificity for sTyr. By choosing more variable positions than before, we attempted to understand how pocket residues evolved to allow tighter binding to a sTyr ligand. To this end, residues at these positions were systematically altered using the soft randomization mutagenesis technique (Li et al., 2021), which allows a 50% mutation rate at each position. The library capacity estimated by monoclonal titration is about 1.0×10^9. This library was defined as the naïve library for the following biopanning.

We amplified the DNA sequences of SH2 variants and performed NGS to characterize the library. As a result, 1.62 million high-quality DNA sequences coding Fyn SH2 variants were retrieved from the sequencing data. Of these sequences identified, 1.54 million (95%) have the designed mutations in 11 evolvable positions, while the rest of them have unexpected mutations, likely due to mutations introduced by PCR amplification or sequencing errors. Further analysis of the amino acid distribution of Fyn SH2 variants showed that the actual mutation of the variant library was in line with the theoretical design with minor exceptions, as shown in **Figure 1D**.

Biopanning

This study used two modified peptides to identify sY superbinders for biopanning: sY peptide and pY counterpart. sY peptide was used to discover sY superbinders from SH2 domain-based phage libraries (Ju et al., 2016; Lawrie et al., 2021). Similarly, pY peptide was employed to explore pY superbinders from SH2 domain-based phage libraries (Li et al., 2021). Therefore, we selected these two modified peptides and constructed an SH2 domain-based phage library for this study.

In previous studies, SH2 variants as sTyr binders evolved by positive selections against sTyr peptide alone or in combination with negative selections against pTyr counterpart (Ju et al., 2016; Lawrie et al., 2021). These selection schemes are the common strategies based on the hypothesis that specific variants of sTyr rather than pTyr are enriched during selections. Nevertheless, all identified sTyr binders can bind pTyr with moderate affinities (Ju et al., 2016; Lawrie et al., 2021). This indicates that SH2 variants



Amino acids	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
G	9.0%	1.2%	1.9%	6.4%	5.4%	7.4%	1.5%	9.0%	1.5%	1.4%	2.3%
S	5.1%	1.3%	9.2%	6.7%	6.1%	1.7%	7.5%	6.7%	2.1%	9.0%	1.2%
A	1.1%	1.1%	1.5%	51.7%	1.0%	8.8%	1.1%	45.0%	1.1%	1.5%	1.2%
T	0.8%	1.0%	1.1%	6.6%	1.2%	1.5%	0.8%	6.8%	1.2%	1.2%	1.0%
V	1.0%	0.9%	8.4%	6.2%	0.9%	45.3%	1.1%	8.4%	0.7%	1.6%	10.1%
D	0.9%	1.5%	0.3%	1.4%	0.7%	1.3%	6.1%	1.4%	5.0%	7.4%	0.2%
N	0.7%	1.3%	0.2%	0.2%	0.8%	0.2%	2.6%	0.2%	4.2%	4.4%	0.1%
L	5.5%	7.3%	37.4%	1.4%	8.0%	12.8%	2.3%	1.7%	5.9%	2.4%	48.6%
I	0.4%	0.7%	4.8%	0.7%	0.8%	5.6%	0.7%	0.3%	0.5%	0.5%	4.1%
E	0.2%	4.7%	1.0%	5.5%	0.1%	5.0%	1.7%	6.6%	1.5%	1.8%	0.7%
Q	1.1%	37.4%	0.9%	0.7%	1.2%	0.9%	1.6%	0.9%	11.4%	1.8%	4.9%
Y	0.7%	1.1%	1.2%	0.1%	0.7%	0.2%	35.5%	0.2%	4.3%	26.6%	0.1%
F	0.6%	0.2%	8.7%	0.1%	0.7%	1.3%	4.8%	0.1%	0.4%	2.5%	0.4%
H	5.3%	10.0%	0.2%	0.2%	6.4%	0.3%	5.8%	0.2%	36.5%	6.7%	1.4%
P	7.1%	8.5%	1.5%	8.0%	10.4%	2.0%	1.0%	6.7%	9.0%	1.6%	7.4%
M	0.1%	0.1%	0.8%	0.1%	0.1%	0.5%	0.1%	0.7%	0.1%	0.1%	0.6%
W	0.7%	0.1%	1.2%	0.1%	0.5%	0.1%	1.0%	0.7%	0.1%	1.8%	0.3%
K	0.1%	4.3%	0.6%	0.6%	0.1%	0.7%	1.3%	0.9%	1.4%	1.3%	0.5%
C	4.9%	0.9%	1.7%	0.2%	4.5%	0.2%	5.9%	0.2%	1.0%	7.8%	0.2%
R	53.6%	9.2%	2.5%	1.5%	49.6%	2.2%	1.3%	1.9%	9.0%	2.1%	12.1%
SH2-58	R	Q	L	A	R	V	Y	A	H	Y	L

FIGURE 1 | The construction and quality control of a Fyn Src Homology 2 (SH2) domain variant library. **(A)** A workflow diagram of the selection of the sulfated tyrosine (sTyr) binders. **(B)** The amino acid sequence and secondary structure of the Fyn SH2 domain (SH2-WT), the SH2 superbinder (SH2-trm) for phosphotyrosine (pTyr) (Kaneko et al., 2012), and the variant SH2-58.6 as the library template in this study (Ju et al., 2016). The 11 evolvable residues were numbered on SH2-58.6. **(C)** Cross-reactivity of SH2-58.6 binding to pTyr peptide and sTyr counterpart measured by phage enzyme-linked immunosorbent assay (ELISA). Streptavidin precoated in the microplate was taken as the background. Three replicates were performed, and the *p*-value was performed by Student's *t*-test (**$p < 0.01$). **(D)** The amino acid distribution at evolvable positions was calculated based on the next-generation sequencing (NGS) data. The corresponding residues of the template SH2-58.6 were shown at the bottom.

that bind to both modifications are challenging to filter out using such selection schemes. To solve this problem, we developed a novel selection theme to find SH2 variants with a superior affinity and specificity for sTyr. This novel selection theme included selections against both pTyr and sTyr peptides separately, NGS sequencing of the selection results, and the comparison between their NGS data to prioritize potential sTyr binders with a high affinity. We hypothesized that such sTyr binders were depleted in the process of selections against the pTyr target but enriched in the selections against the sTyr target.

The library was subjected to phage biopanning against a biotinylated pTyr peptide derived from the protein MidT, a cognate ligand of wild-type Fyn SH2 (Dunant et al., 1997). The phage library was preincubated with the biotin-labeled pTyr peptide, which was immobilized in a microplate precoated with streptavidin. Bound phages were eluted, amplified, and applied as input to the next round of panning. After four rounds of panning, the amplified phage pools from these rounds were applied to an enzyme-linked immunosorbent assay (ELISA) to test their binding to pTyr peptide and sTyr counterpart. **Figure 2A** shows that the phages are explicitly bound to pTyr peptide after the second round and interact with pTyr and sTyr peptides after the third and fourth rounds. This indicates that SH2 variants that interact with both modified peptides were enriched in the round of panning.

Similarly, the library was subject to phage biopanning against the biotinylated sTyr peptide. **Figure 2B** shows that phage pools did not bind to either peptide until the fourth round, at which interactions of the phages with both modified peptides had similar signal intensities. Unexpectedly, the apparent background signal intensity in the fourth round indicates that non-specific phages that bind to the background was also enriched

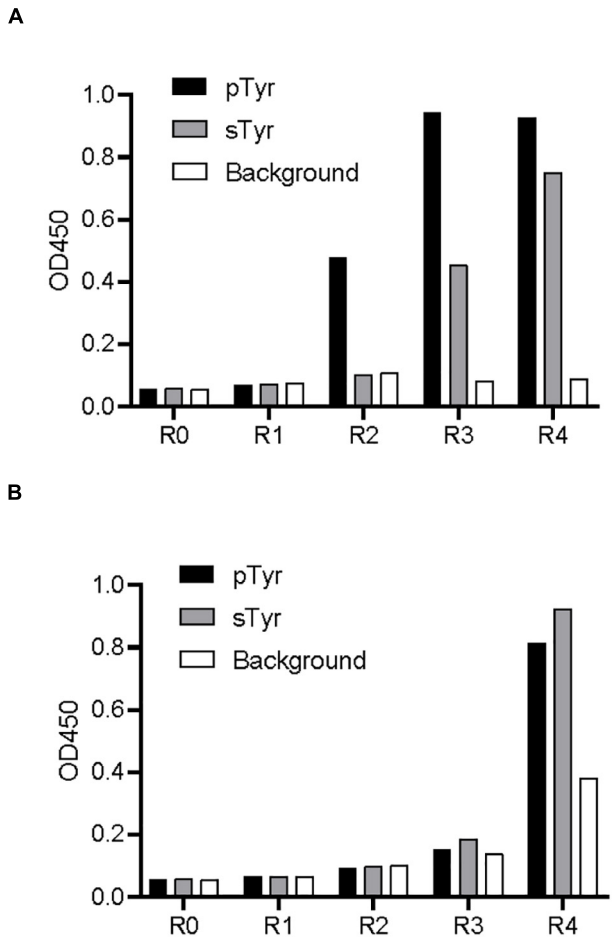


FIGURE 2 | Biopanning of the SH2 variant library against pTyr and sTyr peptides. **(A)** Phage ELISA of the selected phages from each biopanning round against pTyr and the naïve library binding to pTyr peptide and sTyr counterpart. R0 refers to the naïve library. R1–R4 stands for phage panning cycle Rounds 1–4. **(B)** Phage ELISA of the selected phages from each biopanning round against sTyr and the naïve library binding to pTyr peptide and sTyr counterpart. Streptavidin precoated in the microplate was taken as the background.

(**Figure 2B**). The enrichment of non-specific phages and those bound to both modified targets suggest the difficulty of finding sTyr-specific binders through random selections from pools.

Computational Analysis of Next-Generation Sequencing Data to Prioritize Potential Sulfated Tyrosine Binders

Enzyme-linked immunosorbent assay showed similar intensities of pTyr and sTyr binding to phage pools in the fourth selection against the pTyr target and the sTyr counterpart (**Figure 2**). Such observations raised the question, "what are the differences between the phage pools biopanned using pTyr and sTyr targets?" We figured it out by performing NGS of phage-displayed SH2 variants panned from all rounds. First, we investigated the residue types that were enriched or depleted by panning against the pTyr target. To this end, we calculated proportion changes in the residue type at each evolvable position from the naïve library to the library after the fourth round of selection (**Figure 3A**). The residue type was defined as enriched if its proportion change was greater than 5% or depleted if it was less than −5%. **Figure 3A** shows the enrichment or depletion of various residue types at different positions. The results were similar if the naïve library was compared with the library after the third round (**Supplementary Figure 2**). Next, after the fourth round, we performed a similar analysis for panning against the sTyr target by comparing the naïve library with the selected library. Similarly, many residue types were enriched or depleted at distinct positions. A comparison of **Figures 3A,B** illustrated that many residue types were enriched for selections against both modified targets, whereas a few showed the opposite trend. For instance, the amino acids R@P1, C@P2, S@P3, E@P4, Y@P7, V@P8, H@P9, Y@P10, and L@P11 were enriched in both selections (**Figure 3**). In contrast, the amino acid S@P5 was enriched by panning against pTyr but depleted by panning using sTyr, while the amino acid L@P6 was the opposite (**Figure 3**). These observations indicate that while both phage libraries panned against sTyr and pTyr have similar enriched residues at many positions, each library has its characteristics. Therefore, SH2 variants specific to sTyr are likely to exist in the library panned against sTyr.

We hypothesized that sTyr-specific binders had three features: (1) their percentages in the pool tend to increase during selections against sTyr but decrease for selections against pTyr; (2) they are liable to include a large number of residues enriched in selections against sTyr; and (3) the number of such enriched residues is more significant than that of enriched residues in selections against pTyr. We took advantage of the NGS data to identify variants with these features. Practically, we only considered variants as potential sTyr-specific binders whose proportion was larger than 0.1% at the last few rounds of selection against sTyr. Additionally, we defined the term "enrichment index (EI)" to quantify the enrichment status of a given variant. For each position, an enriched residue type (proportion change $>$ 5%) corresponds to the value of 1, a depleted residue type (proportion change $<$ −5%) is related to −1, and the other residue types are 0 (**Figures 3A,B**). The EI value is the sum of the values at the evolvable positions. For instance, the variant with "RCSEGLYVHYL" had the sTyr EI value as 11 but 8 for pTyr (**Figures 3A,B**). A potential sTyr-specific binder is likely to have a considerable sTyr EI value, but a small pTyr EI value. In other words, such a binder has a significant EI difference between the sTyr and pTyr EI values. Accordingly, we used the sTyr EI value and EI difference and the tendency of the proportion in selections to prioritize SH2 variants.

We predicted and selected 12 variants to examine their affinity for pTyr and sTyr peptides. These variants were classified into three categories: the first four (SH1-1–SH2-4) were potential sTyr-specific binders, the second four (SH2-5–SH2-8) were sTyr-specific non-binders, and the remaining four (SH2-58.6 to SH2-WT) were derived from the literature (Ju et al., 2016; Lawrie et al., 2021) for comparison (**Table 1**). SH1-1–SH2-4 variants had large

A

Amino acids	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
A	-1.1%	0.2%	1.4%	1.7%	1.0%	-8.1%	-1.1%	35.0%	-1.1%	-1.5%	-1.2%
C	-4.8%	80.9%	1.7%	0.0%	4.4%	-0.2%	8.5%	-0.2%	-0.9%	-7.2%	-0.2%
D	-0.9%	-1.4%	-0.3%	1.1%	-0.7%	-1.3%	-6.0%	-1.4%	-4.9%	-7.3%	-0.2%
E	-0.2%	-2.8%	-1.0%	5.4%	-0.1%	-4.1%	-1.7%	-6.4%	-1.4%	-1.7%	-0.7%
F	-0.6%	-0.1%	8.5%	-0.1%	-0.7%	-1.3%	-4.4%	-0.1%	-0.4%	-2.4%	-0.4%
G	-8.9%	-1.2%	-1.9%	4.6%	5.2%	-6.9%	-1.5%	-8.8%	-1.4%	-1.4%	-2.2%
H	-5.2%	-9.1%	-0.2%	-0.2%	-6.2%	0.0%	-2.9%	-0.2%	61.9%	-6.0%	-1.3%
I	-0.4%	-0.6%	4.7%	-0.6%	-0.8%	-4.4%	-0.7%	-0.1%	-0.5%	-0.5%	-3.8%
K	-0.1%	-4.0%	-0.6%	-0.6%	-0.1%	-0.6%	-1.3%	-0.9%	-1.3%	-1.3%	-0.5%
L	-5.4%	-4.4%	-36.6%	-1.3%	-7.8%	-9.1%	4.6%	-1.7%	-5.7%	-2.4%	41.1%
M	-0.1%	-0.1%	-0.8%	-0.1%	-0.1%	-0.5%	1.2%	-0.7%	-0.1%	-0.1%	-0.6%
N	-0.7%	-1.2%	-0.2%	-0.2%	-0.8%	-0.2%	-2.6%	-0.2%	-4.1%	-4.4%	-0.1%
P	-7.0%	-7.6%	-1.5%	-6.1%	-10.3%	-2.0%	-1.0%	-6.6%	-8.8%	-1.6%	-6.9%
Q	-1.1%	-28.3%	-0.8%	-0.6%	-1.1%	-0.8%	-1.6%	-0.9%	-11.1%	-1.8%	-4.7%
R	45.5%	-8.6%	-2.5%	-1.3%	-34.0%	-2.0%	-1.3%	-1.8%	-8.7%	-2.0%	-3.5%
S	-5.0%	-1.1%	89.2%	-5.3%	75.8%	-1.6%	-7.4%	-6.6%	-2.1%	-8.8%	-1.2%
T	-0.8%	-1.0%	-1.0%	12.2%	0.2%	-1.4%	-0.8%	-6.7%	-1.2%	-1.2%	-0.9%
V	-1.0%	-0.8%	-8.3%	-5.4%	-0.9%	46.5%	-1.1%	10.5%	-0.7%	-1.5%	-9.6%
W	-0.7%	-0.1%	-1.2%	-0.1%	-0.4%	-0.1%	-1.0%	-0.7%	-0.1%	-1.7%	-0.3%
Y	-0.7%	-1.0%	-1.2%	-0.1%	-0.7%	-0.2%	38.0%	-0.2%	-4.2%	71.1%	-0.1%
Enriched	R	C	S	AET	S	V	CY	AV	H	Y	L
Depleted	GHLPS	HPQR	FLV	SPV	GHLPR	AGL	DS	EGPST	LPQR	CDHS	PV

B

Amino acids	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
A	-0.6%	-0.6%	1.2%	-2.5%	-0.4%	-7.0%	-0.9%	-10.3%	-0.6%	-1.1%	-0.1%
C	-3.0%	15.0%	1.5%	1.0%	-2.3%	-0.1%	-1.5%	2.5%	5.1%	39.8%	1.4%
D	-0.9%	-1.3%	-0.3%	-0.7%	-0.7%	-1.3%	-5.7%	-0.8%	-3.6%	-6.4%	-0.1%
E	-0.2%	-4.3%	-1.0%	16.9%	-0.1%	-4.8%	-1.6%	-4.6%	-1.1%	-1.5%	-0.6%
F	-0.4%	0.8%	-2.6%	0.0%	-0.1%	0.8%	2.8%	0.0%	-0.1%	-1.6%	-0.1%
G	-6.5%	-1.1%	1.8%	-4.7%	5.2%	-6.9%	-1.3%	0.1%	-0.9%	-0.8%	-1.1%
H	-1.5%	-7.7%	-0.2%	0.0%	-3.4%	-0.2%	12.2%	-0.1%	13.1%	-4.7%	-0.7%
I	-0.4%	1.0%	-1.1%	-0.4%	-0.5%	4.0%	-0.4%	-0.2%	-0.4%	-0.5%	-1.8%
K	-0.1%	-4.0%	-0.5%	-0.5%	-0.1%	-0.6%	-1.2%	-0.7%	-0.2%	-1.2%	-0.1%
L	-4.4%	44.0%	-14.1%	0.5%	0.3%	21.3%	0.3%	0.5%	-2.6%	-2.0%	8.2%
M	0.0%	0.5%	-0.1%	0.1%	0.0%	0.4%	0.0%	0.1%	-0.1%	-0.1%	-0.1%
N	-0.3%	-8.8%	-0.1%	0.1%	-0.7%	-0.2%	-2.2%	-0.1%	-1.0%	-3.2%	-0.1%
P	-6.3%	-8.2%	-1.5%	2.5%	-10.2%	-1.9%	-0.9%	-2.5%	-2.7%	-1.3%	-2.8%
Q	-0.7%	-34.3%	9.1%	0.8%	-1.0%	-0.9%	-1.3%	-0.2%	-3.8%	-1.5%	-1.9%
R	23.7%	-8.0%	-2.3%	-1.0%	15.4%	-1.8%	-1.0%	-0.1%	3.0%	-1.5%	8.1%
S	-2.8%	-1.2%	8.2%	-3.2%	-4.7%	-1.6%	-6.4%	-3.1%	-0.9%	-7.0%	-0.5%
T	-0.6%	-0.8%	-1.0%	-2.9%	-1.0%	-1.4%	-0.6%	-3.0%	-0.4%	-0.9%	-0.3%
V	-1.0%	0.4%	-6.5%	-5.0%	-0.7%	3.0%	-0.8%	20.8%	-0.6%	-1.3%	-4.8%
W	6.4%	4.5%	4.7%	0.4%	4.3%	1.1%	4.0%	3.0%	0.1%	0.8%	0.3%
Y	0.5%	2.7%	0.6%	0.1%	1.5%	0.0%	22.6%	-0.1%	0.6%	12.3%	-0.1%
Enriched	RW	CLN	LQS	E	GR	L	HY	V	CH	CY	LR
Depleted	GP	HPQR	V		P	AG	AG	A		DS	

FIGURE 3 | Proportion changes of residue types at evolvable positions from the naïve library to the library at the fourth round of selection against pTyr **(A)** and sTyr **(B)**. Residue types with significant proportion changes (5% as cutoff) were considered enriched or depleted and were listed at the bottom.

sTyr EI values (9 or 10) and EI differences (ranging from 5 to 9), and their proportions constantly increased during selections against sTyr but decreased or had a slight change in selections against pTyr (**Figures 4A–D**). Therefore, these variants were predicted as potential sTyr binders. The variants SH2-5–SH2-8 have minor EI differences (−1 to 2) (**Table 1**). The percentages of SH2-5 and SH2-6 increased in selections against sTyr and pTyr. The percentages of SH2-7 and SH2-8 increased during selections against pTyr but decreased or slightly changed against sTyr (**Figures 4E–H**). Therefore, SH2-5–SH2-8 were not thought of as potential sTyr-specific binders. Moreover, the template SH2-58.6 had the EI difference of 0, and its proportions generally decreased during selections against sTyr and pTyr, indicating that it was overcome during selections. Furthermore, we failed to find sTyr binders (i.e., SH2-1.8 and SH2-3.1) and the wild-type SH2 domain documented in the NGS data. They all had slight EI differences (from −2 to 1), so they were not considered sTyr-specific binders.

SH2-4 Had a Stronger Affinity to Sulfated Tyrosine Than Phosphotyrosine and Was Superior to Others

Using a Biolayer Interferometry assay, we expressed 12 SH2 variants and measured their affinities for the sTyr and pTyr peptides. Interestingly, we successfully expressed nine of them *via* the expression vector pHH0239, except SH2-1–SH2-3. To determine why the three variants failed to express, we attempted to express SH2-1–SH2-4 using the phagemid pFN-OM6 and measure their binding to the modified peptides through phage ELISA. The variants SH2-1–SH2-3 were expressed successfully but were bound to the background, whereas SH2-4 did not interact with the background (**Figure 5**). As they had the common C@P10, we suspected that C@P10 was the reason for solid background ELISA signals in selections against sTyr and explored its association with background signals accordingly (**Figures 2**, **4**). The proportion of C@10-containing variants increased from 7.8% in the naïve library to 47.6% in the fourth round of selection against sTyr and was significantly correlated with background signal intensities (Pearson's $r = 0.88$). In contrast, such a proportion decreased to 0.5% in selections on pTyr, and background signal intensities were constantly low. These observations suggest that C@P10 may bind to the background and result in non-specific binding.

As the last potential sTyr binders, SH2-4 showed a strong affinity for sTyr peptide (K_d = 97 nM) than pTyr peptide (K_d = 131 nM) (**Table 1** and **Supplementary Figure 3A**). Four variants in the second category had weaker affinities for sTyr (K_d = 117–147 nM) than pTyr (K_d = 55–60 nM) (**Supplementary Figures 3B–E**). In the third category, the three documented sTyr binders had a weaker affinity for sTyr (K_d = 190–280 nM) than pTyr (K_d = 140–215 nM) (**Supplementary Figures 3F–H**); SH2-WT bound to pTyr peptide with the affinity K_d as 267 nM but failed to interact with sTyr peptide (**Supplementary Figure 3J**). Therefore, our data analysis method to prioritize sTyr binders is effective. In summary, SH2-4 has a strong affinity for sTyr compared to others and high specificity for sTyr. As previous sTyr binders have been shown to be useful for detecting and enriching sulfoproteins, SH2-4 can potentially be applied to the study of protein tyrosine O-sulfation with proper experimental designs.

Nature commonly utilizes multiple binding sites to interact with targets, including tandem reader domains (Moore et al., 2013). In a previous work where domains were used as detection reagents, the fusion of multiple copies of the same domain was employed to increase the affinity (Albanese et al., 2020). Accordingly, we explored whether linking multiple copies of SH2-4 would provide additional improvements beyond a single copy. Accordingly, two SH2-4 were connected using the Gly4Ser linker. As expected, the tandem SH2-4 had a higher affinity for both modified peptides (K_d: sTyr = 74 nM; pTyr = 90 nM) than for the single SH2-4 (**Figure 3I**). It demonstrates that multiplexing modules can enhance the binding to the target.

DISCUSSION

In this study, the variant SH2-4 was found to be a sTyr superbinder because it had a high affinity and specificity to sulfotyrosine compared to the documented sTyr binders. Unlike previous studies, this identification was based on a novel

TABLE 1 | Residues at the evolvable positions and EI values and binding affinities of Src Homology 2 (SH2) variants to phosphotyrosine (pTyr) and sulfated tyrosine (sTyr) peptides.

SH2[#]	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	Detected in NGS?	EI		EI difference[$]	Kd(nM)	
													(sTyr)	(pTyr)		(sTyr)	(pTyr)
Potential sTyr binders																	
SH2-1	R	L	L	A	R	L	Y	A	H	C	L	Yes	10	2	8	No expressed	
SH2-2	R	L	L	A	R	V	Y	A	H	C	L	Yes	9	4	5	No expressed	
SH2-3	R	L	L	A	R	L	Y	A	Q	C	L	Yes	9	0	9	No expressed	
SH2-4	R	N	Q	P	G	V	Y	V	H	Y	R	Yes	9	4	5	97.9	131
Other variants																	
SH2-5	R	C	S	E	R	V	H	V	H	Y	L	Yes	10	8	2	117	55.3
SH2-6	R	C	S	A	R	V	H	V	H	Y	L	Yes	9	8	1	126	65.4
SH2-7	R	Q	S	A	R	V	F	V	H	Y	L	Yes	6	6	0	145	55.5
SH2-8	R	C	S	E	R	R	Y	V	H	Y	L	Yes	10	8	2	147	61.6
Reported sTyr binders or SH2 wildtype																	
SH2-58.6	R	Q	L	A	R	V	Y	A	H	Y	L	Yes	5	5	0	188	141
SH2-1.8	R	E	A	E	R	V	Y	A	H	Y	L	No	6	7	−1	266	215
SH2-3.1	R	E	H	P	F	V	Y	A	H	Y	L	No	4	6	−2	282	163
SH2-WT	R	E	S	E	T	T	Y	S	H	Y	L	No	6	5	1	ND	267

[#]*SH2 variants were classified into three categories: (1) potential sTyr binders, (2) other variants, and (3) reported sTyr binders plus wild-type SH2 domain.*
[$]*The EI difference means the difference between the sTyr EI value and the pTyr EI value. EI means enrichment index.*
ND, No detected.

strategy that integrates different library constructs, distinctive selection schemes, NGS, and model construction to identify sTyr superbinders. Previous sTyr binders (Ju et al., 2016; Lawrie et al., 2021) were based on the sequence of the SH2 superbinder SH2-trm (**Figure 1B**; Kaneko et al., 2012) that had a super affinity for pTyr due to the three critical residues (V@P6, A@P8, and L@P11) but had a weak affinity for sTyr (Ju et al., 2016; Lawrie et al., 2021). These critical residues were fixed in sTyr binders, indicating their reliance on the binding mechanism of SH2-trm. They were developed by randomizing four or five positions (P1–P5) to change the specificity of SH2-trm to sTyr (Ju et al., 2016; Lawrie et al., 2021). In contrast, the two critical residues of SH2-trm were mutated in SH2-4 (A8V and L11R), suggesting that SH2-4 has a different binding mechanism from previous sTyr binders. In addition, the residues at the randomized positions of previous sTyr binders were mutated in SH2-4 (Q2N, L3Q, A4P, and R5G), and none of them were identical to previous sTyr binders, suggesting that SH2-4 has a different specificity mechanism. In summary, SH2-4 and previous sTyr binders have different binding and specificity mechanisms.

We previously built a phage library by randomizing several positions in the pTyr-binding pocket of the SH2 domain and identified SH2 superbinders after a few rounds of selections against pTyr peptides (Kaneko et al., 2012; Li et al., 2021). As some of them were found to interact with sTyr peptides (Ju et al., 2016; Lawrie et al., 2021), we thought that the variants bound to pTyr and sTyr peptides were enriched during selections against pTyr. This thought was proven in this study (**Figure 2A**). This observation is comprehensible as sTyr and pTyr have structural similarities (**Supplementary Figure 1**). Therefore, it is expected that variants bound to pTyr and sTyr peptides were also enriched during selections against sTyr in this study (**Figure 2B**). Despite this, pTyr-specific variants are likely to be enriched in selections against pTyr, and sTyr-specific variants tend to be enriched in selections on sTyr. To find sTyr-specific variants effectively, we did not use the random phage selection technique but performed NGS for all rounds of selections, followed by a systematic comparison of the variants enriched in selections against pTyr with those enriched in the selection against sTyr. NGS data analyses showed that the variants enriched in the selected pools had different characteristics (**Figures 3A,B**), indicating the existence of variants with high specificity for sTyr. Accordingly, we prioritized the variants based on the characteristics and found that SH2-4 had a superior affinity and specificity for sTyr. Nevertheless, SH2-4 still binds pTyr with a moderate affinity. In the future, we will optimize the prediction model or take SH2-4 as a template for rebuilding a phage library to identify variants with a higher affinity and specificity for sTyr.

MATERIALS AND METHODS

Construction of the Fyn SH2 Domain Phage Display Library

Fyn SH2-58.6 was subcloned into pFN-OM6 phagemid, which was used to construct the PIII fusion protein. Fyn SH2-58.6 in pFN-OM6 was used as a template for library construction. Primer1 (AGAGGTACCTTTCTTATC**TAACCATGGTAA**CTTTCTATC CGTGATTGG) and Primer2 (AAAGGAGACCATGTCAAA **TAACCATGGTAA**ATTCGCAAACTTGACAAT) were used to construct the template "TAA CCATGG TAA" (TAA: stop codon; CCATGG: Nco1 site) *via* the Kunkel reaction (Kunkel et al., 1991). The ssDNA template was prepared for

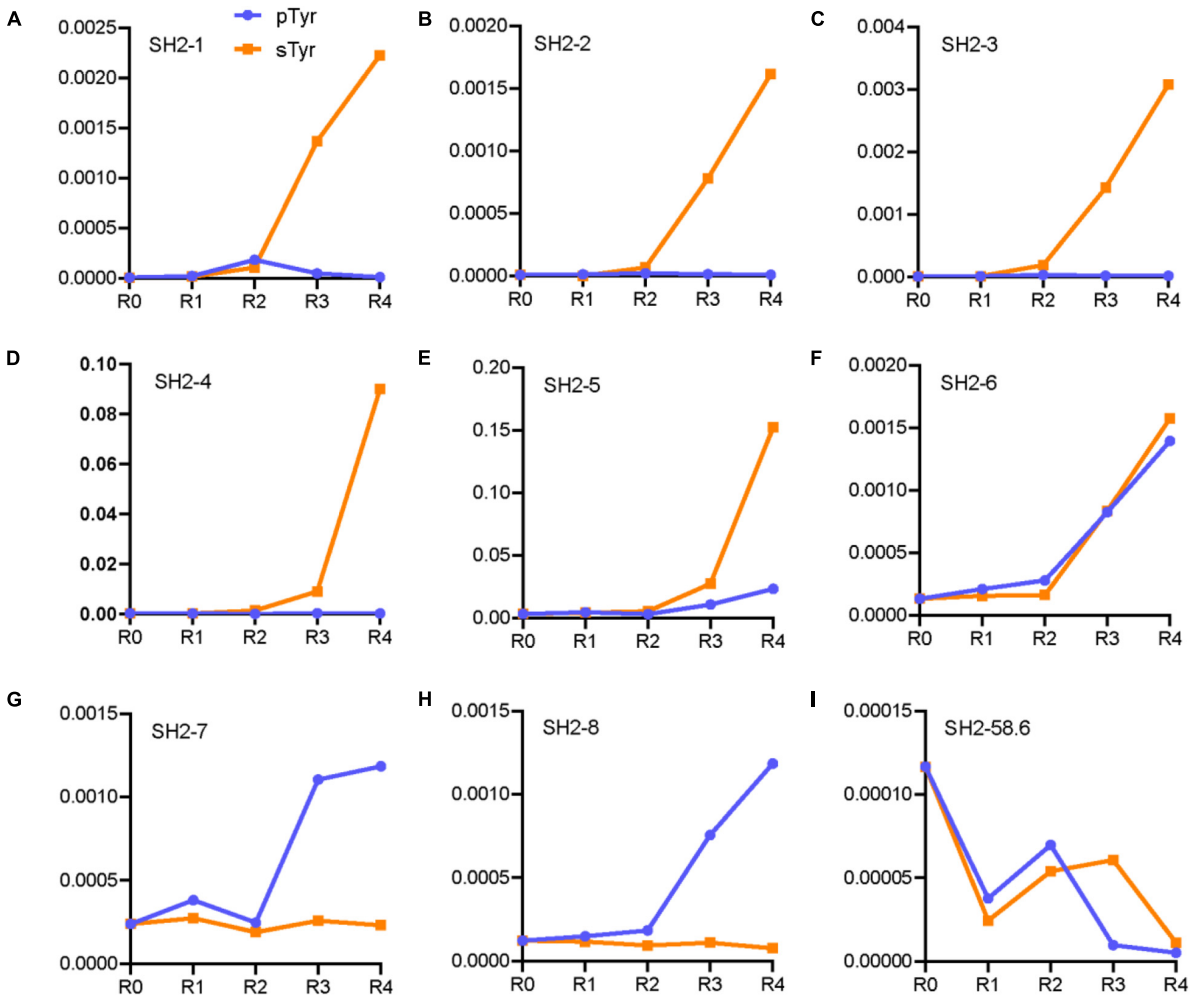


FIGURE 4 | Line chart of proportions of the variants in the rounds of selections against pTyr and sTyr peptides. These variants are SH2-1 **(A)**, SH2-2 **(B)**, SH2-3 **(C)**, SH2-4 **(D)**, SH2-5 **(E)**, SH2-6 **(F)**, SH2-7 **(G)**, SH2-8 **(H)** and SH2-58.6 **(I)**.

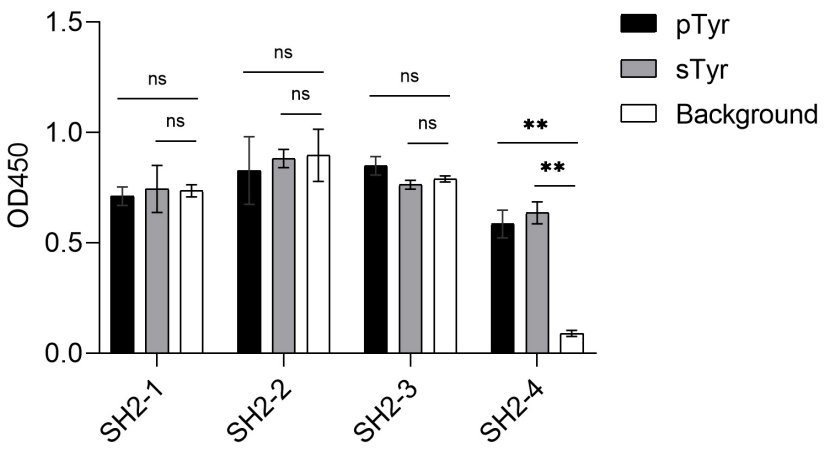


FIGURE 5 | Phage ELISA of the variants SH2-1–SH2-4 binding to pTyr and sTyr peptides and the background. Streptavidin precoated in the microplate was taken as the background. Three replicates were performed, and the *p*-value was performed with Student's *t*-test (ns, no significant; $^{**}p < 0.01$).

Kunkel reaction. The random mutation regions of Primer3 (AGAGGTACCTTTCTTATC$N_3N_4N_3N_3N_1N_1N_2N_2N_1N_4N_3N_1N_3N_4N_3N_4N_2N_1$AAAGGTGCG$N_2N_1N_2N_4N_3N_4$CTTTCTATCCGTGATTGG) and Primer4 (AAAGGAGACCATGTCAAA$N_3N_1N_2N_2N_1N_2N_3N_2N_1$ATTCGCAAACTTGACAAT) replaced the ssDNA template using the Kunkel method. In Primer3 and Primer4, N1, N2, N3, and N4 represent the mixture of A, T, C, and G with different percentages, respectively (**Supplementary Table 1**). Kunkel products (heteroduplex double strand) were purified and prepared for *Escherichia coli* SS320 (Genentech) electroporation. After being cultured overnight, dsDNA was extracted from electroporated *E. coli* SS320. After that, dsDNA was digested by Nco1 (New England Biolabs) and transferred to *E. coli* SS320 (pre-infected by M13KO7) by electroporation. The electroporated *E. coli* SS320 was cultured in 2YT (10 g yeast extract, 16 g tryptone, 5 g NaCl, water was added to make up the volume to 1.0 L, adjusted pH to 7.0 with NaOH, autoclaved) at 32°C overnight. The phage display library was precipitated with PEG/NaCl (20% PEG 8000, 2.5 M NaCl) and resuspended with polybutylene terephthalate [PBT; 0.05% Tween, 0.5% bovine serum albumin (BSA) in phosphate-buffered saline (PBS)] buffer.

Next-Generation Sequencing and Data Analysis

DNA samples from nine Fyn SH2 libraries were amplified by PCR using the forward primer (5′- GAGGTACCTTTCTTATC-3) and the reverse primer (5′-TTGTCAAGTTTGCGAAT-3′) with different labels. About > 50 ng of the purified PCR fragment was used for library preparation. These PCR products were treated with End Prep Enzyme Mix for end repair. After purification with magnetic beads, the amplified products were amplified by P5 and P7 primers with sequences that can anneal with flow cells to perform bridge PCR and index, allowing multiplexing. The library was then obtained by magnetic bead purification. The library was sequenced with a pair-end of 150 bp in the Illumina NovaSeq 6000 sequencing system. The method of data analysis was performed as described (Li et al., 2021).

This study's raw sequencing data have been deposited in the NCBI Sequence Read Archive database (Accession No. PRJNA790495).

The Fyn SH2 Domain Library Panning

Biopanning for negative selection against streptavidin (Solarbio) was performed by incubating a phage library (10^{13} CFU/ml) in a well of a streptavidin-coated 96-well plate (Thermo Fisher Scientific F96 Maxisorp NUNC-immuno plate, Lot No. 168854) for 1 h. Next, the phage library was separately transformed into the streptavidin-coated wells, with the immobilized 16 pmoL/well biotin-sulfopeptide (biotin-ahx-ahx-EPQsYEEIPIYL) and the corresponding wells with the immobilized biotin-phosphopeptide (biotin-ahx-ahx-EPQpYEEIPIYL). After 1 h of incubation, positive wells were washed 10 times using the PT buffer (PBS with 0.05% Tween). Bound phages were eluted by 100 mM HCl 100 μl/well and neutralized by adding 1/8 volume of Tris-HCl (1 M, pH = 11). The amplification of the phage library and its selection stringency were performed as described (Li et al., 2021).

Phage Enzyme-Linked Immunosorbent Assay

About 6 pmoL/well streptavidin in 50 μl PBS was coated in a 96-well NUNC microplate at 4°C overnight. The next day, the solution in the wells was discarded. Around 2% milk was added to wells for blocking. After incubation for 1 h, the wells were washed three times with the PT buffer. Peptides of 8 pmoL/well (biotin-ahx-ahx-EPQ-sTyr-EEIPIYL) in PBS were added for immobilization at room temperature for 1 h. Wells containing no peptides were used as the negative control. After washing five times with the PT buffer, phages in 50 μl PBS were incubated in wells for 1 h at room temperature. Anti-M13/HRP monoclonal antibody conjugate (50 μl, 1:8,000 dilution, Sino Biological) in PBS buffer (1:8,000 dilution, 50μl) was added to the wells and incubated for 30 min. According to the manufacturer's instructions, after washing with the PT buffer eight times, the tetramethylbenzidine (TMB) buffer (Biopanda, 50 μl/well) was added for development. Within 5 min, 50 μl/well of 1.0 M H_3PO_4 was added to stop the reaction. After that, OD450 values were measured.

Protein Expression and Purification

cDNA sequences of SH2 domain variants in pFN-OM6 were subcloned into *Sfi*I and *Not*I sites of pHH0239 to express 6xHis-tag proteins at the N-terminus. The constructed pHH0239 was transformed into *E. coli* SR320. Single colonies were added to a 2YT medium containing 50 μg/ml carbenicillin and cultured at 37°C to OD600 = 0.6, followed by adding 0.5 mM IPTG (Solarbio) and induction overnight at 18°C. After that, *E. coli* SR320 was harvested by centrifugation (16,000 g, 10 min, 4°C) and lysed by sonication. Debris was removed by centrifugation (16,000 g, 30 min, 4°C). According to the manufacturer's manual, proteins were purified by Ni-NTA agarose (Qiagen). Methods for changing elution buffer and measuring protein concentration were described elsewhere (Li et al., 2021).

Biolayer Interferometry Assay

The Octet RED96 System (ForteBio) was used for the Biolayer interferometry assay. A streptavidin biosensor (ForteBio) was used to perform the measurement. The peptides biotin-ahx-ahx-EPQ-pTyr-EEIPIYL and biotin-ahx-ahx-EPQ-sTyr-EEIPIYL were immobilized on the biosensor. The concentration (300, 150, and 0 nM) of SH2 variants was used for the measurement. All steps were described elsewhere (Li et al., 2021).

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI BioProject repository, accession number PRJNA790495.

AUTHOR CONTRIBUTIONS

LL: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, supervision, writing—original draft, and writing—review and editing. DZ: data curation, investigation, validation, and writing—original draft. HJ: data curation and methodology. YQ: investigation, validation, and writing—original draft. HH: supervision and funding acquisition. YY and CZ: investigation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by funding from the National Natural Science Foundation of China (Grant No. 32071430 to LL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.901558/full#supplementary-material

REFERENCES

Albanese, K. I., Krone, M. W., Petell, C. J., Parker, M. M., Strahl, B. D., and Brustad, E. M. (2020). Engineered Reader Proteins for Enhanced Detection of Methylated Lysine on Histones. *ACS. Chem. Biol.* 15, 103–111. doi: 10.1021/acschembio.9b00651

Bian, Y., Li, L., Dong, M., Liu, X., Kaneko, T., and Cheng, K. (2016). Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder. *Nat. Chem. Biol.* 12, 959–966. doi: 10.1038/nchembio.2178

Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., and Ponath, P. D. (1996). The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* 85, 1135–1148. doi: 10.1016/s0092-8674(00)81313-6

Dunant, N. M., Messerschmitt, A. S., and Ballmer-Hofer, K. (1997). Functional interaction between the SH2 domain of Fyn and tyrosine 324 of hamster polyomavirus middle-T antigen. *J. Virol.* 71, 199–206. doi: 10.1128/JVI.71.1.199-206.1997

Hoffhines, A. J., Damoc, E., Bridges, K. G., Leary, J. A., and Moore, K. L. (2006). Detection and purification of tyrosine-sulfated proteins using a novel anti-sulfotyrosine monoclonal antibody. *J. Biol. Chem.* 281, 37877–37887. doi: 10.1074/jbc.M609398200

Hornbeck, P. V., Kornhauser, J. M., Latham, V., Murray, B., Nandhikonda, V., and Nord, A. (2019). 15 years of PhosphoSitePlus(R): integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids. Res.* 47, D433–D441. doi: 10.1093/nar/gky1159

Huttner, W. B. (1988). Tyrosine sulfation and the secretory pathway. *Annu. Rev. Physiol.* 50, 363–376. doi: 10.1146/annurev.ph.50.030188.002051

Ju, T., Niu, W., and Guo, J. (2016). Evolution of Src Homology 2 (SH2) Domain to Recognize Sulfotyrosine. *ACS Chem. Biol.* 11, 2551–2557. doi: 10.1021/acschembio.6b00555

Kaneko, T., Huang, H., Cao, X., Li, X., Li, C., and Voss, C. (2012). Superbinder SH2 domains act as antagonists of cell signaling. *Sci. Signal.* 5:ra68. doi: 10.1126/scisignal.2003021

Kehoe, J. W., Velappan, N., Walbolt, M., Rasmussen, J., King, D., and Lou, J. (2006). Using phage display to select antibodies recognizing post-translational modifications independently of sequence context. *Mol. Cell Proteom.* 5, 2350–2363. doi: 10.1074/mcp.M600314-MCP200

Kunkel, T. A., Bebenek, K., and McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* 204, 125–139. doi: 10.1016/0076-6879(91)04008-c

Lawrie, J., Waldrop, S., Morozov, A., Niu, W., and Guo, J. (2021). Engineering of a Small Protein Scaffold To Recognize Sulfotyrosine with High Specificity. *ACS Chem. Biol.* 16, 1508–1517. doi: 10.1021/acschembio.1c00382

Leung, A. W., Backstrom, I., and Bally, M. B. (2016). Sulfonation, an underexploited area: from skeletal development to infectious diseases and cancer. *Oncotarget* 7, 55811–55827. doi: 10.18632/oncotarget.10046

Li, S., Zou, Y., Zhao, D., Yin, Y., Song, J., and He, N. (2021). Revisiting the phosphotyrosine binding pocket of Fyn SH2 domain led to the identification of novel SH2 superbinders. *Protein Sci.* 30, 558–570. doi: 10.1002/pro.4012

Liu, B. A., Jablonowski, K., Raina, M., Arcé, M., Pawson, T., and Nash, P. D. (2006). The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Mol. Cell* 22, 851–868. doi: 10.1016/j.molcel.2006.06.001

Moore, K. E., Carlson, S. M., Camp, N. D., Cheung, P., James, R. G., and Chua, K. F. (2013). A general molecular affinity strategy for global detection and proteomic analysis of lysine methylation. *Mol. Cell* 50, 444–456. doi: 10.1016/j.molcel.2013.03.005

Moore, K. L. (2003). The biology and enzymology of protein tyrosine O-sulfation. *J. Biol. Chem.* 278, 24243–24246. doi: 10.1074/jbc.r300008200

Moore, K. L. (2009). Protein tyrosine sulfation: a critical posttranslation modification in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14741–14742. doi: 10.1073/pnas.0908376106

Naider, F., and Anglister, J. (2018). The Synthesis of Sulfated CCR5 Peptide Surrogates and their Use to Study Receptor-Ligand Interactions. *Protein Pept. Lett.* 25, 1124–1136. doi: 10.2174/0929866525666181101103834

Conflict of Interest: HH and YY were the employees of Noventi Biopharmaceuticals Co., Ltd. CZ was the employee of Sino Genomics Technology Co., Ltd. LL, DZ, and CZ filed a provisional Chinese patent application related to this work. DZ was an intern of Noventi Biopharmaceuticals.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 July 2022
DOI 10.3389/fmicb.2022.928774

OPEN ACCESS

EDITED BY
Qi Zhao,
University of Science and Technology Liaoning, China

REVIEWED BY
Marilena Hall,
Stonehill College, United States
Puey Ounjai,
Mahidol University, Thailand

*CORRESPONDENCE
Heng Chen
hchen13@gzu.edu.cn
Jian Huang
hj@uestc.edu.cn

SPECIALTY SECTION
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

RECEIVED 26 April 2022
ACCEPTED 27 June 2022
PUBLISHED 15 July 2022

CITATION
He B, Li B, Chen X, Zhang Q, Lu C, Yang S, Long J, Ning L, Chen H and Huang J (2022) PDL1Binder: Identifying programmed cell death ligand 1 binding peptides by incorporating next-generation phage display data and different peptide descriptors.
Front. Microbiol. 13:928774.
doi: 10.3389/fmicb.2022.928774

PDL1Binder: Identifying programmed cell death ligand 1 binding peptides by incorporating next-generation phage display data and different peptide descriptors

Bifang He[1], Bowen Li[1], Xue Chen[1], Qianyue Zhang[1], Chunying Lu[1], Shanshan Yang[1], Jinjin Long[1], Lin Ning[2], Heng Chen[1]* and Jian Huang[3]*

[1]Medical College, Guizhou University, Guiyang, China, [2]School of Healthcare Technology, Chengdu Neusoft University, Chengdu, China, [3]School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, China

Monoclonal antibody drugs targeting the PD-1/PD-L1 pathway have showed efficacy in the treatment of cancer patients, however, they have many intrinsic limitations and inevitable drawbacks. Peptide inhibitors as alternatives might compensate for the drawbacks of current PD-1/PD-L1 interaction blockers. Identifying PD-L1 binding peptides by random peptide library screening is a time-consuming and labor-intensive process. Machine learning-based computational models enable rapid discovery of peptide candidates targeting the PD-1/PD-L1 pathway. In this study, we first employed next-generation phage display (NGPD) biopanning to isolate PD-L1 binding peptides. Different peptide descriptors and feature selection methods as well as diverse machine learning methods were then incorporated to implement predictive models of PD-L1 binding. Finally, we proposed PDL1Binder, an ensemble computational model for efficiently obtaining PD-L1 binding peptides. Our results suggest that predictive models of PD-L1 binding can be learned from deep sequencing data and provide a new path to discover PD-L1 binding peptides. A web server was implemented for PDL1Binder, which is freely available at http://i.uestc.edu.cn/pdl1binder/cgi-bin/PDL1Binder.pl.

KEYWORDS

PD-1/PD-L1 pathway, PD-L1 binding peptides, machine learning, next-generation phage display (NGPD) biopanning, support vector machine (SVM)

Introduction

Blocking the immune checkpoint pathway is a highly promising therapeutic modality to fight cancer. Programmed cell death protein 1 (PD-1) is an immune checkpoint protein, which is mainly up-regulated on activated T cells, natural killer cells and B cells (Freeman et al., 2000). Programmed cell death ligand 1 (PD-L1) is

a ligand for PD-1, which is highly expressed on many different malignancy cells and antigen-presenting cells (APCs) (Talantova et al., 2013). The interaction between PD-1 on T cells and PD-L1 on tumor cells leads to the inhibition of T-cell responses and loss of the cytotoxic T-cells' functions and thereby mediates tumor cells to escape from the host immune surveillance. Blockade of this pathway can activate tumor-infiltrating T cells and restore their anti-tumor activity (Tumeh et al., 2014; Wolchok and Chan, 2014). Therefore, PD-1 and PD-L1 have become attractive therapeutic targets against cancer. Neoadjuvant anti-PD-1/PD-L1 therapy also achieved satisfactory clinical results in tumors (Li et al., 2021b).

Six PD-1/PD-L1 monoclonal antibody (mAb) blockers to date have been approved by FDA for cancer treatment (Postow et al., 2015; Robert et al., 2015; Bang et al., 2018; Tang et al., 2018). Moreover, most of PD-1/PD-L1 inhibitors in the clinical development are mAbs (Lin et al., 2020). Although mAbs targeting either PD-1 or PD-L1 have showed certain anti-tumor efficacy in cancer patients (Hamanishi et al., 2016), current mAb agents have many intrinsic limitations such as expensive production, still poor therapeutic responses (only approximately 20% of patients with a durable response) (Yang et al., 2016), and considerable individual differences as well as immunotherapy-induced improper immune-related responses (Fishman et al., 2019). Additionally, mAb therapeutics are accompanied by inevitable drawbacks including inferior organ or tumor penetration, poor oral bioavailability and immunogenicity. Compared to mAbs, peptides as drug candidates have several advantages, including higher tissue or tumor penetration, lower production costs and decreased immunogenicity. Peptides can also be subjected to chemical modifications to improve their pharmaceutical properties. However, PD-L1 binding peptides discovery through random peptide library screening is time consuming, expensive, and laborious.

In order to improve the efficiency of phage display selection, researchers have employed computational methods to aid analysis of results of random peptide library screening. For example, SAROTUP integrates a suite of tools which can be used to scan, report and exclude possible target-unrelated peptides from phage display biopanning results (He et al., 2019b). Sun et al. (2016) have proposed an epitope prediction method based on random peptide library screening. Machine learning methods have been used in mining and designing peptides of specific function (Tallorin et al., 2018; Ma et al., 2022). Obtaining therapeutic molecules is cheap and fast with the help of machine learning approaches (Liu et al., 2020; Laustsen et al., 2021). However, there are currently no bioinformatics tools to identify PD-L1 binding peptides.

Phage display permits high-throughput screening of peptide ligands with high affinity and specificity for almost any target of interest through several rounds of target-binding (selection) and amplification of phage display peptide libraries (Jaroszewicz et al., 2022; Ledsgaard et al., 2022). Moreover, phage display coupled with next-generation sequencing (NGPD) offers a more powerful tool to identify peptide ligands (Matochko and Derda, 2015; He et al., 2016, 2018a, 2019c; Asar et al., 2020; Pleiko et al., 2021). Fewer biopanning rounds powered by deep sequencing can discover robust target-binding peptides that are not identified by Sanger sequencing (Juds et al., 2020). In addition, NGPD has been revealed very effective to suppress false-positive hits from amplification-induced bias (Matochko et al., 2014). Many researchers have employed traditional phage display technology to identify PD-L1 binding peptides (Li et al., 2018, 2021a; Liu et al., 2019; Tooyserkani et al., 2021), however, current existing PD-L1 binding peptides are not enough to implement a computational model for identifying PD-L1 binding peptides. NGPD can help to discover more novel PD-L1 binding peptide ligands. Illumina sequencing is a massively parallel sequencing technology and can produce large amounts of data (Quail et al., 2008). We screened the Ph.D.-12 phage display library against PD-L1, and here the selection output was investigated using Illumina sequencing.

In the present study, we aimed to develop a novel computational classifier for identifying peptides targeting PD-L1. We took advantage of NGPD to isolate PD-L1 binding peptides and used them to construct the predictive model *via* machine learning methods. First, we used PD-L1 as bait to screen the Ph.D.-12 phage display library. Second, the PD-L1 binding peptides isolated by phage display selection were paired with non-PD-L1 binding peptides. They were used to implement machine learning based models for predicting PD-L1 binding peptides. Third, we utilized two independent testing datasets to evaluate the generalization ability of the models. The PD-L1 binding peptides identified in this work hold high potential to be developed as anti-tumor therapeutics. The predictor for identifying peptides targeting PD-L1, called PDL1Binder, is valuable in accelerating PD-L1 binding peptides discovery and freely available at http://i.uestc.edu.cn/pdl1binder/cgi-bin/PDL1Binder.pl. Our study demonstrates that predictive models of PD-L1 binding can be learned from deep sequencing data and provides an efficient approach to discover PD-L1 binding peptides.

Dataset and methods

Phage display peptide library biopanning

We performed two rounds of phage display selection using recombinant human PD-L1 extracellular domain (ECD) protein (Cat# 10084-H08H, Sino Biological Inc., Beijing, China) as bait. The selection of Ph.D.-12 phage display library (New England Biolabs, Ipswitch, MA, United States) against PD-L1 was performed in six replicates. The control selections, i.e., Ph.D.-12 against Dynabeads (Cat# 10-103-D, Invitrogen) and

Ph.D.-12 against unrelated anti-FLAG M2 monoclonal antibody (Cat# F3165, Sigma-Aldrich), were performed in triplicate.

Round 1

In a microcentrifuge tube, 20 μL of Dynabeads were coated with a solution of PD-L1 (100 μL, 100 μg/mL) in PBS for overnight at 4 C. The solution was added to 900 μL of PBS and then transferred to a well in the KingFisher 96 deep-well plate (Cat# 95040450, Thermo Fisher Scientific). The Dynabeads with PD-L1 were rinsed 3 times with 0.1% Tween-20 in PBS, and then blocked with 2% (w/v) BSA in PBS for 1 h at room temperature, followed by an incubation with 3×10^{11} PFU Ph.D.-12 phage display library for 1.5 h at room temperature. The unbound phage was rinsed with 0.1% Tween-20 in PBS. Phage remained on the beads were eluted for 9 min at room temperature by adding 20 μL of HCl (pH 2). The elution buffer along with the beads were transferred into a 1.5 mL microcentrifuge tube and immediately neutralized with 10 μL of neutralization buffer (Phusion HF Buffer, NEB B0518S). The recovered phage was amplified in *E. coli* ER2738 (New England Biolabs, Ipswitch, MA, United States) for the second round of biopanning.

Round 2

Six microcentrifuge tubes containing 20 μL of Dynabeads were coated with a solution of PD-L1 (100 μL, 100 μg/mL) in PBS for overnight at 4 C. An additional three microcentrifuge tubes containing 20 μL of Dynabeads were coated with a solution of Protein G (100 μg/mL) along with anti-FLAG M2 monoclonal antibody (150 μg/mL) in 100 μL PBS for overnight at 4 C. In parallel, three more microcentrifuge tubes containing 20 μL of Dynabeads were suspended in 100 μL PBS for overnight at 4 C. Solution from all 12 microcentrifuge tubes were then, respectively, added to 900 μL of PBS and transferred to 12 wells in a KingFisher 96 deep-well plate. The Dynabeads were rinsed with 0.1% Tween-20 in PBS, and then blocked with 2% (w/v) BSA in PBS for 1 h at room temperature, followed by incubation with 3×10^{10} PFU enriched Ph.D.-12 phage display library from Round 1 for 1.5 h at room temperature. The unbound phage was rinsed with 0.1% Tween-20 in PBS five times. Phage remained on the beads were resuspended in DNase free water and boiled at 90 C for 10 min. The single-stranded DNAs (ssDNAs) from discovered phage were extracted and subjected to polymerase chain reaction (PCR) amplification and Illumina sequencing, and those from Ph.D.-12 libraries before and after the first round of biopanning were also sequenced to serve as additional controls. The steps for Illumina sequencing of phage display libraries were described previously (He et al., 2018b). Briefly, PCR amplification was first performed to transform the ssDNA of the amplified phage into Illumina-compatible double-stranded DNA (dsDNA). The detailed PCR protocol for Illumina sequencing can be found in the **Supplementary Information**. After PCR amplification, the dsDNA PCR fragments corresponding to the expected size were confirmed and quantified using agarose gel electrophoresis. The PCR products from multiple experiments were then mixed together allowing 20 ng of each product in the mixture and purified by E-Gel (Thermo Fisher Scientific, Waltham, MA, United States). The purified dsDNAs were finally sequenced using the Illumina NextSeq paired-end 500/550 High Output Kit v2 (150 cycles).

Deep-sequencing analysis

Raw FASTQ data were processed by using MatLab scripts described in a previous publication (He et al., 2018b) and filtered to find significantly enriched sequences using MatLab scripts previously reported on a computational server (Sugon I840-G20, Dawning Information Industry Co., LTD., Beijing, China). Sequences isolated from the PD-L1 screen that increased significantly in abundance against sequences isolated from the control selections were labeled PD-L1 binding peptides. Significance of the ratio was assessed using one-tailed, unequal variance Student *t*-test. Only sequences with ratio ≥ 2 and p-value ≤ 0.05 were considered as PD-L1 binding peptides. Deep sequencing the library before round 1 (R0), the output of two selection rounds (R1 and R2) and the control selection experiments identified 80 peptide sequences that exhibited high normalized abundance in R2 and low normalized abundance in R0, R1, and the control experiments R2-DB (Dynabeads), and R2-UF (unrelated anti-FLAG M2 monoclonal antibody).

Database search for target-unrelated peptides

All sequences that were identified as potential PD-L1 binding peptides were searched against the BDB database[1] (He et al., 2016) to check if they have been previously discovered in other phage display screens with distinct targets (MimoSearch[2] and MimoScan[3]). Peptides that were identified by four or more entirely different targets were putative target-unrelated peptides (He et al., 2019a).

Benchmark dataset for training

The positive dataset was composed of 80 PD-L1 binding peptides identified by NGPD. The remaining peptides were

1 http://i.uestc.edu.cn/bdb/
2 http://i.uestc.edu.cn/bdb/index.php/site/tools?type=MimoSearch
3 http://i.uestc.edu.cn/bdb/index.php/site/tools?type=MimoScan

non-PD-L1 binding peptides, which consisted of the negative dataset. Redundant peptides were then removed by using CD-HIT (Li and Godzik, 2006; Fu et al., 2012), with a sequence identity threshold of 0.8 for the PD-L1 and non-PD-L1 binding peptides, respectively. We used this value based on the size of our dataset. More stringent criteria, such as 0.4 or 0.3, were not adopted because machine learning algorithms could not acquire abundant information to learn with a relatively small sample. After this analysis, no redundant peptides were found and excluded, and the positive training dataset consisted of 80 PD-L1 peptides. To balance the positive and negative training dataset, we randomly selected 800 peptides from the negative dataset and divided them into 10 sub-datasets. Each negative sub-dataset was paired with the positive training dataset. Finally, 10 pairs of sub-datasets were constructed and each pair was composed of 80 PD-L1 and 80 non-PD-L1 binding peptides (**Table 1**). The training dataset is provided in Trainingdataset.xslx in the **Supplementary Material**.

Independent testing dataset construction

The literature data related to PD-L1 binding peptides were extracted from the PubMed database. A typical text mining query is given below: (anti-PD-L1 peptide) OR (PD-L1 binding peptide). The search returned 652 articles published before July 08, 2021. PD-L1 binding peptide sequences were then manually extracted from the above peer-reviewed papers. Modified peptides (peptides with non-natural amino acids) were first excluded since no modified peptides were in the training dataset. Inclusion criteria were as follows: (1) peptides containing only 20 natural amino acids and less than 50 residues were collected since peptides having more than 50 amino acids were considered as proteins; (2) peptides that have been experimentally verified to bind with PD-L1 *in vitro* or *in vivo* were collected. Finally, 34 experimentally validated PD-L1 binding peptides were obtained. After removing redundant peptides by using CD-HIT with a sequence identity cutoff of 0.8, 30 PD-L1 binding peptides were retained. Consequently, we constructed two independent testing datasets, i.e., TestDataset_1 (30 PD-L1 binding peptides) and TestDataset_2 [221405 non-PD-L1 binding peptides from the remaining negative dataset (not for training)] (**Table 1**). As the sources of the two datasets are different, they were tested separately. Datasets TestDataset_1 and TestDataset_2 are provided in Testingdataset.xlsx in the **Supplementary Material**.

TABLE 1 Number of PD-L1 and non-PD-L1 binding peptides in each dataset.

Dataset	Number of PD-L1 binding peptides	Number of non-PD-L1 binding peptides
Training dataset	80	80/80/80/80/80/80/80/80/80/80
TestDataset_1	30	/
TestDataset_2	/	221405

For the training dataset, each negative sub-dataset with 80 non-PD-L1 binding peptides was paired with the positive training dataset composed of 80 PD-L1 binding peptides.

Sequence encoding and peptide descriptor analysis

Four peptide descriptors, including the amino acid composition (AAC), pseudo amino acid composition (PseAAC), dipeptide composition (DPC) and the composition of *k*-spaced amino acid group pairs (CKSAAGP), were used to encode each peptide in the training dataset. The calculation of the above descriptors were performed by the codes within the iLearnPlus (Chen et al., 2021). AAC and DPC are defined as follows:

$$AAC\,(i) = \frac{x(i)}{\sum_{i=1}^{20} x(i)} \quad (1)$$

$$DPC\,(j) = \frac{y(j)}{\sum_{j=1}^{400} y(j)} \quad (2)$$

where *AAC(i)* is the percent of the *ith* ($i = 1, 2, \ldots, 20$) amino acid, and *x(i)* represents the number of the *ith* amino acid in a peptide sequence. *DPC(j)* is the frequency of the *jth* ($j = 1, 2, \ldots, 400$) dipeptide, and *y(j)* represents the number of the *jth* dipeptide in a peptide sequence.

The CKSAAGP descriptor is modified from the composition of *k*-spaced amino acid pairs (CKSAAP) in which the occurrences of the amino acid pairs that are separated by *k*-residues are calculated. For CKSAAGP, 20 amino acids are first divided into five groups according to their physicochemical properties: aromatic, aliphatic, negative-charged, positive-charged, and uncharged residues. The frequencies of the 25 amino acid pairs (5×5) separated by *k*-residues with group annotations were then calculated. For a peptide with *L* residues, if the *k*-spaced residue group pair *AE* appears *n* times, the frequency of the corresponding residue pair is $n/[L-(k+1)]$. In this study, $k = 0$, 1 and 2 were jointly considered due to the peptide length of 12. Finally, the CKSAAGP descriptor with $25 \times 3 = 75$ dimensions was comprised of the frequencies of 0-spaced, 1-spaced, 2-spaced residue group pairs.

The sequence-order information would be completely ignored if AAC is used to encode a sequence. To compensate for AAC, PseAAC was proposed by introducing discrete factors for incorporating some sort of sequence-order or pattern information (Chou, 2001). The detailed calculation of PseAAC can be found at (Chou, 2009). In the formula for PseAAC, the weight factor ω and discrete counted-rank correlation factor λ

are two key parameters. Considering the limited sequence length and to ensure the diversity of key components, we set $\lambda = 4$ and $\omega = 0.4$ to generate PseAAC with $20 + \lambda$ dimensions.

Feature selection

In this study, feature selection was implemented by using the iLearnPlus platform (Chen et al., 2021). Chi-square test (CHI2) (Forman, 2003), Information gain (IG) (Yu and Liu, 2003), F-score value (FScore) (Chen et al., 2018), Mutual information (MIC) (Peng et al., 2005), and Pearson's correlation coefficient (Pearson) (Stigler, 1989) feature selection strategies were used to identify key features. The selected feature number was set to be 160 as each sub-dataset was comprised of 160 peptide sequences. The MinMax normalization approach was then utilized to scale the selected features to the unit range between 0 and 1. To select the optimal feature set, we further used various machine learning methods to construct models with each of the feature sets selected by CHI2, IG, FScore, MIC and Pearson feature selection approaches, respectively, *via* fivefold cross-validation. The feature set, which achieved the best classification performance, was utilized for further model construction.

Machine learning algorithm selection

The optimal feature set obtained by feature selection was used to construct classifiers based on 12 state-of-the-art machine learning algorithms in the iLearnPlus-AutoML module *via* fivefold cross-validation (select the "Auto optimization" option to optimize parameters automatically), including Support vector machine (SVM) (Cortes and Vapnik, 1995), Random forest (RF) (Breiman, 2001), Decision tree (DecisionTree) (Breimann et al., 1984), K-nearest neighbors (KNN) (Altman, 1992), Logistic regression (LR) (Freedman, 2009), Gradient boosting decision tree (GBDT) (Friedman, 2001), Light gradient boosting machine (LightBGM) (Ke et al., 2017), Extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016), Stochastic gradient descent (SGD) (Pedregosa et al., 2011), Naïve Bayes (NaïveBayes) (Rennie et al., 2003), Linear discriminant analysis (LDA) (McLachlan, 1992), and Quadratic discriminant analysis (QDA) (McLachlan, 1992).

Performance evaluation

The fivefold cross-validation test was selected to evaluate the performance of the constructed classifiers. In the fivefold cross-validation test, the sequence dataset is randomly divided into five equally sized folds. Four folds of these folds are used to develop the machine learning model and optimize its parameters, and the remaining one fold is employed to assess the performance of the model. The process was repeated five times until each fold is used for testing once. In this study, eight commonly used metrics were utilized to quantify the model predictive performance, including sensitivity (Sn), specificity (Sp), Precision (Pr), F1 score (F1), accuracy (Acc), Matthews correlation coefficient (MCC), the area under the receiver operating characteristic (ROC) curve (AUROC) and the area under the precision-recall curve (AUPRC). The former six performance indicators are calculated by the following equations:

$$Sn = \frac{TP}{TP + FN} \tag{3}$$

$$Sp = \frac{TN}{TN + FP} \tag{4}$$

$$Pr = \frac{TP}{TP + FP} \tag{5}$$

$$F1 = \frac{2 \times TP \times TP}{TP \times (TP + FN) + TP \times (TP + FP)} \tag{6}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{8}$$

where *TP*, *FP*, *TN*, *FN*, respectively, are the number of true positives, the number of the false positives, the number of true negatives, the number of the false negatives. We also computed the AUROC and AUPRC values for comparing the model performance.

Final model construction and web service implementation

Ten submodels were constructed based on SVM by using the LIBSVM 3.25 package (Chang and Lin, 2011), which is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/. The radial basis function (RBF) was selected as the kernel function to develop SVM-based models. The kernel width factor *gamma* and the regularization factor *c* were automatically optimized by selecting the "Auto optimization" option *via* the grid search method in iLearnPlus (Chen et al., 2021). To reduce the generalization error of the prediction, we adopted the voting strategy to implement an ensemble predictor, called PDL1Binder. The ensemble model aggregates the predictive result of each submodel. In this study, we used the averaging voting technique, which takes an average of predictions from ten submodels and uses it to make the final prediction. Each peptide for prediction will be subjected to the prediction of ten submodels. Each submodel corresponds to a possibility value of the peptide being a PD-L1 binding

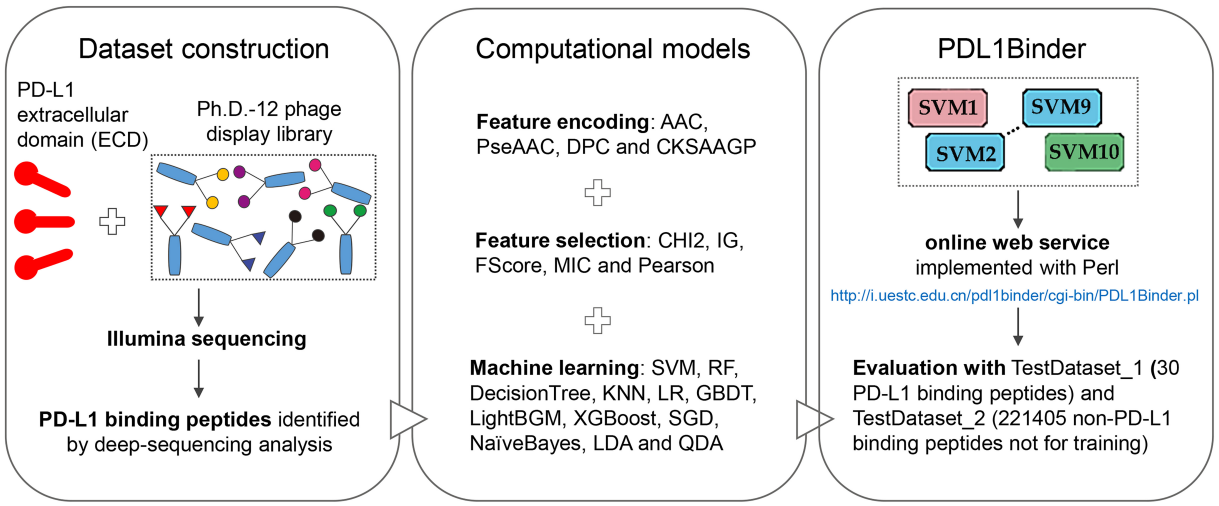


FIGURE 1
Overview of this study.

peptide. The final probability value was computed by averaging the probability values of ten submodels. If the value is greater than or equal to the threshold of possibility value (0.5 by default), the peptide will be identified as a PD-L1 binding peptide.

For ease of use, the PDL1Binder classifier was further implemented into an online web service, which is freely available at http://i.uestc.edu.cn/pdl1binder/cgi-bin/PDL1Binder.pl. The web interface of PDL1Binder was developed by using Perl. The web service was tested in the Mozilla Firefox, Google Chrome, and Internet Explorer browsers.

Results

The workflow of this study is shown in Figure 1. We first isolated 80 PD-L1 binding peptides by using NGPD and utilized them as the benchmark dataset to develop computational models for identifying PD-L1 binding peptides. Four different peptide descriptors were employed to encode each peptide sequence. The optimal feature selection approach chosen from five feature selection strategies and 12 machine learning methods were combined to implement predictive models. Fivefold cross-validation results showed that the SVM-based model outperformed models developed with 11 other machine learning algorithms. Therefore, an ensemble SVM-based computational model, called PDL1Binder, was implemented. Moreover, two independent testing dataset: TestDataset_1 (30 PD-L1 binding peptides) and TestDataset_2 (221405 non-PD-L1 binding peptides not for training) were used to evaluate PDL1Binder.

Selection and analysis of peptides that bind to programmed cell death ligand 1

We used the Ph.D.-12 phage display library to discover peptide ligands for PD-L1. Two rounds of phage display selection were performed using PD-L1 ECD as bait. In round 2, we also performed two control selections; in the first control, we panned the enriched Ph.D.-12 library from Round 1 against the Dynabeads (R2-DB) and in the second control, we panned the enriched Ph.D.-12 library from Round 1 against unrelated anti-FLAG M2 monoclonal antibody (R2-UF). Deep sequencing the library before round 1 (R0), the output of two selection rounds (R1 and R2) and the control selection experiments identified 80 peptide sequences that exhibited high normalized abundance in R2 and low normalized abundance in R0, R1, and the control experiments R2-DB, and R2-UF.

All 80 potential peptide binders for PD-L1 were significantly enriched ($p < 0.05$, $R \geq 2$) in the selection of the Ph.D.-12 phage display library on PD-L1 but not in any of the control screens (Supplementary Figure 1). We clustered the hit sequences based on their features described by the BLOSUM62 matrix and found that 29 peptides were clustered into five groups (Figure 2). The remaining un-clustered sequences were assigned to their nearest clusters (Supplementary Figure 1).

Finally, we investigated whether 80 PD-L1 binders could be target-unrelated peptides which are enriched for other reasons other than target specificity. MimoSearch and MimoScan confirmed that YPGSQSWMPSDF has been previously selected by IgE from patients, while the remaining 79 peptides have not been identified in other phage display biopanning datasets which are curated in BDB. As YPGSQSWMPSDF was only identified with two different targets so far, it was not considered as a target-unrelated peptide and remained for further analysis.

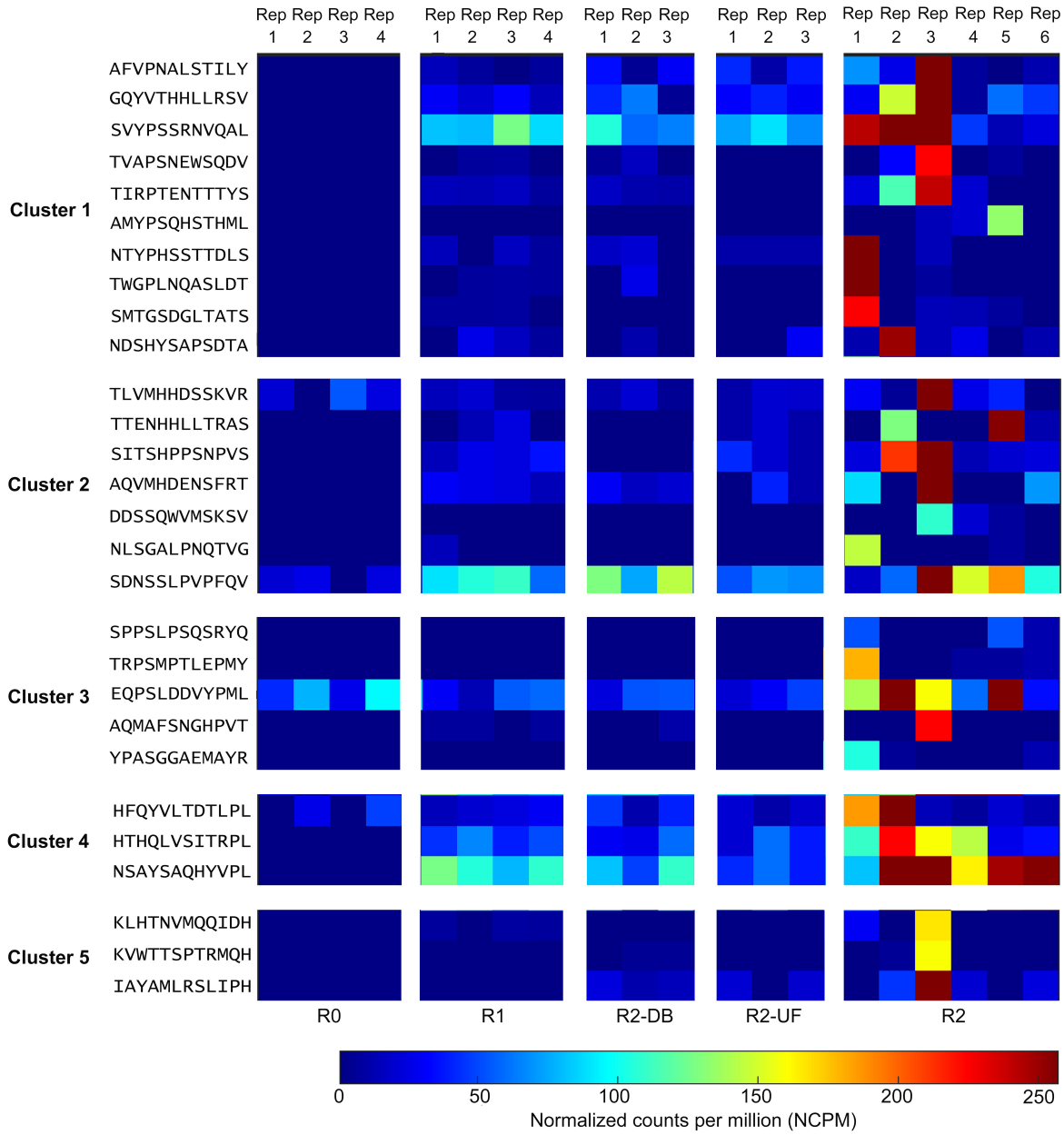


FIGURE 2
Deep sequencing the output of all selection rounds and the control experiments identified peptide sequences that exhibited high normalized abundance in R2 and low normalized abundance in R0, R1, and the control screens R2-DB, and R2-UF. Twenty-nine sequences from the deep sequencing results were clustered into five groups. Rep, replicate; R0, the library before round 1; R1, the first round of panning against PD-L1 ECD; R2-DB, panning the enriched Ph.D.-12 library from R1 against the Dynabeads; R2-UF, panning the enriched Ph.D.-12 library from R1 against unrelated anti-FLAG M2 monoclonal antibody (R2-UF); R2, panning the enriched Ph.D.-12 library from R1 against PD-L1 ECD.

Performance analysis of models trained with diverse machine learning and feature selection methods

The AAC, DPC, CKSAAGP, and PseAAC descriptors were used to encode each peptide in the training dataset. We directly concatenated four types of peptide descriptors. As a result, the dimension of the feature vector of each peptide is 519 (Table 2). For each of the ten sub-datasets, feature selection was performed by using five popular feature selection approaches, respectively. The number of selected features was determined to be 160 to keep the same as the number of peptides in the training dataset.

The feature subsets obtained through various feature selection methods were then used to develop predictors

TABLE 2 List of 519 features.

Peptide descriptor	Feature dimension
Amino acid composition (AAC)	20
Dipeptide composition (DPC)	400
Pseudo amino acid composition (PseAAC)	24
Composition of *k*-spaced amino acid group pairs (CKSAAGP)	75
(AAC, DPC, PseAAC, CKSAAGP)	519

Each peptide was represented by four types of peptide descriptors, which were conflated into a feature vector with 519 dimensions.

with 12 different traditional machine learning methods implemented in iLearnPlus (Chen et al., 2021). As shown in **Figure 3**, the results of fivefold cross-validation showed that the SVM-based classifier trained with the feature set selected by Pearson's correlation coefficient achieved an average accuracy of 82.13% with an average of 86.13% sensitivity and 78.13% specificity. For all ten submodels, the AUROC and AUPRC values of the SVM-based model are the highest. The model under this combination outperformed other models developed with different feature subsets and machine learning algorithms (see Machinelearningresult.xlsx in **Supplementary Material**). Therefore, feature subsets selected by Pearson's correlation coefficient and SVM were utilized for further model construction. Performance metrics of each submodel under each combination were provided in Machinelearningresult.xlsx in **Supplementary Material**.

Ensemble predictor for identifying programmed cell death ligand 1 binding peptides

Based on the above results, we proposed an ensemble SVM-based predictor for identifying PD-L1 binding peptides, called PDL1Binder, the framework of which is illustrated in **Figure 4**. For a given peptide, it will be predicted by ten submodels separately. PDL1Binder then uses the averaging voting method and makes final predictions based on the average probability value. The fivefold cross-validation results in **Figure 3** showed that PDL1Binder achieved an average accuracy of 82.13% with an average of 86.13% sensitivity, 80.42% precision, and 78.13% specificity, and an average of 0.6528 MCC, 0.8271 F1, 0.8978 AUROC, and 0.8989 AUPRC.

For the convenience of users in using PDL1Binder, an online web service has been developed, which is freely available at http://i.uestc.edu.cn/pdl1binder/cgi-bin/PDL1Binder.pl. As shown in **Figure 5**, a professional and user-friendly web architecture for PDL1Binder was implemented. PDL1Binder allows users to submit peptide sequences in fasta or plain text format and set the threshold of probability value to differentiate

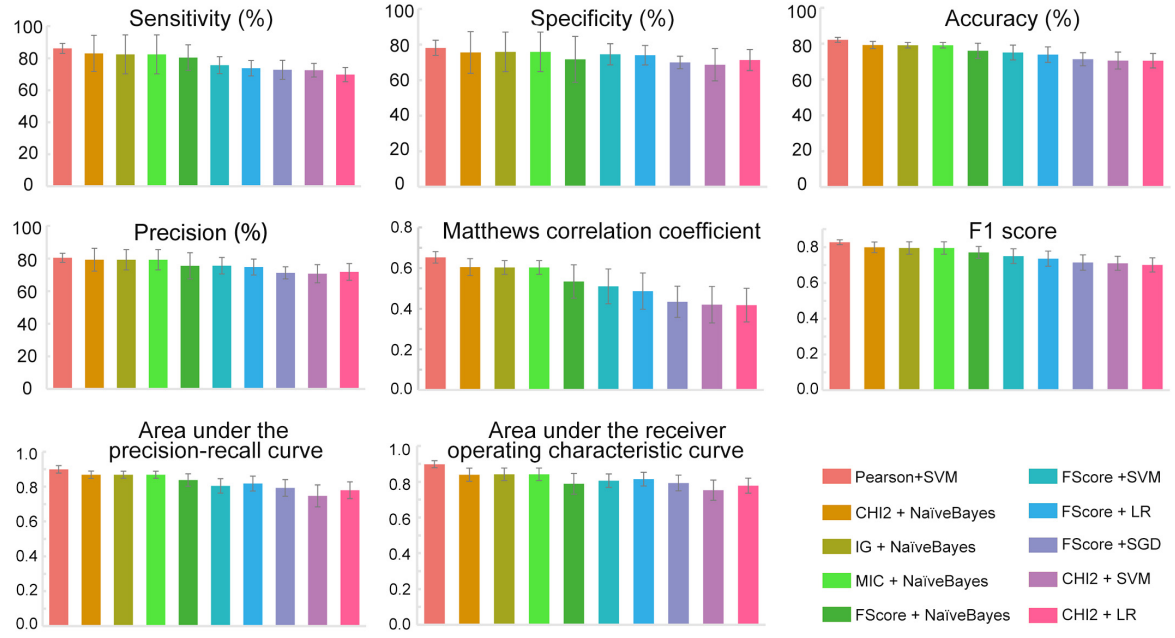


FIGURE 3
The performance metrics of each submodel. All data were expressed as mean ± standard deviation. SVM, Support vector machine; LR, Logistic regression; SGD, Stochastic gradient descent; NaïveBayes, Naïve Bayes; Pearson, Pearson's correlation coefficient; CHI2, Chi-square test; IG, Information gain; FScore, F-score value; MIC, Mutual information.

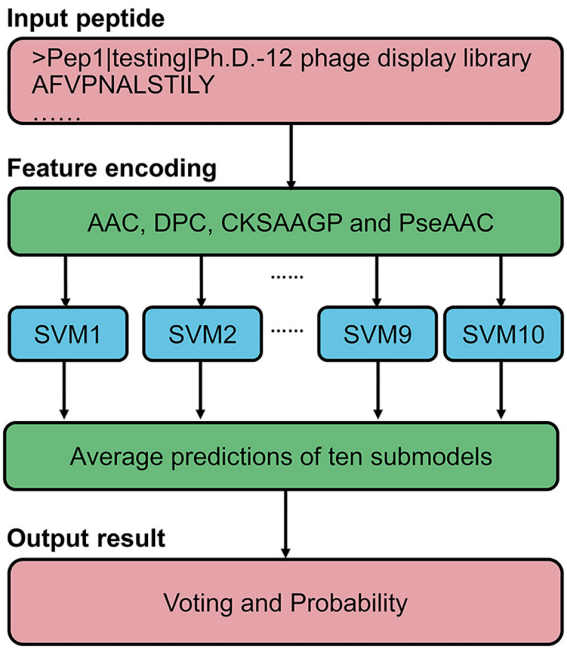


FIGURE 4
Framework of the proposed scheme for PD-L1 binding peptide prediction.

between predicted positives and negatives (*tp*) (**Figure 5A**), which makes it more convenient and flexible for future users. To simplify the representation of PDL1Binder prediction, predictive results are displayed in a table (**Figure 5B**). Users can sort the results in ascending or descending order by a specific column.

Evaluation of PDL1Binder with independent testing datasets

Two independent testing datasets, one with 30 non-redundant PD-L1 binding peptides and the other one with 221405 non-redundant non-PD-L1 binding peptides, were employed to evaluate the generalization ability of PDL1Binder under different *tp*-values. As shown in **Table 3**, with the increase of the *tp*-value, the value of sensitivity decreases, while the specificity value increases. When the *tp*-value was set to 0.55 within PDL1Binder, 83.33% PD-L1 binding peptides in the TestDataset_1 were correctly identified as PD-L1 binding peptides, while 53.29% non-PD-L1 binding peptides were precisely predicted as non-PD-L1 binding peptides in the TestDataset_2 (**Table 3**).

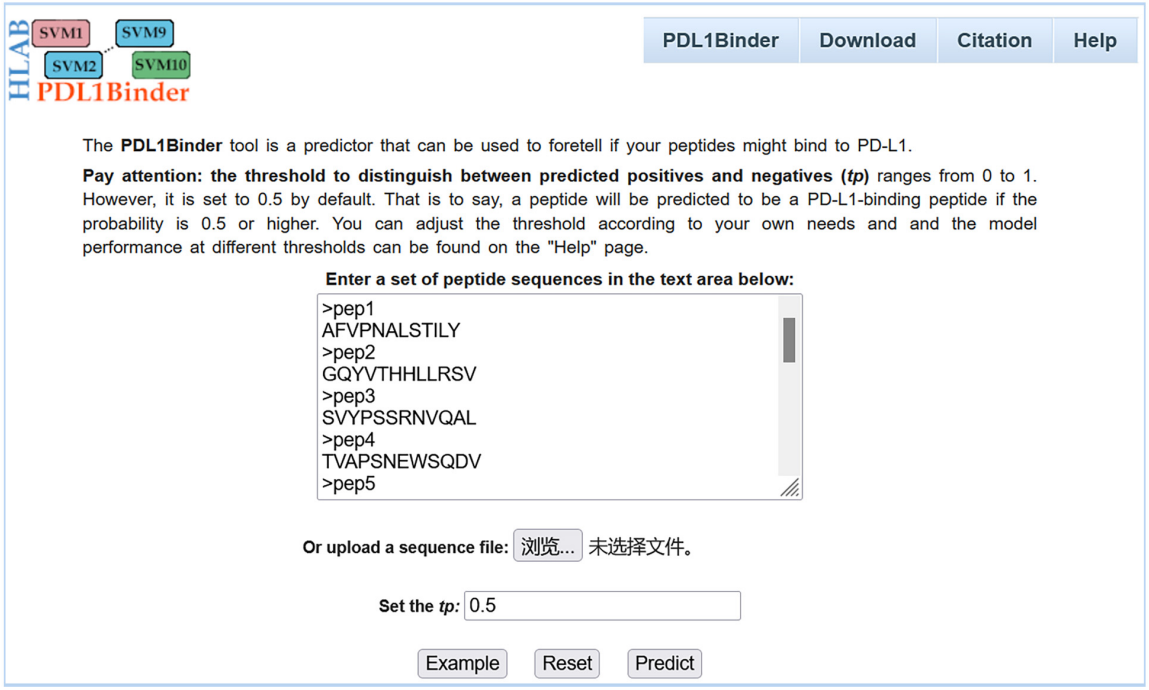
Discussion

Many studies have demonstrated that PD-L1 binding peptides are promising for the treatment of cancers (Pan et al., 2021). PD-L1 binding peptides screened by phage display selection in this study could serve as peptide drug candidates for blocking the PD-1/PD-L1 interaction. However, only a few tens of PD-L1 binding peptides have been experimentally identified. In fact, many candidate molecules are needed to develop a peptide drug for cancer immunotherapy. Therefore, it is urgently needed to employ computational methods to rapidly identify more novel PD-L1 binding peptides.

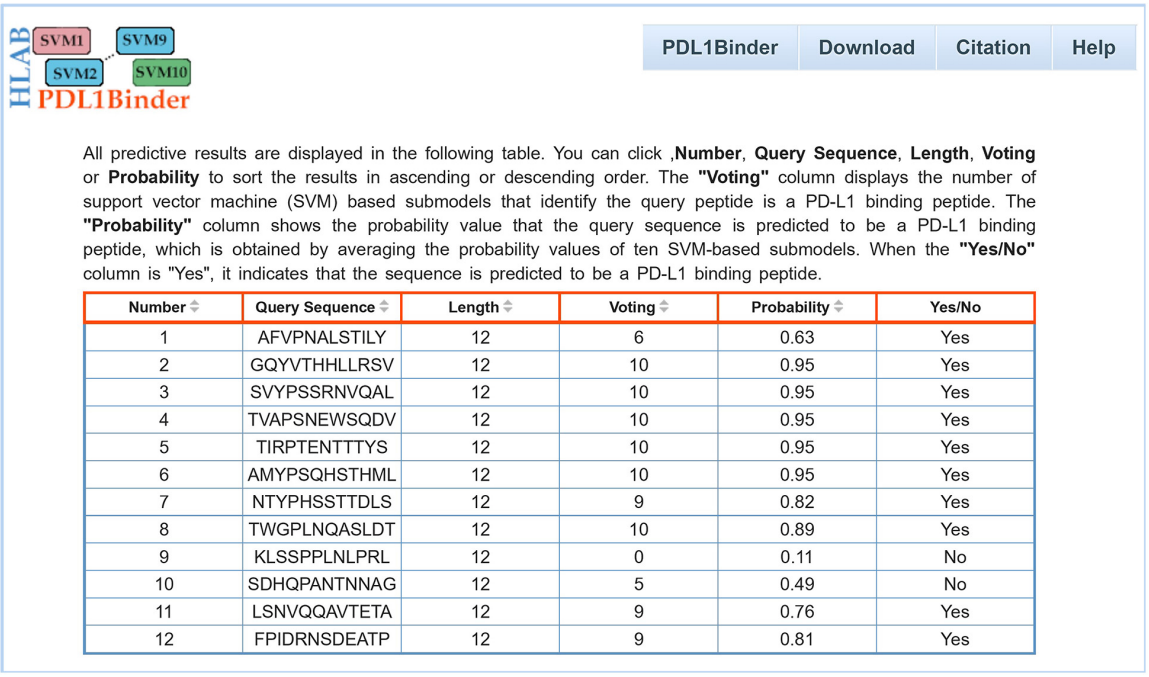
At present, no computational models have been proposed for efficiently discovering PD-L1 binding peptides. To pursue identifying PD-L1 binding peptides from pools of peptides with unknown functions, we designed a SVM-based classifier based on sequence information, called PDL1Binder, which could help to eliminate false positive peptides and improve the efficiency of obtaining PD-L1 binding peptides. The classifier integrates 10 SVM submodels. Two independent testing datasets were constructed to test the performance of PDL1Binder. Here, 83.33% of PD-L1 binding peptides in the TestDataset_1 were correctly identified as PD-L1 binding peptides, while 53.29% of non-PD-L1 binding peptides were precisely predicted as non-PD-L1 binding peptides in the TestDataset_2 when the *tp*-value was set to 0.55 within PDL1Binder. The proposed approach is considered as an applicable scheme for assisting the development of novel PD-L1 binding peptides.

PDL1Binder could be beneficial for both panning experiments and subsequent affinity determination experiments. The model can help researchers remove as many non-PD-L1 binding peptides as possible, thereby

A



B



Number	Query Sequence	Length	Voting	Probability	Yes/No
1	AFVPNALSTILY	12	6	0.63	Yes
2	GQYVTHHLLRSV	12	10	0.95	Yes
3	SVYPSSRNVQAL	12	10	0.95	Yes
4	TVAPSNEWSQDV	12	10	0.95	Yes
5	TIRPTENTTTYS	12	10	0.95	Yes
6	AMYPSQHSTHML	12	10	0.95	Yes
7	NTYPHSSTTDLS	12	9	0.82	Yes
8	TWGPLNQASLDT	12	10	0.89	Yes
9	KLSSPPLNLPRL	12	0	0.11	No
10	SDHQPANTNNAG	12	5	0.49	No
11	LSNVQQAVTETA	12	9	0.76	Yes
12	FPIDRNSDEATP	12	9	0.81	Yes

FIGURE 5
Webpage of PDL1Binder. **(A)** Input interface of PDL1Binder. Users can submit query sequences in FASTA or plain text format. The *tp* can be set by users, ranging from 0 to 1. **(B)** Output interface of PDL1Binder. PDL1Binder outputs the number of SVM-based submodels that identify the query peptide is a PD-L1 binding peptide and the probability value that the query sequence is predicted to be a PD-L1 binding peptide. The output likelihood value is obtained by averaging the probability values of 10 SVM-based submodels.

reducing both time and costs involved in getting PD-L1 binding peptide candidates. After testing on two independent test sets, we found that PDL1Binder was able to help remove around 5% of non-PD-L1 binding peptides (TestDataset_2) while retaining almost all PD-L1 binding peptides (TestDataset_1) ($tp = 0.2$). When the tp-value was set to 0.5, PDL1Binder correctly

TABLE 3 Performance of PDL1Binder in two independent testing datasets under different *tp*-values.

tp	0.1	0.15	0.20	0.25	0.30	0.35
TestDataset_1	100.00%	100.00%	100.00%	96.67%	93.33%	93.33%
TestDataset_2	0.80%	2.79%	5.03%	8.92%	15.18%	20.29%
tp	0.40	0.45	0.50	**0.55**	0.60	0.65
TestDataset_1	93.33%	86.67%	83.33%	**83.33%**	63.33%	53.33%
TestDataset_2	27.52%	37.39%	44.31%	**53.29%**	62.34%	71.17%
tp	0.70	0.75	0.80	0.85	0.90	0.95
TestDataset_1	43.33%	43.33%	30.00%	10.00%	3.33%	0.00%
TestDataset_2	80.84%	86.43%	92.26%	96.52%	99.09%	100.00%

tp, threshold of probability value to differentiate between predicted positives and negatives. Performance metric is the predictive accuracy of PD-L1 binding for TestDataset_1 and that of non-PD-L1 binding for TestDataset_2. Bold: The predictive accuracy of PD-L1 binding and that of non-PD-L1 binding have reached their maximum under tp = 0.55.

predicted 83.33% of PD-L1 binding peptides (TestDataset_1) while clearing away 44.31% of non-PD-L1 binding peptides (TestDataset_2). Additionally, our tool could successfully eliminate more than half of non-PD-L1 binding peptides (53.29%, TestDataset_2) while reserving 83.33% of PD-L1 binding peptides (TestDataset_1) ($tp = 0.55$). In the actual situation of random peptide library screening experiment, PD-L1 binding peptides are fewer and more precious, so researchers wish to keep as many putative PD-L1 binding peptides as possible in an experiment, while the proportion of non-PD-L1 binding peptides is much larger than that of PD-L1 binding peptides, thereby they wish to remove as many non-PD-L1 binding peptides as possible. We recommend users to set *tp* at 0.55 when using PDL1Binder since both the predictive accuracy of PD-L1 binding and that of non-PD-L1 binding have reached their maximum under this threshold. The above results indicate that PDL1Binder might save a significant amount of time and cost, greatly improving the efficiency of discovering PD-L1 binding peptides.

In the process of removing redundant peptides from 80 PD-L1 binding peptides identified by phage display screen, no redundant peptides were found and excluded. This suggests that these PD-L1 binding peptides seem to have a low sequence identity (below 0.8), which indicates that there are fewer features that are consistent within the PD-L1 binding sequences in the low-dimensional space. The SVM algorithm first projects the features in a low-dimensional space to those in a high-dimensional feature space through the RBF kernel function, and more consistent features are found in the high-dimensional feature space. We speculate that this might be a reason why SVM is superior to other machine learning algorithms. Another possible reason might be that LIBSVM utilizes L1 regularization (Chang and Lin, 2011), which could effectively avoid overfitting on a small training dataset. SVM with RBF kernel (RBFSVM) can handle the overfitting problem through selecting appropriate kernel width factor *gamma* and regularization factor *c*.

Our dataset for training is relatively small. More experimentally validated PD-L1 binding peptides will be needed to improve the performance of the computational model for identifying PD-L1 binding peptides. In the future, we will continue to improve the model and synthesize potential PD-L1 binding peptides predicted by the model to experimentally show if they can bind with PD-L1.

Conclusion

PD-L1 binding peptides are potential therapeutic agents for treating cancers. The PD-L1 binding peptides identified by phage display screen in this study are promising to become peptide drug candidates for blocking the PD-1/PD-L1 interaction to combat cancer. Computational models for identifying PD-L1 binding peptides can accelerate the discovery of these novel drug candidates. This study proposes the first SVM-based computational model, PDL1Binder, for effectively predicting peptides targeting PD-L1. We implemented PDL1Binder into an online web-server, which is freely accessible at http://i.uestc.edu.cn/pdl1binder/cgi-bin/PDL1Binder.pl. Our study showcases the potential of machine learning approaches for mining PD-L1 binding peptides from peptide pools of unknown bioactivities and provides promising PD-L1 binding peptide candidates for in-depth investigations.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

Author contributions

JH carried out the concept and design of this study. BH was responsible for data acquisition, constructed models, and drafted the manuscript. BL constructed models and prepared the manuscript for submission. QZ, BL, and CL repeated the model construction. SY and JL prepared the figures and tables. LN and HC guided modeling. All authors contributed to manuscript revision.

Funding

This work was supported by the National Natural Science Foundation of China (grant nos. 61901130, 61901129, 62071099, and 61571095), the Guizhou Provincial Science and Technology Projects [grant nos. (2020)1Y407, ZK(2022)-general-056, and ZK(2022)-general-038], Health Commission of Guizhou

Province (grant no: gzwkj2022-473), and the Guizhou University [grant nos. (2018) 54, (2018) 55, and (2020) 5].

Acknowledgments

We are grateful to the reviewers for their valuable suggestions and comments, which will lead to the improvement of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.928774/full#supplementary-material

References

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* 46, 175–185. doi: 10.2307/2685209

Asar, M. C., Franco, A., and Soendergaard, M. (2020). Phage display selection, identification, and characterization of novel pancreatic cancer targeting peptides. *Biomolecules* 10:714. doi: 10.3390/biom10050714

Bang, Y. J., Ruiz, E. Y., Van Cutsem, E., Lee, K. W., Wyrwicz, L., Schenker, M., et al. (2018). Phase III, randomised trial of avelumab versus physician's choice of chemotherapy as third-line treatment of patients with advanced gastric or gastro-oesophageal junction cancer: primary analysis of JAVELIN Gastric 300. *Ann. Oncol.* 29, 2052–2060. doi: 10.1093/annonc/mdy264

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Breimann, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. New York, NY: Taylor & Francis.

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transac. Intel. Syst. Technol.* 2, 1–27. doi: 10.1145/1961189.1961199

Chen, J., Guo, M., Wang, X., and Liu, B. (2018). A comprehensive review and comparison of different computational methods for protein remote homology detection. *Brief Bioinform.* 19, 231–244. doi: 10.1093/bib/bbw108

Chen, T., and Guestrin, C. (2016). "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, New York, NY.

Chen, Z., Zhao, P., Li, C., Li, F., Xiang, D., Chen, Y. Z., et al. (2021). iLearnPlus: a comprehensive and automated machine-learning platform for nucleic acid and protein sequence analysis, prediction and visualization. *Nucleic Acids Res.* 49, e60. doi: 10.1093/nar/gkab122

Chou, K.-C. (2009). Pseudo Amino Acid Composition and its Applications in Bioinformatics, Proteomics and System Biology. *Curr. Prot.* 6, 262–274. doi: 10.2174/157016409789973707

Chou, K. C. (2001). Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins* 43, 246–255. doi: 10.1002/prot.1035

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Fishman, J. A., Hogan, J. I., and Maus, M. V. (2019). Inflammatory and infectious syndromes associated with cancer immunotherapies. *Clin. Infect. Dis.* 69, 909–920. doi: 10.1093/cid/ciy1025

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3, 1289–1305. doi: 10.1162/153244303322753670

Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. Cambridge, MA: Cambridge University Press.

Freeman, G. J., Long, A. J., Iwai, Y., Bourque, K., Chernova, T., Nishimura, H., et al. (2000). Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J. Exp. Med.* 192, 1027–1034. doi: 10.1084/jem.192.7.1027

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi: 10.2307/2699986

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Hamanishi, J., Mandai, M., Matsumura, N., Abiko, K., Baba, T., and Konishi, I. (2016). PD-1/PD-L1 blockade in cancer treatment: perspectives and issues. *Int. J. Clin. Oncol.* 21, 462–473. doi: 10.1007/s10147-016-0959-z

He, B., Chai, G., Duan, Y., Yan, Z., Qiu, L., Zhang, H., et al. (2016). BDB: biopanning data bank. *Nucleic Acids Res.* 44, D1127–D1132. doi: 10.1093/nar/gkv1100

He, B., Chen, H., and Huang, J. (2019a). PhD7Faster 2.0: predicting clones propagating faster from the Ph.D.-7 phage display library by coupling PseAAC and tripeptide composition. *PeerJ* 7:e7131. doi: 10.7717/peerj.7131

He, B., Chen, H., Li, N., and Huang, J. (2019b). SAROTUP: a suite of tools for finding potential target-unrelated peptides from phage display data. *Int. J. Biol. Sci.* 15, 1452–1459. doi: 10.7150/ijbs.31957

He, B., Dzisoo, A. M., Derda, R., and Huang, J. (2019c). Development and application of computational methods in phage display technology. *Curr. Med. Chem.* 26, 7672–7693. doi: 10.2174/0929867325666180629123117

He, B., Jiang, L., Duan, Y., Chai, G., Fang, Y., Kang, J., et al. (2018a). Biopanning data bank 2018: hugging next generation phage display. *Database* 2018:bay032. doi: 10.1093/database/bay032

He, B., Tjhung, K. F., Bennett, N. J., Chou, Y., Rau, A., Huang, J., et al. (2018b). Compositional bias in naive and chemically-modified phage-displayed libraries uncovered by paired-end deep sequencing. *Sci. Rep.* 8:1214. doi: 10.1038/s41598-018-19439-2

Jaroszewicz, W., Morcinek-Orlowska, J., Pierzynowska, K., Gaffke, L., and Wegrzyn, G. (2022). Phage display and other peptide display technologies. *FEMS Microbiol. Rev.* 46:fuab052. doi: 10.1093/femsre/fuab052

Juds, C., Schmidt, J., Weller, M. G., Lange, T., Beck, U., Conrad, T., et al. (2020). Combining Phage Display and Next-Generation Sequencing for Materials Sciences: a Case Study on Probing Polypropylene Surfaces. *J. Am. Chem. Soc.* 142, 10624–10628. doi: 10.1021/jacs.0c03482

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). "Lightgbm: A highly efficient gradient boosting decision tree," in *Proceedings of*

the 31st International Conference on Neural Information Processing Systems, Red Hook, NY. doi: 10.5555/3294996.3295074

Laustsen, A. H., Greiff, V., Karatt-Vellatt, A., Muyldermans, S., and Jenkins, T. P. (2021). Animal immunization, *in vitro* display technologies, and machine learning for antibody discovery. *Trends Biotechnol.* 39, 1263–1273. doi: 10.1016/j.tibtech.2021.03.003

Ledsgaard, L., Ljungars, A., Rimbault, C., Sorensen, C. V., Tulika, T., Wade, J., et al. (2022). Advances in antibody phage display technology. *Drug Discov. Today* 27, 2151–2169. doi: 10.1016/j.drudis.2022.05.002

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Li, C., Zhang, N., Zhou, J., Ding, C., Jin, Y., Cui, X., et al. (2018). Peptide Blocking of PD-1/PD-L1 Interaction for Cancer Immunotherapy. *Cancer Immunol. Res.* 6, 178–188. doi: 10.1158/2326-6066.CIR-17-0035

Li, Z., Wu, X., Zhao, Y., Xiao, Y., Zhao, Y., Zhang, T., et al. (2021b). Clinical benefit of neoadjuvant anti-PD-1/PD-L1 utilization among different tumors. *MedComm* 2, 60–68. doi: 10.1002/mco2.61

Li, W., Zhu, X., Zhou, X., Wang, X., Zhai, W., Li, B., et al. (2021a). An orally available PD-1/PD-L1 blocking peptide OPBP-1-loaded trimethyl chitosan hydrogel for cancer immunotherapy. *J. Control Release* 334, 376–388. doi: 10.1016/j.jconrel.2021.04.036

Lin, X., Lu, X., Luo, G., and Xiang, H. (2020). Progress in PD-1/PD-L1 pathway inhibitors: from biomacromolecules to small molecules. *Eur. J. Med. Chem.* 186, 111876. doi: 10.1016/j.ejmech.2019.111876

Liu, G., Zeng, H., Mueller, J., Carter, B., Wang, Z., Schilz, J., et al. (2020). Antibody complementarity determining region design using high-capacity machine learning. *Bioinformatics* 36, 2126–2133. doi: 10.1093/bioinformatics/btz895

Liu, H., Zhao, Z., Zhang, L., Li, Y., Jain, A., Barve, A., et al. (2019). Discovery of low-molecular weight anti-PD-L1 peptides for cancer immunotherapy. *J. Immunother. Cancer* 7:270. doi: 10.1186/s40425-019-0705-y

Ma, Y., Guo, Z., Xia, B., Zhang, Y., Liu, X., Yu, Y., et al. (2022). Identification of antimicrobial peptides from the human gut microbiome using deep learning. *Nat. Biotechnol.* 40, 921–931. doi: 10.1038/s41587-022-01226-0

Matochko, W. L., Cory, Li, S., Tang, S. K., and Derda, R. (2014). Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res.* 42, 1784–1798. doi: 10.1093/nar/gkt1104

Matochko, W. L., and Derda, R. (2015). Next-generation sequencing of phage-displayed peptide libraries. *Methods Mol. Biol.* 1248, 249–266. doi: 10.1007/978-1-4939-2020-4_17

McLachlan, G. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. New York, NY: John Wiley & Sons.

Pan, C., Yang, H., Lu, Y., Hu, S., Wu, Y., He, Q., et al. (2021). Recent advance of peptide-based molecules and nonpeptidic small-molecules modulating PD-1/PD-L1 protein-protein interaction or targeting PD-L1 protein degradation. *Eur. J. Med. Chem.* 213:113170. doi: 10.1016/j.ejmech.2021.113170

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.48550/arXiv.1201.0490

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/TPAMI.2005.159

Pleiko, K., Posnograjeva, K., Haugas, M., Paiste, P., Tobi, A., Kurm, K., et al. (2021). *In vivo* phage display: identification of organ-specific peptides using deep sequencing and differential profiling across tissues. *Nucleic Acids Res.* 49:e38. doi: 10.1093/nar/gkaa1279

Postow, M. A., Chesney, J., Pavlick, A. C., Robert, C., Grossmann, K., McDermott, D., et al. (2015). Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. *N. Engl. J. Med.* 372, 2006–2017. doi: 10.1056/NEJMoa1414428

Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., et al. (2008). A large genome center's improvements to the Illumina sequencing system. *Nat. Methods* 5, 1005–1010. doi: 10.1038/nmeth.1270

Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. (2003). "Tackling the poor assumptions of naive bayes text classifiers," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, Washington, DC. doi: 10.5555/3041838.3041916

Robert, C., Schachter, J., Long, G. V., Arance, A., Grob, J. J., Mortier, L., et al. (2015). Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* 372, 2521–2532. doi: 10.1056/NEJMoa1503093

Stigler, S. M. (1989). Francis Galton's account of the invention of correlation. *Stat. Sci.* 4, 73–79. doi: 10.1214/ss/1177012580

Sun, P., Qi, J., Zhao, Y., Huang, Y., Yang, G., Ma, Z., et al. (2016). A novel conformational B-cell epitope prediction method based on mimotope and patch analysis. *J. Theor. Biol.* 394, 102–108. doi: 10.1016/j.jtbi.2016.01.021

Talantova, M., Sanz-Blasco, S., Zhang, X., Xia, P., Akhtar, M. W., Okamoto, S., et al. (2013). Aβ induces astrocytic glutamate release, extrasynaptic NMDA receptor activation, and synaptic loss. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2518–E2527. doi: 10.1073/pnas.1306832110

Tallorin, L., Wang, J., Kim, W. E., Sahu, S., Kosa, N. M., Yang, P., et al. (2018). Discovering de novo peptide substrates for enzymes using machine learning. *Nat. Commun.* 9:5253. doi: 10.1038/s41467-018-07717-6

Tang, J., Yu, J. X., Hubbard-Lucey, V. M., Neftelinov, S. T., Hodge, J. P., and Lin, Y. (2018). Trial watch: the clinical trial landscape for PD1/PDL1 immune checkpoint inhibitors. *Nat. Rev. Drug Discov.* 17, 854–855. doi: 10.1038/nrd.2018.210

Tooyserkani, R., Rasaee, M. J., Bandehpour, M., and W P M Löwik, D. (2021). Novel anti-PD-L1 peptide selected from combinatorial phage library inhibits tumor cell growth and restores T-cell activity. *J. Drug Target* 29, 771–782. doi: 10.1080/1061186X.2021.1879087

Tumeh, P. C., Harview, C. L., Yearley, J. H., Shintaku, I. P., Taylor, E. J., Robert, L., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* 515, 568–571. doi: 10.1038/nature13954

Wolchok, J. D., and Chan, T. A. (2014). Cancer: antitumour immunity gets a boost. *Nature* 515, 496–498. doi: 10.1038/515496a

Yang, C. Y., Lin, M. W., Chang, Y. L., Wu, C. T., and Yang, P. C. (2016). Programmed cell death-ligand 1 expression is associated with a favourable immune microenvironment and better overall survival in stage I pulmonary squamous cell carcinoma. *Eur. J. Cancer* 57, 91–103. doi: 10.1016/j.ejca.2015.12.033

Yu, L., and Liu, H. (2003). "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, Washington, DC. doi: 10.5555/3041838.3041946

frontiers | Frontiers in Microbiology

METHODS
published: 15 July 2022
doi: 10.3389/fmicb.2022.932661

Ensemble Learning-Based Feature Selection for Phage Protein Prediction

Songbo Liu[1], *Chengmin Cui*[2], *Huipeng Chen*[1]* *and Tong Liu*[1]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, [2] Beijing Institute of Control Engineering, China Academy of Space Technology, Beijing, China

Phage has high specificity for its host recognition. As a natural enemy of bacteria, it has been used to treat super bacteria many times. Identifying phage proteins from the original sequence is very important for understanding the relationship between phage and host bacteria and developing new antimicrobial agents. However, traditional experimental methods are both expensive and time-consuming. In this study, an ensemble learning-based feature selection method is proposed to find important features for phage protein identification. The method uses four types of protein sequence-derived features, quantifies the importance of each feature by adding perturbations to the features to influence the results, and finally splices the important features among the four types of features. In addition, we analyzed the selected features and their biological significance.

Keywords: machine learning, ensemble learning, feature selection, phage, protein classification

OPEN ACCESS

Edited by:
Jian Huang,
University of Electronic Science and Technology of China, China

Reviewed by:
Zhiwei Ji,
Nanjing Agricultural University, China
Yi Xiong,
Shanghai Jiao Tong University, China

***Correspondence:**
Huipeng Chen
chp@ir.hit.edu.cn

Specialty section:
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

Received: *30 April 2022*
Accepted: *14 June 2022*
Published: *15 July 2022*

Citation:
Liu S, Cui C, Chen H and Liu T (2022) Ensemble Learning-Based Feature Selection for Phage Protein Prediction. Front. Microbiol. 13:932661. doi: 10.3389/fmicb.2022.932661

INTRODUCTION

Phages, which are the most abundant and widespread organisms on the Earth, can replicate within and destroy the host cell. Phages play an important role in microbial physiology, population dynamics, evolution, and therapy (Clokie et al., 2011), affecting biochemical systems worldwide (Jahn et al., 2019).

Phages also influence the development of anti-cancer drugs. The use of phages to target cancer cells for a specific binding for therapeutic purposes has been applied in clinical trials for cancer treatment due to differences in surface markers of tumor cells from normal cells (Yu et al., 2021). The identification of phage proteins is important for understanding the relationship between phages and host bacteria and for developing novel drugs or antibiotics (Lekunberri et al., 2017), and therefore, thorough investigations must be performed to identify the specific components recognized by phages.

Traditional physical experimental methods such as mass spectrometry (MS), sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and protein arrays (Lavigne et al., 2009; Yuan and Gao, 2016; Jara-Acevedo et al., 2018), which have been used to identify phage viral proteins, are expensive and often time-consuming. Traditional biological methods such as cell separation, electron microscopy, and fluorescence microscopy are less feasible for analyzing large-scale biological data (Mei, 2012; Li et al., 2015). Computational models can not only analyze large amounts of biological data but also make preliminary predictions of unknown protein sequences, which is an excellent complement to traditional experimental methods.

In recent years, protein function prediction has been a hot topic in the field of computational biology (Ding et al., 2020; Fu et al., 2020; Guo et al., 2020). With the increasing amount of protein data, the techniques of applying machine learning and data mining to protein function prediction have gradually matured (Liu et al., 2019; Zhao et al., 2021). Several researchers have used machine learning to predict the function of protein sequences through sequence analysis (Chou, 2009; Cui et al., 2019; Jin et al., 2021), position-specific scoring matrix (PSSM) (Jones, 1999), various physicochemical and biochemical properties of amino acids, sequence conservation, amino acid composition, domain interactions, and geometrical and biophysical properties (Kawashima and Kanehisa, 2000; Cai et al., 2003; Zulfiqar et al., 2021). Feng et al. (2013) developed a Naive Bayes-based model for protein classification, which used amino acid composition (AAC) and dipeptide combination (DPC) as input features. Ding et al. (2014) developed a support vector machine (SVM) prediction model. In this method, analysis of variance was applied to select significant features from the g-gap DPC. Recently, Zhang et al. (2015) developed a random forest classification method to distinguish phage virus protein (PVP) from non-PVP. A novel feature extraction method with a two-layered structure is proposed (Xiong et al., 2018; Jiang et al., 2021). First, the features irrelevant to the results are removed by the filter or wrapper method, and then, the results of the previous step are used in the model for classification and prediction.

Since each feature extraction method extracts only part of the protein sequence information, the method of combining multiple protein information for classification is proposed in the absence of a clear sequence or structural information. Jiao and Du (2017) proposed the functional domain enrichment score with position-specific physicochemical properties (PSPCP). Li et al. (2015) proposed to fuse the position-specific scoring matrix (PSSM) and gene ontology to extract feature sequences.

Protein sequences often have high feature dimensionality and contain a large amount of redundant information, which reduces the prediction performance of a model. Dimensionality reduction of feature vectors with high-dimensional data is performed to eliminate unnecessary features. The cmmonly used methods are principal component analysis (PCA) (Ahmad et al., 2016), information gain (Wen et al., 2016), maximum correlation and minimum redundancy (MRMR) (Khan et al., 2017), maximum correlation maximum distance (MRMD) (Zou et al., 2016), singular value decomposition (SVD) (Silvério-Machado et al., 2015), local linear discriminant analysis (Yu et al., 2018), and dipeptide composition (DPC) (Ahmad et al., 2016). Xie et al. (2021) proposed a method for k-size optimal parsimony features based on the rough set theory, which found the effective features by fixing the parsimony size and dynamic weighting strategy. NMFBFS reduced the dimensionality of the data by decomposing the non-negative matrix of the data (Ji et al., 2015). Despite the specific advantages of existing methods for PVP prediction, there is a need to improve the accuracy and transferability of predictive models.

In the present study, we propose an ensemble learning-based feature selection method for phage virus protein classification that uses a four-step pipeline for protein prediction, (I) extracting the amino acid composition content (AAC), physicochemical properties CTD, dipeptide composition CKSAAP, and reduced position specificity scoring matrix (RPSSM) of the protein; (II) using ensemble learning to measure the importance of each feature component from each type of feature; (III) using an incremental strategy to select the most important feature subset; and (IV) combining the optimal feature subset derived from each type of feature to retrain the data to be filtered again and finally applying the obtained optimal feature subset to predict the protein type. Instead of the PCA-like feature dimension reduction method, our method can directly obtain important features for further biological analysis. Experimental results demonstrate the effectiveness of our method.

MATERIALS AND METHODS

Data

In this study, we used the dataset constructed by Ding et al. (2014). This dataset was processed by UniProt in the following ways. First of all, the phage proteins whose subcellular location is a virion were considered positive sample and *vice versa*. The sequences containing unknown amino acids such as "B," "J," "O," "U," "X," or "Z" were removed. To eliminate the influence of homologous sequences, more than 40% homologous sequences were removed by CD-HIT. Eventually, 99 phage proteins and 208 non-phage proteins were obtained.

The data from various literature studies (Feng et al., 2013; Ding et al., 2014; Zhang et al., 2015) were processed in the same way to build an independent test set. Besides, more than 40% homologous sequences with the training set were removed.

Feature Extraction

The functions of a protein consists of the amino acid type, quantity, arrangement order of the peptide chain, and the spatial structure of the protein. Therefore, the main description methods of a protein can be divided into the global description of the protein and the description of the amino acid level (Xu et al., 2020). The global description of the protein includes the first-class features and spatial features of the protein. However, the acquisition of spatial features is expensive. Under certain conditions, the primary structure of the protein can determine the secondary, tertiary, and quaternary structures. Therefore, in this study, we only extracted the primary features of the protein and also extracted simple spatial structural features from the primary structure.

For the convenience of discussion, we define a protein sequence P as follows:

$$P = p_1 p_2 p_3 \ldots p_L, p_i \in \{A,\ C,\ D,\ \ldots Y\} \tag{1}$$

where p_i is an amino acid, i is the position of the amino acid in the sequence, and L is the length of the amino acid sequence.

AAC

The Amino Acid Composition (AAC) is to calculate the content of each amino acid in a protein sequence. The AAC feature of an

amino acid sequence is as follows:

$$AAC_i = \frac{N\left(p_i\right)}{L},\ 0 < i \leq 20 \tag{2}$$

where AAC_i represents the proportion of p_i in a protein sequence, $N\left(p_i\right)$ is the number of the amino acid p_i in a protein sequence, and L is the number of amino acids in a sample.

CTD

Protein-related chemical reactions commonly occur in the cell and tissue fluids, so the physicochemical properties of a protein are closely related to the function of the protein. There are eight types commonly used with the following physicochemical properties: hydrophobicity, polarity, surface tension, polarizability, charge, van der Waals force, secondary structure, and solubility (Cai et al., 2003).

For each physicochemical property, the amino acids are divided into three groups (positive, neural, and negative), and then, the three values of Composition (C), Transition (T), and Distribution (D) of each property are calculated. C is the percentage of the composition of a certain physicochemical property. T describes the following three types of residue pairs: a negative residue followed by a neural residue; a positive residue followed by a negative residue; and a positive followed by a neural residue. The percentage of the amino acids of a particular property located at the first, 25, 50, 75, and 100% is measured as the distribution of the protein (D). Finally, 168 physicochemical features are obtained.

CKSAAP

The composition of k-space amino acid pairs (CKSAAP) is encoded by the proportion of amino acid pairs separated by any k residues.

$$\begin{aligned} &CKSAAP = \frac{N_{p_i p_j}}{L}, p_i, p_j \in \{A,\ C,\ D,\ \ldots,\ Y\} \\ &j = i + k + 1,\ i,\ j \leq L, \end{aligned} \tag{3}$$

where $N_{p_i p_j}$ is the proportion of residue pairs p_i and p_j. According to Ding et al. (2014), the best results of extracting phage protein features for classification are obtained when $k = 1$.

RPSSM

In the process of biological evolution, amino acid sequences mutate corresponding to common changes, including deletion, substitution, and insertion of amino acid residues. However, these changed protein sequences still have similar structures and functions (Ding et al., 2014). The position-specific scoring matrix (PSSM) can describe this change.

For the convenience of description, we defined the sequence P_A as follows:

$$P_A = (P_{1A}, P_{2A},\ \ldots\ , P_{LA})^T \tag{4}$$

where P_{iA} indicates the score of amino acid mutation to amino acid A at the *ith* position of the protein sequence.

According to the similarity of amino acids, Li et al. (2003) found that 10 residues can construct a set with the smallest reasonable folding model, i.e., {F, Y, W}, {M, L}, {I, V}, {A, T, S}, {N, H}, {Q, E, D}, {R, K}, {P}, {C}, and {G}. We combined the same categories based on the similarity of amino acids as follows:

$$\begin{cases} P_1 = \frac{(P_F + P_Y + P_W)}{3}, \\ P_2 = \frac{(P_M + P_L)}{2}, \\ \ldots, \\ P_{10} = P_G. \end{cases} \tag{5}$$

The reduced position-specific scoring matrix (RPSSM) is represented as follows:

$$PSSM_S = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,j} & \cdots & P_{1,10} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,j} & \cdots & P_{2,10} \\ \vdots & \vdots & & \vdots & & \\ P_{i,1} & P_{i,2} & \cdots & P_{i,j} & \cdots & P_{i,10} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{L,1} & P_{L,2} & \cdots & P_{L,j} & \cdots & P_{L,10} \end{bmatrix}. \tag{6}$$

The variance of each column is calculated to get the feature D_s,

$$D_S = \frac{1}{L}\sum_{i=1}^{L}\left(p_{is} - \bar{p}_s\right)^2,\ s \leq 10,\ i \leq L \tag{7}$$

However, D_S does not contain the amino acid order information. To get the information on the local sequence order effect, the dipeptide composition of the protein sequence is extended to PSSM. Assuming that p_i is mutated into p_s and that p_{i+1} is mutated into p_t in the sequence, then

$$\begin{aligned} D_{i,i+1} &= \left(p_{i,s} - \frac{p_{i,s} + p_{i+1,t}}{2}\right)^2 + \left(p_{i+1,t} - \frac{p_{i,s} + p_{i+1,t}}{2}\right)^2 \\ &= \frac{\left(p_{i,s} - p_{i+1,t}\right)^2}{2} \end{aligned} \tag{8}$$

where $s, t = 1, 2, \ldots 10,\ i = 1, 2, \ldots\ L\ 1$, and $D_{i,i+1}$ represents the difference between the average values of $p_{i,s}$ and $p_{i+1,t}$, while $D_{i,i+1}$ in the protein sequence is expressed as follows:

$$D_{s,t} = \frac{1}{L}\sum_{i-1}^{L-1}\frac{\left(p_{i,s} - p_{i+1,t}\right)^2}{2},\ s,\ t = 1, 2, \ldots,\ 10 \tag{9}$$

Finally, 110 features were obtained by splicing $D_{s,t}$ and D_s.

Feature Selection

The goal of the feature-selection module is to select as few features as possible with guaranteed classification accuracy so that the model does not degrade significantly when the model is learned using only the subset of features and the learning results are close to or higher than the learning results using the full set of features.

The relationship between the sequence structure and the function of protein is not entirely clear. Therefore, the features based on knowledge extraction are not necessarily related to the function of the protein or are even some irrelevant features.

Redundancy will also affect the fitting of the classifier to protein data and will interfere with the prediction. Therefore, we cannot select features only in a knowledge-driven way. It is necessary to further screen the extracted features in a data-driven way so as to screen out effective feature subsets beneficial to the learning algorithm. This can not only reduce the difficulty of learning tasks but also improve the efficiency of the model. We assume that only some of all the features play a decisive role in model fitting. If this feature is modified, it will have a greater impact on the results. To quantify this effect, we proposed a feature selection model. The specific steps we followed were as follows.

Importance Accumulation

First, the features from four types of protein sequences were extracted by a feature extraction module in **Figure 1A**. Then, the importance score of each feature component was calculated. According to the feature selection module shown in **Figure 1B**, we assumed that all the features extracted are valid for classification. Then, 70% of the training samples were randomly selected to train the classifier, and the classification accuracy rate was obtained by testing the data out-of-bag, which is recorded as $score_1$. If a feature component was very relevant to the protein function, just changing it would have a great impact on the result. According to this idea, we only shuffled the values of a feature component randomly by a permutation way on the testing dataset in order to maintain the data distribution unchanged before and after this modification, then used the model trained on the training set to predict the shuffled data, and obtained the classification accuracy, which is recorded as $score_2$. The importance of this feature component is defined as:

$$imp_i = score_1 - score_2 \tag{10}$$

Each feature component is shuffled to get its important score. To reduce the error, it repeats n rounds to get the average importance scores of each feature.

Incremental Feature Selection

To find the optimal feature subset, we added each feature component incrementally according to its score in descending order, trained the classification model, and calculated the classification accuracy and finally showed the result in **Figure 1C**. In this way, a feature subset that maintains a comparative classification result equivalent to the feature is selected and considered an important feature. Then, the classifier was used in a feature subset to train all the training samples, and the independent test set was used to predict the phage protein. The final processed was shown in **Figure 1D**.

Ensemble Classifier

Due to the unclear sample distribution and various classification boundaries, only using a single classifier may not fit the data well and may not get a good classification result. Therefore, we used seven common classifiers with default parameters as the base classifier in scikit-learn (Pedregosa et al., 2011), linear discriminant analysis (LDA), decision tree classifier (DTC), k-neighbors classifier (KNN), support vector machine (SVM), logistic regression (LR), Gaussian Naive Bayes classifier (GNB), and multilayer perceptron (MLP) using 1,000 iterations. At the same time, seven classifiers were trained in each sampling, and the classifier with the highest accuracy was selected. We believed that this classifier can best fit the distribution of the data. This classifier was used as the best classifier in this round to classify or calculate feature importance. Although the ensemble learning method can be insensitive to the distribution of data, the time

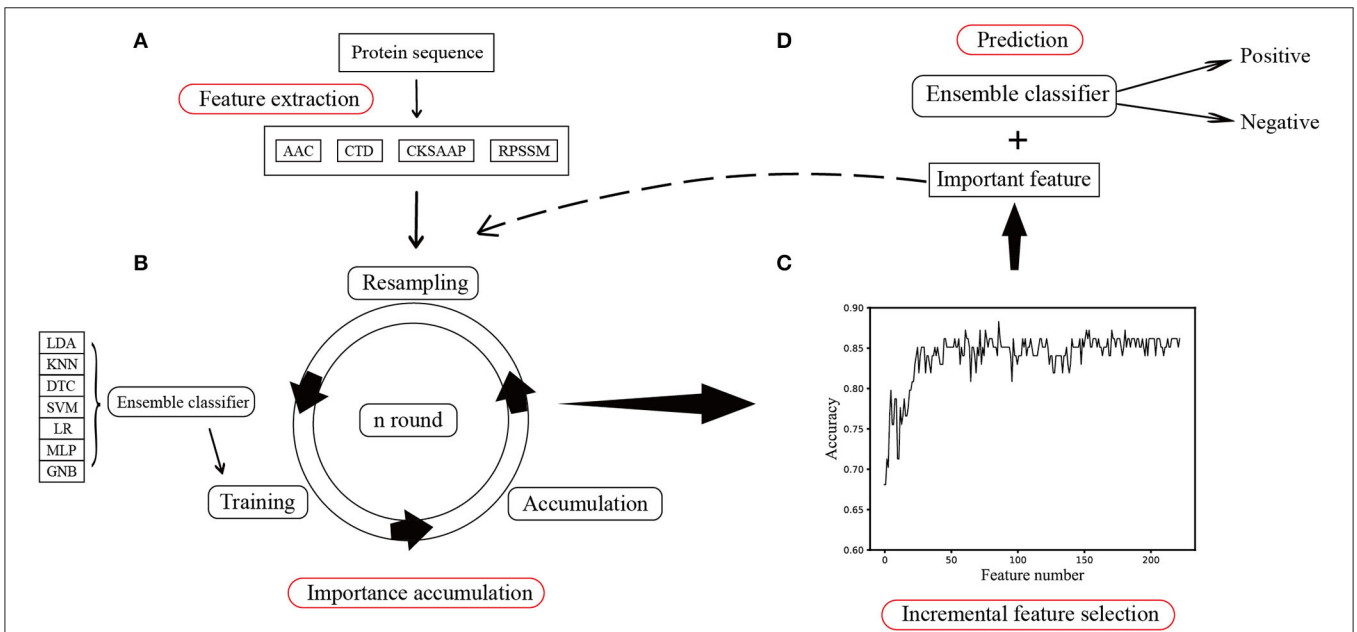


FIGURE 1 | Flowchart of the proposed method. **(A)** Feature extraction method includes extracting AAC, CTD, CKSAAP-1GAP, and RPSSM from the original protein sequences. **(B)** Resampling all samples and then calculating the importance of each feature component using the ensemble classification. **(C)** Selecting a subset of important features using the incremental method. **(D)** Predicting phage proteins on the testing set using the important feature components.

cost will increase due to the use of multiple classifiers. The time complexity of the proposed method $O(rounds* \sum_i classifier_i)$ was mainly to calculate the importance score of features, where rounds refer to the number of iterations. Therefore, if the time complexity needs to be reduced, the preferred method is to choose a classifier with smaller time complexity.

Evaluation Criteria

To evaluate our model comprehensively, we used the common measures, i.e., ACC, SN, and SP. These methods are defined as follows:

$$ACC = \frac{TN + TP}{TP + FN + FN + FP} \quad (11)$$

$$SN = \frac{TP}{TP + FN} \quad (12)$$

$$SP = \frac{TN}{TN + FP} \quad (13)$$

where TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively.

RESULTS

The function of the protein consists of amino acid composition, arrangement order, and spatial structure. In this study, we selected amino acid content (AAC), physicochemical properties (CTD), dipeptide (CKSAAP-1GAP), and PSSM matrix of protein (RPSSM) as the features of protein data. The classification accuracy of each feature was calculated using the ensemble classifier to get the importance of each feature component. According to the importance of feature components, the effective feature subset of each feature of different species was selected in an incremental way. Finally, all feature subsets were spliced and predicted on the independent test set and compared with other methods.

Performance Comparison of Individual Types of Features on Training Dataset Results Using Independent Features

For features of different types, each classification model in the ensemble classifier was trained using the training dataset, and

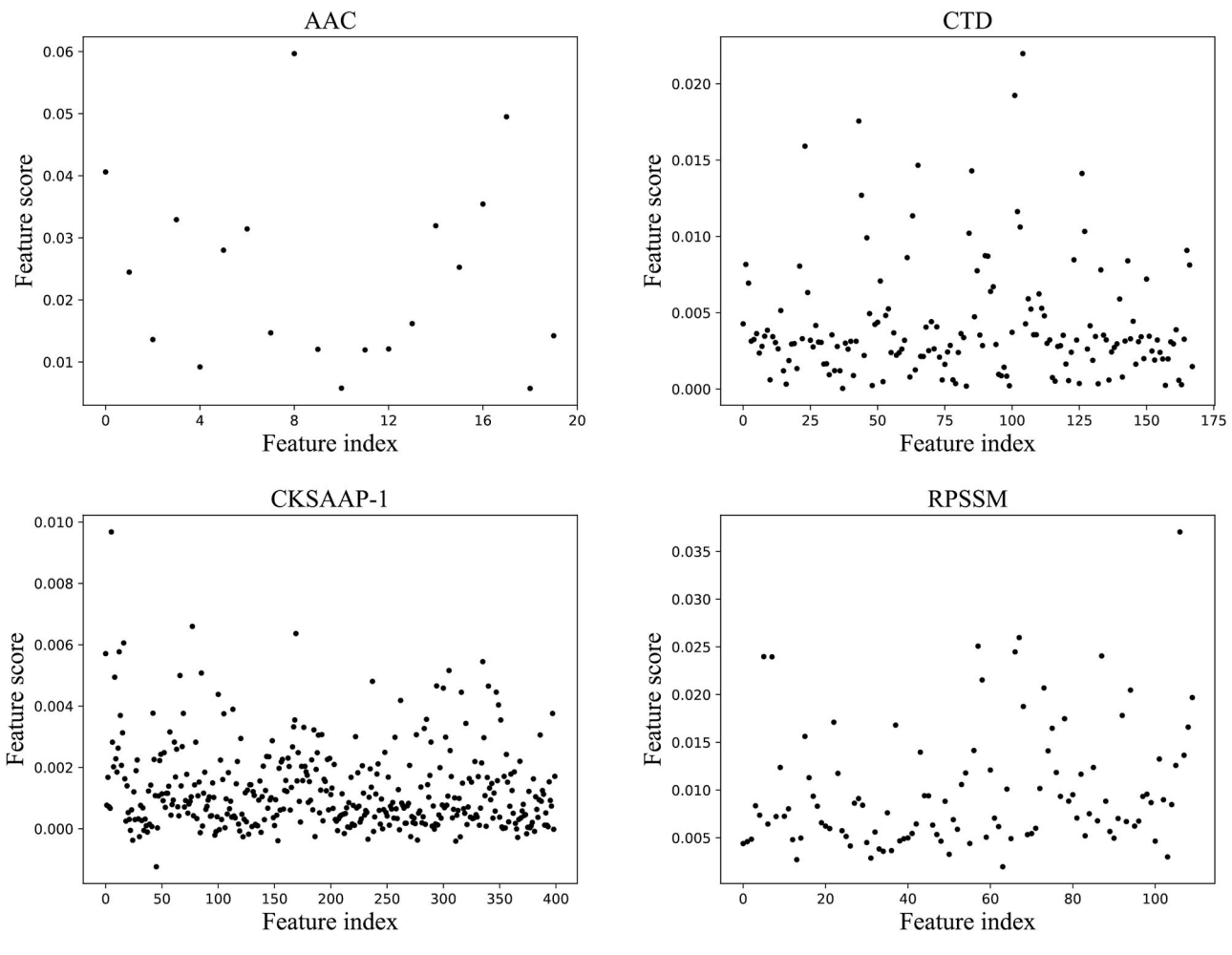


FIGURE 2 | The scatterplots of feature importance using different feature extraction methods.

the model with the best fitting effect was selected by ACC in the ensemble classifier. The feature importance score was calculated using a feature selection module shown in **Figure 1B**. **Figure 2** shows the scatterplots of the accumulated scores according to feature components. It can be found that AAC can affect the results by 6% at the highest and the physicochemical properties (CTD) by 2%, but dipeptide can only affect the results by 1% at the highest. RPSSM has an impact of 3.5%. Some features have no effect on the results after modification.

To get rid of the ineffective feature components, we used the incremental method according to the rank of feature importance scores. After stacking feature components according to their importance, we trained the ensemble classifier and obtained the classification accuracies. The ACC curve was obtained using a 10-fold cross-validation method, as seen in **Figure 3**.

It can be found that the classification accuracy of the model in AAC has shown an increasing trend with the superposition of feature components, but it is around 82%. The accuracy of physicochemical properties (CTD) is more stable when features are superimposed. It can be found that the classification accuracy is the highest using only the first 50 important features, and the classification accuracy does not increase when feature components are superimposing. The classification accuracy of the first 103 feature components in CKSAAP-1GAP is also good with the highest classification accuracy, and the classification accuracy even decreases when new feature components are added. RPSSM is always more stable, and the first 50 feature components with the highest classification accuracy are selected as the optimal feature subset of the features.

Performance Comparison of Concatenated Features on Training Data Set Results Using Integrated Features

We selected the subset of features with the best classification result and the least number of features among different kinds of features and spliced the obtained subset of features to form an important feature with multiple types of superpositions.

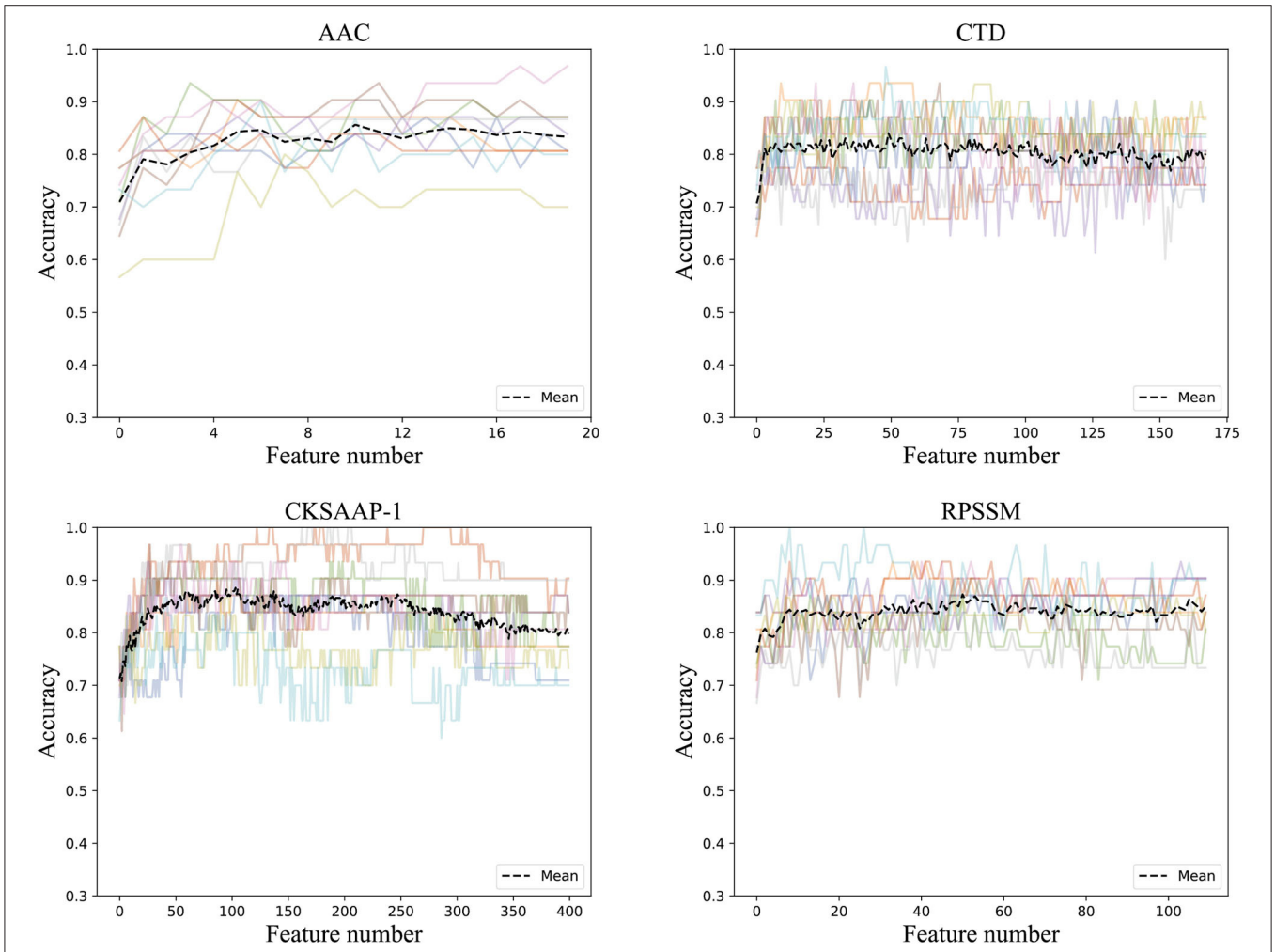


FIGURE 3 | Accuracies derived from the incremental strategy using different feature extraction methods (The colored line shows the accuracy results of the 10-fold data).

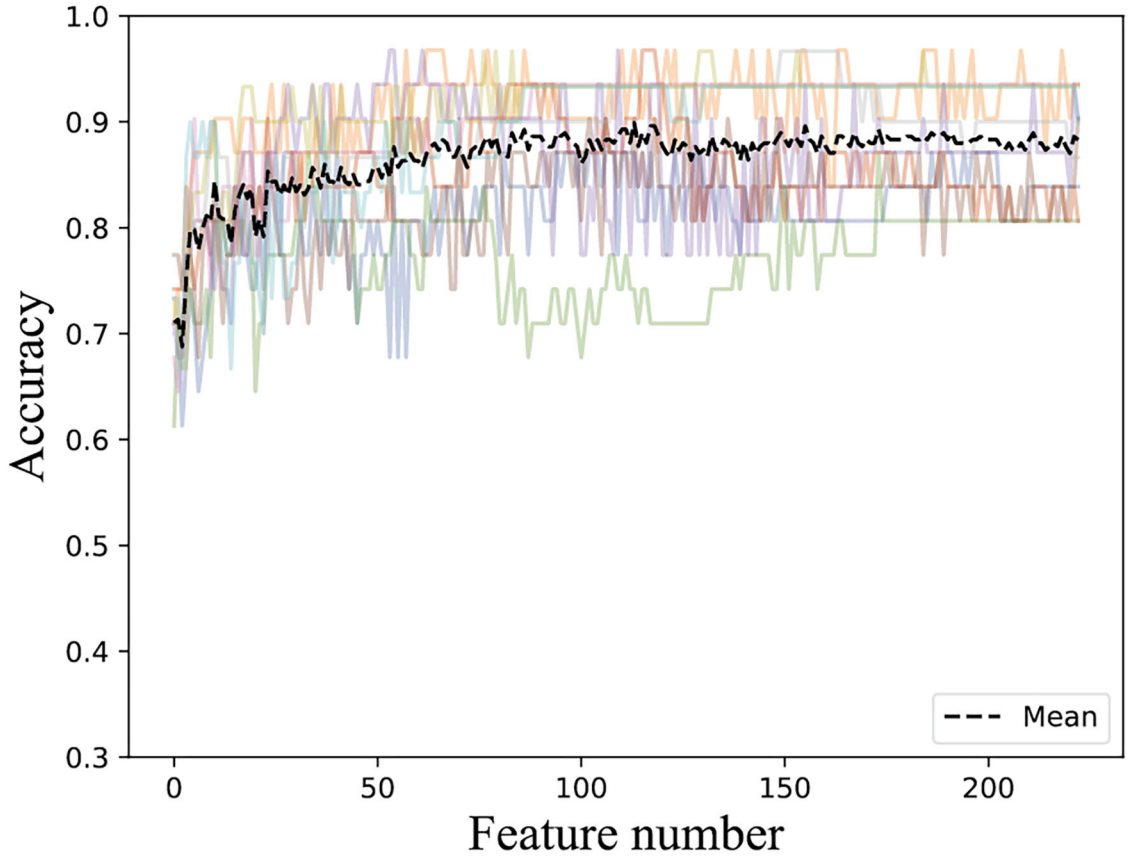


FIGURE 4 | Accuracies derived from the incremental strategy using integrated features.

TABLE 1 | Average accuracy of a 10-fold cross-validation on the training set using different features.

Classifier	Features	SN (%)	SP (%)	ACC (%)
DTC, GNB, LR, MLP	AAC (20D)	71.89	85.02	80.81
MLP, KNN, DTC, LR	CTD (168D)	69.00	86.12	80.41
KNN, MLP, LR, GNB	CKSAAP_1gap (400D)	70.78	80.71	77.51
GNB, LDA, LR, MLP	RPSSM (110D)	82.78	79.81	80.81
MLP, GNB, LR, DTC, KNN	Concatenation (698D)	56.67	90.63	79.79
GNB, MLP, KNN, LR	**CTD (50D)**	**73.88**	**89.40**	**84.29**
GNB, KNN	**CKSAAP_1gap (103D)**	**82.06**	**88.07**	**86.04**
LR, MLP, DTC, GNB, LDA	**RPSSM (50D)**	**81.78**	**87.05**	**85.37**
DTC, LR, MLP, LDA, KNN	**Concatenation (38D)**	**83.00**	**85.05**	**86.00**
GNB, LDA, LR	**Concatenation (87D)**	**89.00**	**86.55**	**89.28**

Bold indicates the result of the processing of the features.

TABLE 2 | Accuracy on independent test sets using different kinds of features.

Classifier	Features	SN (%)	SP (%)	ACC (%)
MLP	AAC (20D)	66.67	85.94	79.79
MLP	CTD (168D)	70.00	79.69	76.60
GNB	CKSAAP_1gap (400D)	63.33	85.94	78.72
GNB	RPSSM (110D)	73.33	76.56	75.53
GNB	**Concatenation (698D)**	56.67	90.63	79.79
LDA	**CTD (50D)**	**70.00**	**82.81**	**78.72**
SVM	**CKSAAP_1gap (103D)**	**46.67**	**92.19**	**77.66**
MLP	**RPSSM (50D)**	**60.00**	**92.19**	**81.92**
MLP	**Concatenation (38D)**	**73.33**	**90.63**	**85.11**
MLP	**Concatenation (87D)**	**73.33**	**93.75**	**87.23**

Bold indicates the result of the processing of the features.

According to the important feature subset selected from the AAC sequence feature, the physicochemical feature, dipeptide 1-gap content, which has been indicated to be the best (Ding et al., 2014), and RPSSM feature, we obtained 223 feature components totally. Accordingly, the average classification accuracy was calculated in turn on the training set using a 10-fold cross-validation. The qualitative results are shown in **Figure 4**. The quantitative results compared with other methods are listed in **Tables 1**, **2**.

It can be seen from the 10-fold results that the classification accuracy of only the first 20 feature components of AAC is better than that of the others. On the independent test set, the classification accuracy rate is 79.79%. It shows that the amino acid content of phage protein is quite different from that of the

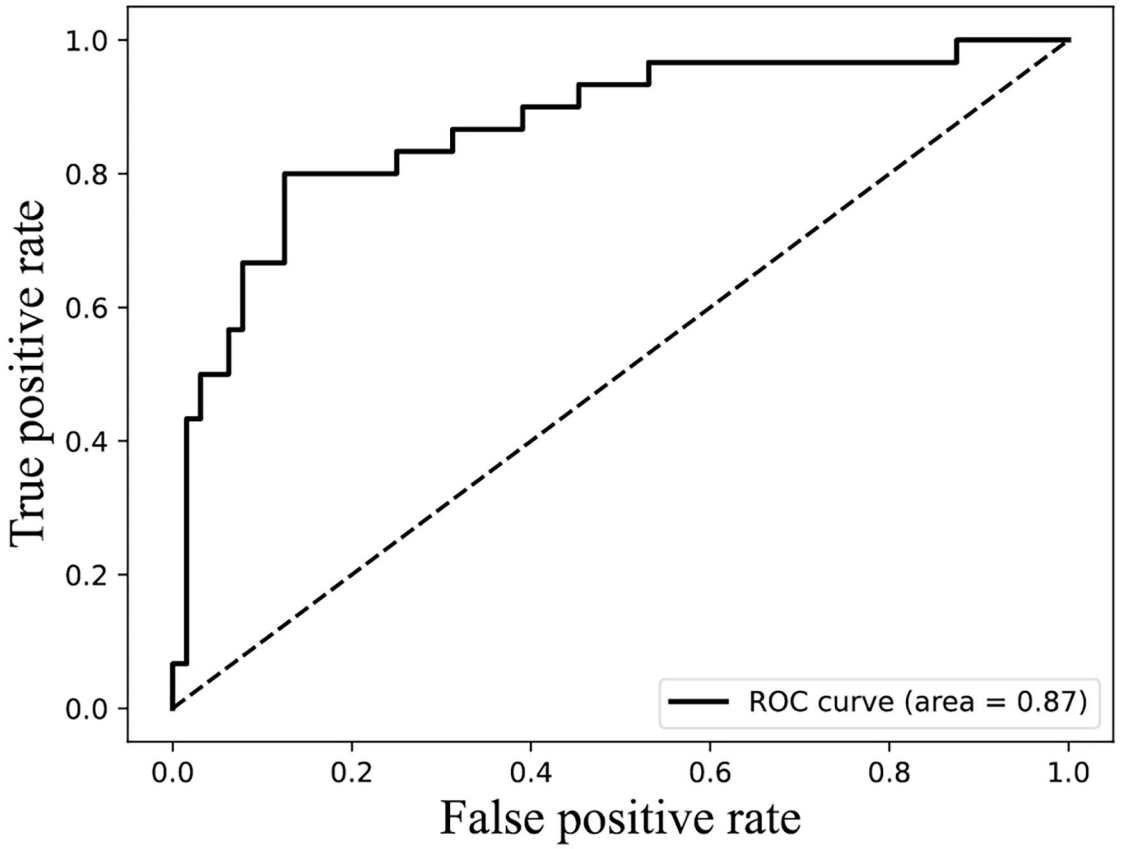


FIGURE 5 | ROC curves of 87 features on independent test data.

non-bacterial protein in these data, especially lysine and valine are the most important residues. The classification accuracy of CKSAAP-1GAP is not as good as that of AAC. Because the number of the feature subsets is 30% more than the total number of training samples, the classification accuracy rate is increased to 86.04% when 103 important feature components are selected, but the classification accuracy rate is only 77.66% when only CKSAAP-1GAP is used in the independent test set. However, the classification results of other individual feature subsets also show that the classification accuracy of only using any individual feature subsets is always low, so we decided to splice different types of feature subsets. To prevent the dimension from being too high, we only selected important features to combine.

For the training set of 223 dimensional features after splicing, we calculated the feature importance again using the feature selection modules shown in **Figure 1B** to calculate the influence of each feature on the results when different feature components were spliced. The features were then further filtered in the training data set using the incremental approach and cross-validation at 10-fold.

The result of **Figure 4** shows that 85% classification accuracy, which is the same classification accuracy as stated in the original article (Ding et al., 2014), can be achieved on the independent test set with 55 feature components selected, and 89.28% on the training set and 86.17% on the independent test with 87 feature components selected. Although the accuracy is 0.65% higher than 87 at the 113-dimension level, the number of feature components is greatly increased. Therefore, we chose 87 features for prediction on the independent test set, and the ROC curve is shown in **Figure 5**.

TABLE 3 | Performance comparison of the different features in independent test sets.

Classifier	Features	SN (%)	SP (%)	ACC (%)
Naïve Bayes	Ding et al., 2014 (38D)	75.76	80.77	79.15
SVM	Ding et al., 2014 (160D)	75.76	89.42	85.02
Bin et al., 2020	Nine feature groups (8D)	50.00	92.19	78.72
MLP	**Concatenation (38D)**	**73.33**	**90.63**	**85.11**
MLP	**Concatenation (87D)**	**73.33**	**93.75**	**87.23**

Bold indicates the result of the processing of the features.

From **Table 1**, it can be seen that the classification accuracy is low using only one type of all features and that using feature selection is significantly higher. As can be seen from **Table 3**, we can achieve the same results with fewer feature components compared to the original article (Ding et al., 2014).

For the 17 selected 1-gap dipeptides (A*G, A*T, A*P, S*T, S*A, V*A, T*S, V*T, G*A, G*G, S*G, V*G, V*I, E*L, K*L, K*E, and

E*E) in the original article (Ding et al., 2014), the order of feature importance is (1, 4, 5, 17, 15, 14, 7, 61, 18, 25, 8, 161, 16, 23, 3, 87, and 54) when determined individually. It can be found that there is a significant overlap between the feature we selected and the original article (Ding et al., 2014). Lin et al. analyzed the amino acid composition of filamentous bacterial virus xf (*Xanthomonas oryzae*) coat protein, which showed that His, Cys, and Phe were absent from the xf protein. This indicates that these three amino acids are not important in phage classification. In AAC features, the rankings of these three amino acids are 7, 9, and 17. It also occupies 1, 0, and 1 amino acids in the top 20 CKSAAPs.

DISCUSSION AND CONCLUSION

By analyzing the selected feature, it can be found that physicochemical properties are important for phage protein identification. In fact, 40 components representing physicochemical properties appear in the 87 features spliced. Of the first 15 features, 12 refer to the physicochemical properties. The most important feature comes from the physicochemical properties. The charge property is the most important, followed by polarity and polarization rate. Besides, the effective physicochemical properties are derived from different feature extraction methods. The contribution of CKSAAP-1GAP to the 87-dimensional feature is limited. Only 9 of CKSAAP-1GAP features are selected, while the secondary structure composed of the physicochemical properties derived from CTD occupies 20. Thus, it can be concluded that CTD makes a greater contribution than CKSAAP-1GAP on the selected feature for the classification of phage protein. As to AAC, its most important components are proline and leucine, which are highly ranked in the 87 features spliced. For RPSSM, its selected features are not top ranked in the 87-dimensional feature, but 30 features appear in the 87 features spliced. The extracted secondary structures using CTD and RPSSM after classifying amino acids in advance according to their properties are highly ranked in the 87 feature components, while the ones derived from CKSAAP-1GAP are not.

In this study, a feature selection framework is proposed for the protein classification of phages. The model improves the classification accuracy of the data by overlaying different types of features. To prevent overfitting caused by high feature dimensionality, the feature importance was quantified and the important features with high scores were selected as the final feature for classification. We performed ensemble learning using different classifiers, which are insensitive to the distribution of the original data, quantified the importance of each feature, and then performed feature selection on it. Finally, only an 87-dimensional feature was used to achieve a high classification accuracy. Compared with the original article (Ding et al., 2014) and PredNeuroP (Bin et al., 2020), a new method for recognizing phage protein, the model can achieve the same classification accuracy using only 38 feature components. The classification accuracy reaches 87.23% when the optimal 87-dimensional feature is used. It also shows that, when the relationship between the structure and function of phage proteins is not fully understood, the knowledge-driven approach to feature extraction alone does not necessarily lead to better prediction results. In contrast, the prediction of phage proteins by a combination of knowledge-driven and data-driven is more accurate, and the functions of phage proteins can be further investigated by analyzing the selected feature.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SL and CC proposed this research topic. TL completed the experiment and wrote the manuscript. HC supervised the experimental process and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

Ahmad, K., Waris, M., and Hayat, M. (2016). Prediction of protein submitochondrial locations by incorporating dipeptide composition into Chou's general pseudo amino acid composition. *J. Membr. Biol.* 249, 293–304. doi: 10.1007/s00232-015-9868-8

Bin, Y., Zhang, W., Tang, W., Dai, R., Li, M., Zhu, Q., et al. (2020). Prediction of neuropeptides from sequence information using ensemble classifier and hybrid features. *J. Proteome Res.* 19, 3732–3740. doi: 10.1021/acs.jproteome.0c00276

Cai, C., Han, L., Ji, Z. L., Chen, X., and Chen, Y. Z. (2003). SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res.* 31, 3692–3697. doi: 10.1093/nar/gkg600

Chou, K.-C. (2009). Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. *Curr. Proteomics* 6, 262–274. doi: 10.2174/157016409789973707

Clokie, M. R., Millard, A. D., Letarov, A. V., and Heaphy, S. (2011). Phages in nature. *Bacteriophage* 1, 31–45. doi: 10.4161/bact.1.1.14942

Cui, X., Yu, Z., Yu, B., Wang, M., Tian, B., and Ma, Q. (2019). UbiSitePred: a novel method for improving the accuracy of ubiquitination sites prediction by using LASSO to select the optimal Chou's pseudo components. *Chemometr. Intell. Lab. Syst.* 184, 28–43. doi: 10.1016/j.chemolab.2018.11.012

Ding, H., Feng, P.-M., Chen, W., and Lin, H. (2014). Identification of bacteriophage virion proteins by the ANOVA feature selection and analysis. *Mol. Biosyst.* 10, 2229–2235. doi: 10.1039/C4MB00316K

Ding, Y., Tang, J., and Guo, F. (2020). Human protein subcellular localization identification via fuzzy model on kernelized neighborhood representation. *Appl. Soft Comput.* 96, 106596. doi: 10.1016/j.asoc.2020.106596

Feng, P.-M., Ding, H., Chen, W., and Lin, H. (2013). Naive Bayes classifier with feature selection to identify phage virion proteins. *Comput. Math. Methods Med.* 2013. doi: 10.1155/2013/530696

Fu, X., Cai, L., Zeng, X., and Zou, Q. (2020). StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency. *Bioinformatics* 36, 3028–3034. doi: 10.1093/bioinformatics/btaa131

Guo, Z., Wang, P., Liu, Z., and Zhao, Y. (2020). Discrimination of thermophilic proteins and non-thermophilic proteins using feature dimension reduction. *Front. Bioeng. Biotechnol.* 8, 584807. doi: 10.3389/fbioe.2020.584807

Jahn, M. T., Arkhipova, K., Markert, S. M., Stigloher, C., Lachnit, T., Pita, L., et al. (2019). A phage protein aids bacterial symbionts in eukaryote immune

evasion. *Cell Host Microbe* 26, 542–550.e545. doi: 10.1016/j.chom.2019.08.019

Jara-Acevedo, R., Díez, P., González-González, M., Dégano, R. M., Ibarrola, N., Góngora, R., et al. (2018). "Screening phage-display antibody libraries using protein arrays," in *Phage Display*. (Totowa, NJ: Springer), 365–380. doi: 10.1007/978-1-4939-7447-4_20

Ji, Z., Meng, G., Huang, D., Yue, X., and Wang, B. (2015). NMFBFS: a NMF-based feature selection method in identifying pivotal clinical symptoms of hepatocellular carcinoma. *Comput. Math. Methods Med.* 2015, 846942. doi: 10.1155/2015/846942

Jiang, M., Zhao, B., Luo, S., Wang, Q., Chu, Y., Chen, T., et al. (2021). NeuroPpred-Fuse: an interpretable stacking model for prediction of neuropeptides by fusing sequence information and feature selection methods. *Brief. Bioinformatics* 22, bbab310. doi: 10.1093/bib/bbab310

Jiao, Y.-S., and Du, P.-F. (2017). Predicting protein submitochondrial locations by incorporating the positional-specific physicochemical properties into Chou's general pseudo-amino acid compositions. *J. Theor. Biol.* 416, 81–87. doi: 10.1016/j.jtbi.2016.12.026

Jin, S., Zeng, X., Xia, F., Huang, W., and Liu, X. (2021). Application of deep learning methods in biological networks. *Brief. Bioinformatics* 22, 1902–1917. doi: 10.1093/bib/bbaa043

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091

Kawashima, S., and Kanehisa, M. (2000). AAindex: amino acid index database. *Nucleic Acids Res.* 28, 374–374. doi: 10.1093/nar/28.1.374

Khan, M., Hayat, M., Khan, S. A., Ahmad, S., and Iqbal, N. (2017). Bi-PSSM: position specific scoring matrix based intelligent computational model for identification of mycobacterial membrane proteins. *J. Theor. Biol.* 435, 116–124. doi: 10.1016/j.jtbi.2017.09.013

Lavigne, R., Ceyssens, P.-J., and Robben, J. (2009). "Phage proteomics: applications of mass spectrometry," in *Bacteriophages*. Totowa, NJ: Humana Press, 239–251. doi: 10.1007/978-1-60327-565-1_14

Lekunberri, I., Subirats, J., Borrego, C. M., and Balcázar, J. L. (2017). Exploring the contribution of bacteriophages to antibiotic resistance. *Environ. Pollut.* 220, 981–984. doi: 10.1016/j.envpol.2016.11.059

Li, L., Yu, S., Xiao, W., Li, Y., Hu, W., Huang, L., et al. (2015). Protein submitochondrial localization from integrated sequence representation and SVM-based backward feature extraction. *Mol. Biosyst.* 11, 170–177. doi: 10.1039/C4MB00340C

Li, T., Fan, K., Wang, J., and Wang, W. (2003). Reduction of protein sequence complexity by residue grouping. *Protein Eng.* 16, 323–330. doi: 10.1093/protein/gzg044

Liu, B., Gao, X., and Zhang, H. (2019). BioSeq-Analysis2. 0: an updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches. *Nucleic Acids Res.* 47, e127–e127. doi: 10.1093/nar/gkz740

Mei, S. (2012). Multi-kernel transfer learning based on Chou's PseAAC formulation for protein submitochondria localization. *J. Theor. Biol.* 293, 121–130. doi: 10.1016/j.jtbi.2011.10.015

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195

Silvério-Machado, R., Couto, B. R., and Dos Santos, M. A. (2015). Retrieval of Enterobacteriaceae drug targets using singular value decomposition. *Bioinformatics* 31, 1267–1273. doi: 10.1093/bioinformatics/btu792

Wen, P.-P., Shi, S.-P., Xu, H.-D., Wang, L.-N., and Qiu, J.-D. (2016). Accurate in silico prediction of species-specific methylation sites based on information gain feature optimization. *Bioinformatics* 32, 3107–3115. doi: 10.1093/bioinformatics/btw377

Xie, X., Gu, X., Li, Y., and Ji, Z. (2021). K-size partial reduct: positive region optimization for attribute reduction. *Knowl. Based Syst.* 228, 107253. doi: 10.1016/j.knosys.2021.107253

Xiong, Y., Wang, Q., Yang, J., Zhu, X., and Wei, D.-Q. (2018). PredT4SE-stack: prediction of bacterial type IV secreted effectors from protein sequences using a stacked ensemble method. *Front. Microbiol.* 9, 2571. doi: 10.3389/fmicb.2018.02571

Xu, Y., Verma, D., Sheridan, R. P., Liaw, A., Ma, J., Marshall, N. M., et al. (2020). Deep dive into machine learning models for protein engineering. *J. Chem. Inf. Model.* 60, 2773–2790. doi: 10.1021/acs.jcim.0c00073

Yu, B., Li, S., Qiu, W., Wang, M., Du, J., Zhang, Y., et al. (2018). Prediction of subcellular location of apoptosis proteins by incorporating PsePSSM and DCCA coefficient based on LFDA dimensionality reduction. *BMC Genomics* 19, 1–17. doi: 10.1186/s12864-018-4849-9

Yu, L., Wang, M., Yang, Y., Xu, F., Zhang, X., Xie, F., et al. (2021). Predicting therapeutic drugs for hepatocellular carcinoma based on tissue-specific pathways. *PLoS Comput. Biol.* 17, e1008696. doi: 10.1371/journal.pcbi.1008696

Yuan, Y., and Gao, M. (2016). Proteomic analysis of a novel bacillus jumbo phage revealing glycoside hydrolase as structural component. *Front. Microbiol.* 7, 745. doi: 10.3389/fmicb.2016.00745

Zhang, L., Zhang, C., Gao, R., and Yang, R. (2015). An ensemble method to distinguish bacteriophage virion from non-virion proteins based on protein sequence characteristics. *Int. J. Mol. Sci.* 16, 21734–21758. doi: 10.3390/ijms160921734

Zhao, X., Wang, H., Li, H., Wu, Y., and Wang, G. (2021). Identifying plant pentatricopeptide repeat proteins using a variable selection method. *Front. Plant Sci.* 12, 298. doi: 10.3389/fpls.2021.506681

Zou, Q., Zeng, J., Cao, L., and Ji, R. (2016). A novel features ranking metric with application to scalable visual and bioinformatics data classification. *Neurocomputing* 173, 346–354. doi: 10.1016/j.neucom.2014.12.123

Zulfiqar, H., Yuan, S.-S., Huang, Q.-L., Sun, Z.-J., Dao, F.-Y., Yu, X.-L., et al. (2021). Identification of cyclin protein using gradient boost decision tree algorithm. *Comput. Struct. Biotechnol. J.* 19, 4123–4131. doi: 10.1016/j.csbj.2021.07.013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 22 July 2022
DOI 10.3389/fmicb.2022.933249

OPEN ACCESS

EDITED BY
Beibei Ru,
National Institutes of Health (NIH),
United States

REVIEWED BY
Changlong Liu,
Shanghai Veterinary Research Institute (CAAS), China
Yusuke Matsumoto,
Tokyo Metropolitan Institute of Medical Science, Japan

*CORRESPONDENCE
Nitat Sookrung
nitat.soo@mahidol.ac.th

SPECIALTY SECTION
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

RECEIVED 30 April 2022
ACCEPTED 28 June 2022
PUBLISHED 22 July 2022

CITATION
Thavorasak T, Chulanetra M, Glab-ampai K, Mahasongkram K, Sae-lim N, Teeranitayatarn K, Songserm T, Yodsheewan R, Nilubol D, Chaicumpa W and Sookrung N (2022) Enhancing epitope of PEDV spike protein. *Front. Microbiol.* 13:933249. doi: 10.3389/fmicb.2022.933249

Enhancing epitope of PEDV spike protein

Techit Thavorasak[1,2], Monrat Chulanetra[2], Kittirat Glab-ampai[2], Kodchakorn Mahasongkram[2], Nawannaporn Sae-lim[2], Karsidete Teeranitayatarn[3], Thaweesak Songserm[4], Rungrueang Yodsheewan[4], Dachrit Nilubol[5,6], Wanpen Chaicumpa[2] and Nitat Sookrung[2,7]*

[1]Graduate Program in Immunology, Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, [2]Center of Research Excellence in Therapeutic Proteins and Antibody Engineering, Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, [3]MORENA Solution Company, Bangkok, Thailand, [4]Department of Veterinary Pathology, Faculty of Veterinary Medicine, Kasetsart University, Bangkok, Thailand, [5]Department of Veterinary Microbiology, Faculty of Veterinary Science, Chulalongkorn University, Bangkok, Thailand, [6]Swine Viral Evolution and Vaccine Development Research Unit, Chulalongkorn University, Bangkok, Thailand, [7]Biomedical Research Incubation Unit, Department of Research, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

Porcine epidemic diarrhea virus (PEDV) is the causative agent of a highly contagious enteric disease of pigs characterized by diarrhea, vomiting, and severe dehydration. PEDV infects pigs of all ages, but neonatal pigs during the first week of life are highly susceptible; the mortality rates among newborn piglets may reach 80–100%. Thus, PEDV is regarded as one of the most devastating pig viruses that cause huge economic damage to pig industries worldwide. Vaccination of sows and gilts at the pre-fertilization or pre-farrowing stage is a good strategy for the protection of suckling piglets against PEDV through the acquisition of the lactating immunity. However, vaccination of the mother pigs for inducing a high level of virus-neutralizing antibodies is complicated with unstandardized immunization protocol and unreliable outcomes. Besides, the vaccine may also induce enhancing antibodies that promote virus entry and replication, so-called antibody-dependent enhancement (ADE), which aggravates the disease upon new virus exposure. Recognition of the virus epitope that induces the production of the enhancing antibodies is an existential necessity for safe and effective PEDV vaccine design. In this study, the enhancing epitope of the PEDV spike (S) protein was revealed for the first time, by using phage display technology and mouse monoclonal antibody (mAbG3) that bound to the PEDV S1 subunit of the S protein and enhanced PEDV entry into permissive Vero cells that lack Fc receptor. The phages displaying mAbG3-bound peptides derived from the phage library by panning with the mAbG3 matched with several regions in the S1-0 sub-domain of the PEDV S1 subunit, indicating that the epitope is discontinuous (conformational). The mAbG3-bound phage sequence also matched with a linear sequence of the S1-BCD sub-domains. Immunological assays verified the phage mimotope results. Although the molecular mechanism of ADE caused by the mAbG3 *via* binding to the newly identified S1 enhancing epitope awaits investigation, the data obtained from this study are helpful and useful in designing a safe and effective PEDV protein subunit/DNA vaccine devoid of the enhancing epitope.

KEYWORDS

antibody-dependent enhancement, enhancing epitope, phage display technology, phage panning, phage mimotope, porcine epidemic diarrhea, porcine epidemic diarrhea virus, spike protein

Introduction

Porcine epidemic diarrhea virus (PEDV) is an etiologic agent of a highly contagious disease of pigs named porcine epidemic diarrhea (PED) which is characterized by acute diarrhea, vomiting, and severe dehydration (Lee, 2015). The virus can infect pigs of all ages, but the disease is highly fatal among neonatal pigs during the first 7–10 days of lives and the mortality rate may reach up to 80–100% (Pensaert and de Bouck, 1978; Lee, 2015; Jung et al., 2020). PEDV-infected neonatal piglet manifests acute viremia and severe atrophic enteritis (mainly jejunum and ileum), with increased pro-inflammatory and innate immune responses (Annamalai et al., 2015; Jung et al., 2018). The PEDV is shed in the watery stool and nasal discharge of the infected pigs and spreads further (Jung et al., 2020). Pigs are more tolerable to PED as they grow older, but asymptomatically infected older pigs on a farm may serve as the virus reservoirs for the subsequent outbreaks (Wang et al., 2019). PED was first recognized in 1971 in England; the disease subsequently spread to other European countries, North and South Americas, and Asia (Lee, 2015; Jung et al., 2020). The virus is now regarded as one of the most devastating pig viruses causing severe economic damage to pig industries worldwide.

Porcine epidemic diarrhea virus is a large, enveloped, plus-sense RNA virus of the genus *Alphacoronavirus*, family *Coronaviridae*, and order *Nidovirales* (Lee, 2015; Jung et al., 2020). The PEDV genome is approximately 28 kb long and consists of a 5′-untranslated region (UTR) with a cap, followed by at least seven open reading frames (ORF1a, ORF1b, and ORFs 2–6), and a 3′-polyadenylated tail (Lee, 2015). The ORFs 1a and 1b that occupy two-thirds of the genome at the 5′ end code for multifunctional polyproteins (pp) la and pp1ab, which are further post-translationally cleaved by the *cis* and *trans* proteases of the virus to generate 16 functionally different non-structural proteins (nsps), nsp1–16. The mature nsps form replicase/transcriptase complex (RTC) to generate full-length genomic RNA and sub-genomic (sg) mRNAs from the remaining ORFs that constitute one-third of the genome at the 3′ end. The sg mRNAs are translated into four structural proteins, including spike (S) protein which is post-translationally glycosylated (150–220 kDa), membrane (M) protein (20–30 kDa), envelope (E) protein (7 kDa), and nucleocapsid (N) protein (58 kDa), and accessory gene ORF3 (Li et al., 2020).

The PEDV uses the S1 subunit located at the N-terminal end of the surface-exposed spike (S) glycoprotein to bind to the receptors on the pig enterocyte, followed by virus–host membrane fusion mediated by the S2 subunit at the C-terminal end of the S ectodomain to release the virus genome into the cytosol for further replication (Lee, 2015). The S1 subunit is composed of the leader peptide sequence (residues 1–18) and sub-domains 0 (residues 19–219), A (residues 220–509), B (residues 510–639), and CD (residues 640–728); the S2 subunit starts at residue 729 and extends up to the C-terminal end (Li et al., 2017). Heterogeneity in the genomic sequences of the PEDV S protein has been reported and used for the classification of isolated strains (Lin et al., 2016). Currently, PEDV isolates are classified into two genotypes (GI and GII) and five sub-genotypes (GIa, GIb, GIIa, GIIb, and GIIc; Guo et al., 2019). The GI is represented by classical PEDV strain CV777 and the CV777-like strains that have emerged since the 1970s, while the GII includes highly virulent strains (field isolates) detected globally after 2010 (Lin et al., 2016). The emerging PEDV strains are also divided into two major groups based on the S gene features and their virulence in young pigs, i.e., the highly virulent non-S INDEL strains (insertions and deletions) that cause pandemic PED outbreaks worldwide and the less pathogenic (cause lower mortality) S INDEL strains [the N-terminal region of the S protein of these strains has an amino acid insertion (residues 161–162) and two deletions (residues 59–62 and 140) compared with those of the highly virulent PEDV strains] (Wang et al., 2014; Lin et al., 2016).

Vaccination against PEDV is a useful strategy to control the disease regardless of the type of immunogens used in the vaccine (killed, live, recombinant subunit, and DNA-vectored) and administration route (intramuscular versus oral; Won et al., 2020). However, the vaccine must be given in multiple, spaced doses to sows or gilts either at pre-fertilization or pre-farrowing stage to induce high levels of neutralizing antibodies (mainly directed toward the PEDV S protein to block the virus entry into the host cells) in colostrum and milk for conferring lactogenic immunity to the suckling newborns. The immunization protocols of the vaccines are complicated, and the protective efficacies among the passively immunized suckling piglets are inconsistent. Some studies obtained satisfactory results including reduced morbidity and mortality, shortened period of virus shedding in the feces, and quick recovery of daily weight gain, while other studies did not (Gerdts and Zakhartchouk, 2017). One of the major obstacles to vaccination against PEDV and other coronaviruses is the vaccine-induced antibody-dependent enhancement (ADE) of new infection (Wen et al., 2020; Xu et al., 2021; Yu et al., 2021). For safety issues, the vaccine should not contain epitopes that elicit enhancing antibodies (Xu et al., 2021; Yu et al., 2021). Several neutralizing B-cell epitopes of the PEDV spike protein have been identified (Thavorasak et al., 2022). On contrary, data on enhancing epitopes, which is the dark side of PEDV protein, are scarce, if there were any. In this study, a novel B-cell epitope of the PEDV S1 protein that enhanced the virus entry into the host cells is reported.

Materials and methods

Cells and virus

African green monkey kidney epithelial (Vero) cells were obtained from American Type Culture Collection (ATCC, Manassas, VA, United States). The cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific, Waltham, MA, United States) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich, St. Louis, MO, United States), 2 mM L-alanyl-L-glutaminase dipeptide, 100 IU/ml penicillin, and 100 µg/ml streptomycin (Thermo Fisher Scientific).

Mouse hybridoma G3 secreting monoclonal antibody (mAbG3; IgG1-kappa isotype) that bound to the S1 protein of the PEDV was generated previously (Thavorasak et al., 2022). The cells were nurtured at 37°C in a humidified atmosphere containing 5% CO_2 in an incubator in a serum-free medium [CD medium supplemented with 8 mM L-alanyl-L-glutaminase dipeptide, 100 IU/ml penicillin, and 100 µg/ml streptomycin (Thermo Fisher Scientific)].

The PEDV (GII variant, strain P70, and isolated from PEDV-infected piglet in Thailand) was propagated in the Vero cells cultured in virus maintaining medium [DMEM supplemented with 2 µg/ml of *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK)-treated trypsin (Sigma-Aldrich)]. Vero cell monolayers at ~80–100% confluent growth in the T175 culture flask were rinsed with sterile phosphate-buffered saline, pH 7.4 (PBS); the PEDV was added to the cells at MOI 0.001 and incubated at 37°C in a humidified atmosphere containing 5% CO_2 in an incubator for 1 h. The fluid was removed, the cells were rinsed twice with PBS, replenished with the virus maintaining medium, and incubated at 37°C in humidified 5% CO_2 atmosphere in an incubator. When the cytopathic effect (CPE; syncytial formation) reached the maximum level (approximately 48 h), the culture supernatant containing the PEDV was harvested, and centrifuged at 4500 × *g* and 4°C for 30 min to remove cells and debris. The clarified culture supernatant was filtered through a 0.2 µm Acrodisc® syringe filter (Pall, Port Washington, NY, United States) and stored in small portions at −80°C until further use. The virus titer was determined by plaque-forming assay (PFA) and reported as plaque-forming units (pfu)/ml.

Preparation of mAbG3

For the production of mouse monoclonal antibody G3 (mAbG3), log phase grown hybridoma G3 cells were seeded into a serum-free medium (CD medium, Thermo Fisher Scientific) at 3×10^5 cells/ml and incubated further at 37°C in humidified 5% CO_2 atmosphere in an incubator for 96 h. The culture supernatant containing the mAbG3 was harvested, and centrifuged at 4500 × *g* and 4°C for 30 min to remove the cells and cell debris. The clarified supernatant was subjected to Hitrap protein G HP (Cytiva, Uppsala, Sweden) chromatography for the purification of the mAbG3. The protein content of the purified mAbG3 was quantified using the BCA assay (Thermo Fisher Scientific).

Virus neutralization assay

The neutralization assay was performed as described previously (Thavorasak et al., 2022). Purified mAbG3 (5–40 µg in 125 µl of PBS) was mixed with 100 pfu of PEDV P70 and incubated at 37°C for 1 h. Immune serum to PEDV S1 subunit (1:100), mAbA3 (Thavorasak et al., 2022), isotype-matched control mAb (IgG1-kappa; BioLegend, San Diego, CA, United States), and culture medium alone were included in the experiments as controls. The mixtures were then added to the confluent monolayers of Vero cells in a 24-well tissue culture plate. After 1 h of incubation, the supernatant was discarded, and the cells were rinsed for two times to remove the unbound virus. Carboxymethyl cellulose (2% CMC in DMEM supplemented with 2 µg of TPCK-treated trypsin) was added. At 48 h post-infection, the cells were fixed with 10% formalin, kept at room temperature (25 ± 2°C) for 1 h, and stained with 1% crystal violet in 10% ethanol. The CPE (syncytial formation) was counted under a light microscope at ×40 magnification. The percentage of mAbG3-mediated PEDV neutralization/enhancement was calculated: [1-(Average CPE count of test/Average CPE count of non-neutralization control)] × 100.

Real-time (quantitative) RT-PCR (qRT-PCR)

Alternatively, the PEDV RNAs recovered from the treated infected cells were determined by qRT-PCR. The virus neutralization assay was performed as described above with a slight modification. The DMEM supplemented with 2 µg of TPCK-treated trypsin was used instead of 2% CMC in DMEM supplemented with 2 µg of TPCK-treated trypsin. After 6 h post-infection, total RNA from the mAbG3/control-treated infected cells was extracted separately using TRIzol® reagent (Thermo Fisher Scientific). The amount of viral RNA was quantified by qRT-PCR using a one-step brilliant III SYBR green RT-qPCR master mix (Agilent Technologies, Santa Clara, CA, United States). The qRT-PCR primers for the PEDV nucleocapsid (N) gene and house-keeping gene control are listed in Supplementary Table S1. The PCR reaction mixture consisted of 6 µl of nuclease-free water, 10 µl of 2× SYBR Green QPCR master mix, 0.4 µl of 10 µM of each specific primer, 0.2 µl of 100 mM dithiothreitol (DTT), 1 µl of RT/RNase block, and 2 µl of RNA template. The conditions for PCR thermal cycles were 50°C for 10 min, 95°C for 3 min, 40 cycles at 95°C for 5 s, and 58°C for 10s, followed by melt curve analysis protocol. The copy numbers of viruses were calculated from Cq values and compared.

Production of S1 subunit of PEDV spike protein and the S1 subunit truncated polypeptides

The full-length S1 subunit of the PEDV was prepared as described previously (Thavorasak et al., 2022). For preparing the S1 subunit truncated polypeptides, i.e., S1-0 (residues M1-V217), S1-A (residues T218-N508), and S1-BCD (residues D509-T728; Li et al.,

2017), total RNA was isolated from culture supernatant containing PEDV using TRIzol reagent (Thermo Fisher Scientific). The isolated RNA was subjected to RevertAid First Strand cDNA Synthesis Kit using a reverse primer specific to the PEDV E gene for synthesizing the cDNA. The cDNA was then used as a template for the amplification of DNA sequences coding for the S1-0, S1-A, and S1-BCD polypeptides using specific primers listed in Supplementary Table S2. The amplified DNA sequences were ligated separately to pJET1.2 plasmids, and the recombinant plasmids were transformed into DH5α *E. coli*. The transformed DH5α *E. coli* colonies with the respective DNA inserts were cultured in Luria-Bertani (LB) broth supplemented with 100 μg/ml ampicillin (LB-A) at 37°C with shaking aeration (250 rpm) and incubated overnight. The recombinant plasmids were then extracted from the DH5α *E. coli* and subcloned into pTriEx1.1 Hygro DNA (Merck KGaA, Darmstadt, Germany) *via Bam*HI and *Xho*I restriction sites. The ligated plasmids were transformed into NiCo21 (DE3) *E. coli*. Transformed NiCo21 (DE3) *E. coli* clones that carried the correct DNA inserted sequences (verified by Sanger sequencing) were grown in LB-A broth at 37°C with shaking aeration until OD at 600 nm was 0.4–0.5. Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to individual cultures to 0.1 mM final concentration, and the cultures were incubated further for 3 h. The 8× His-tagged recombinant polypeptides were purified from the bacterial homogenates by using Ni-NTA affinity resin (Thermo Fisher Scientific) under denaturing conditions.

SDS-PAGE and Western blot analysis

One microgram each of the full-length recombinant S1 (rS1), S1-0, S1-A, and S1-BCD were prepared in protein loading buffer and subjected to 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The separated proteins were blotted onto the nitrocellulose membrane (NC). The empty sites of blotted NC were blocked with 5% skim milk in Tris-buffered saline containing 0.1% Tween-20 (TBST) at room temperature for 1 h. After washing with TBST, the NC was immersed into a solution of either mouse anti-6× His antibody or purified mAbG3 solution (5 μg in 6 ml TBST) and kept at room temperature for 1 h. The membrane was washed with the TBST and placed in a solution of goat anti-mouse Ig-alkaline phosphatase (AP) conjugate (1:3,000; Southern Biotech, Birmingham, AL, United States) in TBST at room temperature for 1 h. After washing, BCIP/NBT substrate (KPL, Gaithersburg, MD, United States) was used for the visualization of the antigen–antibody reactive bands. The enzymatic reaction was stopped by rinsing the NC with distilled water.

ELISAs for determining binding of mAbG3 to the S1 polypeptides

For indirect ELISA, individual wells of 96-well microplate (Nunc MaxiSorp™, Thermo Fisher Scientific) were coated with 1 μg of purified S1, S1-0, S1-A, and S1-BCD in 100 μl of PBS. The plate was kept at 4°C overnight. The supernatants were discarded, and the empty sites of the well surface were blocked with 1% BSA in PBS containing 0.1% Tween-20 (PBST) at 37°C for 1 h. One hundred microliters of PBST containing 1 μg of mAbG3 was added to each antigen-coated well and incubated at 37°C for 1 h. The wells were washed with PBST, added 100 μl of goat anti-mouse Ig-horseradish peroxidase conjugate (HRP; 1:5,000; Southern Biotech) in PBST, and kept at 37°C for 1 h. The fluids were discarded; all wells were washed with PBST, added 100 μl of ABTS substrate (KPL) individually, and the plate was kept at room temperature in darkness for 30 min. The OD of the content of each well was measured at 405 nm against blank (reactants without antibody) by spectrophotometry (BioTek, Winooski, VT, United States). Mouse anti-6× His antibody and PBS (instead of mAbG3) were included as positive and negative controls, respectively.

For dot ELISA, 1 μg of purified S1-0, S1-A, and S1-BCD were dotted separately onto the NC strip by using the Bio-Dot® microfiltration apparatus (Bio-Rad). The membranes were dried for 30 min and blocked with 5% skim milk in TBST for 1 h at room temperature. After washing with TBST, 5 μg of mAbG3 in 6 ml of TBST was added and kept at room temperature for 1 h. Mouse anti-6× His antibody (1:1,000; Bio-Rad) and PBS were included as positive and negative controls, respectively. After washing, the membranes were immersed in goat anti-mouse Ig-alkaline phosphatase conjugate (1:3,000; Southern Biotech) in TBST at room temperature for 1 h. After washing, BCIP/NBT substrate (KPL) was used to visualize the reactive dots on membranes.

Identification of phage mimotopes for determination of the PEDV S1 sub-domains bound by the mAbG3

To identify the region (sub-domains) of the PEDV S1 subunit that was bound by the mAbG3, Ph.D.™- 12 Phage Display Peptide Library (New England Biolab, Ipswich, MA, United States) was used for panning with the mAbG3. For this experiment, 1 μg of purified mAbG3 was added to a well of 96-well microplate (Nunc MaxiSorp™) and kept at 4°C overnight. The empty sites of the well surface were blocked with 1% BSA in TBST. After washing, a diluted phage library (~2×10^{11} peptide displaying phages in 100 μl of 0.5% BSA in TBST) was added to the mAbG3 coated well and incubated at room temperature for 1 h. The unbound phages were removed by washing thoroughly with TBST. The bound phages were eluted with 0.2 M glycine HCl (pH 2.2) for 10 min and immediately neutralized with 1 M Tris–HCl (pH 9.1). The phages were added to an early log phase grown ER2738 *E. coil* to amplify the phages and kept at 37°C for 4.5 h. The culture supernatant was harvested after centrifugation at 12,000× *g* and 4°C for 15 min The clear supernatant was mixed with 1/6 volume of 20% (w/v) PEG8000 and 2.5 M NaCl to precipitate and concentrate the phages. The concentrated phages were titrated on LB/IPTG/X-gal plates and used for the second- and third-round phage panning. The eluted phages of the third-round

panning were diluted in LB broth (10^1–10^4), inoculated into mid-log phase grown ER2738 *E. coli*, mixed with prewarmed (45°C) top agar, and poured onto LB/IPTG/X-gal agar plates. The plates were incubated at 37°C for 16 h. Ten single blue colonies were randomly picked and inoculated into mid-log phase grown ER2738 *E. coli* for phage amplification. The *E. coli* culture supernatants were collected for phage DNA isolation and sequencing. The DNA sequences coding for the peptides displayed by the mAbG3-bound phages were deduced by using QIAGEN CLC workbench program (Version 20.0.4[1]; accessed on 11 March 2022). The phage mimotopes (M types) were then multiply aligned with the monomeric S1 sequence of PEDV GI classical strain CV777 and GII P70 strain by using Jalview software (Version 2.11.2.2; Waterhouse et al., 2009) to determine the region of the S1 subunit that matched with the mimotope sequences, i.e., the mAbG3 epitope.

Peptide binding ELISA

To identify the enhancing epitope of the mAbG3 on the PEDV S1 subunit, peptide binding ELISA was performed. Five biotin-labeled 12 amino acid peptides of the PEDV S1 subunit that matched with the mAbG3-bound phage mimotopes and a control peptide were synthesized (Genscript, Piscataway, NJ, United States). Individual peptides were diluted in PBS to 10 μg/ml, and 100 μl of the diluted peptides were added to the separate wells (triplicate) of the streptavidin-coated microplate (Pierce™ Streptavidin Coated HighCapacity Plates; Thermo Fisher Scientific), and the plate was kept at 4°C overnight. The plate was washed three times with TBST, and the empty sites of the well surface were blocked with 5% skim milk in TBST. After 1 h at 37°C, the fluids were discarded. The wells were washed again, and 1 μg of purified mAbG3 in 100 μl of TBST was added to the individual wells. Buffer was included as a negative antibody control. The plate was incubated at 37°C for 1 h. After washing, 100 μl of HRP-labeled mouse IgG-kappa binding protein (Santa Cruz Biotechnology, Dallas, TX, United States; 1:1,000 in TBST) was added to each well, and the plate was incubated further at 37°C for 1 h. After washing, ABTS substrate (KPL; 100 μl) was added to each well, and the plate was kept in darkness at room temperature for 30 min. The OD of individual wells was measured at 405 nm (BioTek spectrometer).

Multiple sequence alignment

Amino acid sequences of the S1 subunits of the PEDV strains that have been submitted to the National Center of Biotechnology Information (NCBI) database within the past 5 years, including PEDV SD2021 (GenBank: UJZ92254.1), PEDV TRS2021 (GenBank: UJZ92268.1), TW/PT01/2020 (GenBank: UOK15705.1), SC-YB73 (GenBank: QNL15619.1), XJ1904-34 (GenBank: UGN13709.1), PEDV-H18-Barcelona-Vic (GenBank: QKV43853.1), PEDV-2118-1-Orense-Covelas (GenBank: QKV43823.1), CT P10 (GenBank: QHB92364.1), EdoMex/205/2018 (GenBank: QQK84868.1); and 0100/5P (GenBank: UAJ21031.1), as well as CV777 (classical strain, GenBank: AEX92968.1), USA/Colorado/2013 (prototypic strain used for producing baculovirus-derived S subunit vaccine that caused severe PEDV enhancement in immunized piglets in a challenge study, Yu et al., 2021; GenBank: KF272920.1), and our P70 strain, were multiply aligned and compared.

Localization of the mAbG3 epitope on the PEDV spike protein

To locate the S1 epitope that bound to mAbG3, a cryoelectron structure of PEDV spike protein (PDB: 6vv5; Kirchdoerfer et al., 2021) was subjected to PyMOL software (Version 4.6; The PyMOL Molecular Graphics System, Schrödinger LLC, New York, NY, United States) for generating the spike protein structure, and the locations of the mAbG3 epitope were mapped on the S protein structure.

Statistical analysis

Statistical analysis of data was performed using GraphPad Prism software (Version 9.2; San Diego). Data were analyzed by using one-way ANOVA and were considered significantly different when $p < 0.05$.

Results

The mAbG3-mediated enhancement of PEDV infectivity instead of neutralization

About 5, 10, 20, and 40 μg concentrations of purified mAbG3 were mixed with PEDV GII field isolated P70 strain and incubated at 37°C for 1 h before adding to the confluent monolayers of Vero cells. After allowing the virus adsorption for 1 h, the fluids were removed, and the cells were rinsed two times before adding 2% CMC in 500 μl of DMEM supplemented with 2 μg/ml of TPCK-treated trypsin. Two days post-infection, the cells were fixed and stained with 10% formalin and 1% crystal violet in 10% ethanol, respectively. Mouse immune serum to PEDV S1 subunit (1:100), 40 μg of mAbA3 (Thavorasak et al., 2022), 40 μg of isotype-matched control mAb (IgG1-kappa), and culture medium alone were included in the experiments as controls. Cytopathic effect (CPE) was determined under light microscopy at ×40 magnification. It was found that instead of mediating neutralization of the PEDV infectivity, on contrary, the mAbG3 at 40 μg concentration significantly enhanced the PEDV infectivity by increasing CPE (the numbers of syncytia) compared to the medium ($p < 0.0001$) and the isotype control ($p < 0.01$), while the immune serum and the mAbA3 neutralized the

1 https://digitalinsights.qiagen.com/

PEDV infectivity (Figure 1A). The mAbG3 at 5, 10, and 20 μg concentrations showed a trend toward the enhancement of infectivity, but the CPE counts were not significantly different from the medium alone ($p > 0.05$). The results of qRT-PCR conformed to the CPE results (Figure 1B). Examples of the PEDV-mediated CPE in the Vero cell monolayer of different virus treatments are shown in Figure 1C. A comparison of the Vero syncytium to normal Vero cells is shown in Figure 1D. Three independent experiments were performed with reproducibility.

Identification of tentative mAbG3 epitope by phage mimotope search and phage peptide alignment with the PEDV S1 linear sequence

Because the mAbG3 was found to enhance the PEDV infectivity, it was interesting to identify the epitope of this monoclonal antibody for the future design of a safe PEDV subunit/DNA vaccine. The Ph.D.™-12 Phage Display Peptide Library was used as a tool for the identification of the phage peptides that were bound by the mAbG3, and the mAbG3-bound phage peptides were aligned with the PEDV S1 linear sequence to identify the mAbG3-bound S1 sub-domain region, i.e., mAbG3 epitope. After panning the phage library with the mAbG3, the mAbG3-bound phages were propagated; the phage DNAs were isolated, sequenced, and deduced. The deduced amino acid sequences, designated mimotope (M) types, were aligned with monomeric S1 segments of CV777 classical strain (GenBank: AEX92968.1) and PEDV GII variant P70 strain. The results revealed that there were four phage mimotope (M) types (M1-M4) that matched with the residues in the linear S1 subunit, including M1: FFDYMYLPGFAA, M2: SFDWPTHKMFNL, M3: NLYNYMFADLYT, and M4: TDLCHLYNMHGC. The M1 phage mimotope sequence matched with 203AMQYVYEPTYYM214 of S1-0 (designated M1S1–0), M2 sequence matched with 165GITWDNDRVTVF176 of S1-0 (designated M2S1–0), M3 sequence matched with 179KIYHFYFKNDWS190 of S1-0 (designated M3S1–0), and M4 sequence matched with 191RVATKCYNSGGC202 of S1-0 (designated M4S1–0) and 646LDVCTKYTIYGF657 of S1-BCD (designated M4S1-BCD; Figure 2).

Truncated recombinant S1 polypeptides

The truncated polypeptides of the PEDV S1 subunit (S1-0, S1-A, and S1-BCD) were prepared for verification of the S1 regions bound by the mAbG3. Amplicons of the DNA sequences coding for the S1-0, S1-A, and S1-BCD (651, 873, and 660 bp, respectively) are shown in Figure 3A. The recombinant S1 polypeptides with 8× His tag expressed by the transformed NiCo21(DE3) *E. coli* clones were then purified by Ni-NTA affinity resin under denaturing conditions with 8 M urea buffer. The purified polyproteins were then refolded by drop-wise dialysis against PBS. The polypeptides, i.e., S1-0 (residues 1–217; ~26 kDa), S1-A (residues 218–508; ~32 kDa), and S1-BCD (residues 509–728; ~24 kDa) after SDS-PAGE and staining with Coomassie Brilliant Blue G-250 (CBB) dye are shown in Figure 3B. Figure 3C shows Western blot patterns of the SDS-PAGE-separated polypeptides probed with mouse anti-6× His antibody.

Binding of mAbG3 to recombinant polypeptides of the PEDV S1 subunit

The mAbG3 was tested for binding to the purified full-length S1 and S1-0, S1-A, and S1-BCD polypeptides in the indirect ELISA, dot ELISA, and Western blot analysis. The indirect and dot ELISA results showed that the mAbG3 bound to the full-length S1 subunit and the S1-0 and S1-BCD polypeptides but not the S1-A polypeptide (Figures 4A,B). When the mAbG3 was tested for binding to the S1 polypeptides by Western blot analysis under reducing conditions using mouse anti-6× His antibody as positive binding control, the mAbG3 bound to all three S1 sub-domain polypeptides (Figure 4C).

Verification of the mAbG3 epitope by using peptide binding ELISA

Biotin-labeled M1S1–0, M2S1–0, M3S1–0, M4S1–0, M4S1-BCD, and control peptides were commercially synthesized and used as antigens in the indirect ELISA for testing the mAbG3 binding. In the indirect ELISA process, the synthesized peptides were used to coat separate wells of a microplate, and the coated wells were added with the mAbG3. The mAbG3 yielded a strong binding signal to M3S1–0, a significant signal to the M4S1-BCD, negligible signals to M1S1–0, M2S1–0, and M4S1–0, and no signal to control peptide (Figure 4D).

Multiple sequence alignment of S1 proteins of different PEDV isolates

The results revealed that the M3S1–0 and M4S1-BCD that were bound by the mAbG3 are highly conserved (Figure 5). Locations of the mAbG3 epitope on the PEDV trimeric S, monomeric S1-0, and monomeric S1-BCD proteins are shown in Figures 6A–C, respectively.

Discussion

Antibodies are important immunological factors for host defense against virus infections. The antibodies neutralize the virus infectivity by binding to the virus particles and interfering with the virus attachment/binding to the host receptors and/or preventing the uncoating of the viral genome and release into the cytoplasm, thus inhibiting further replication. The antibody-mediated antiviral activities may also involve complement-mediated lysis of the virus

10.3389/fmicb.2022.933249

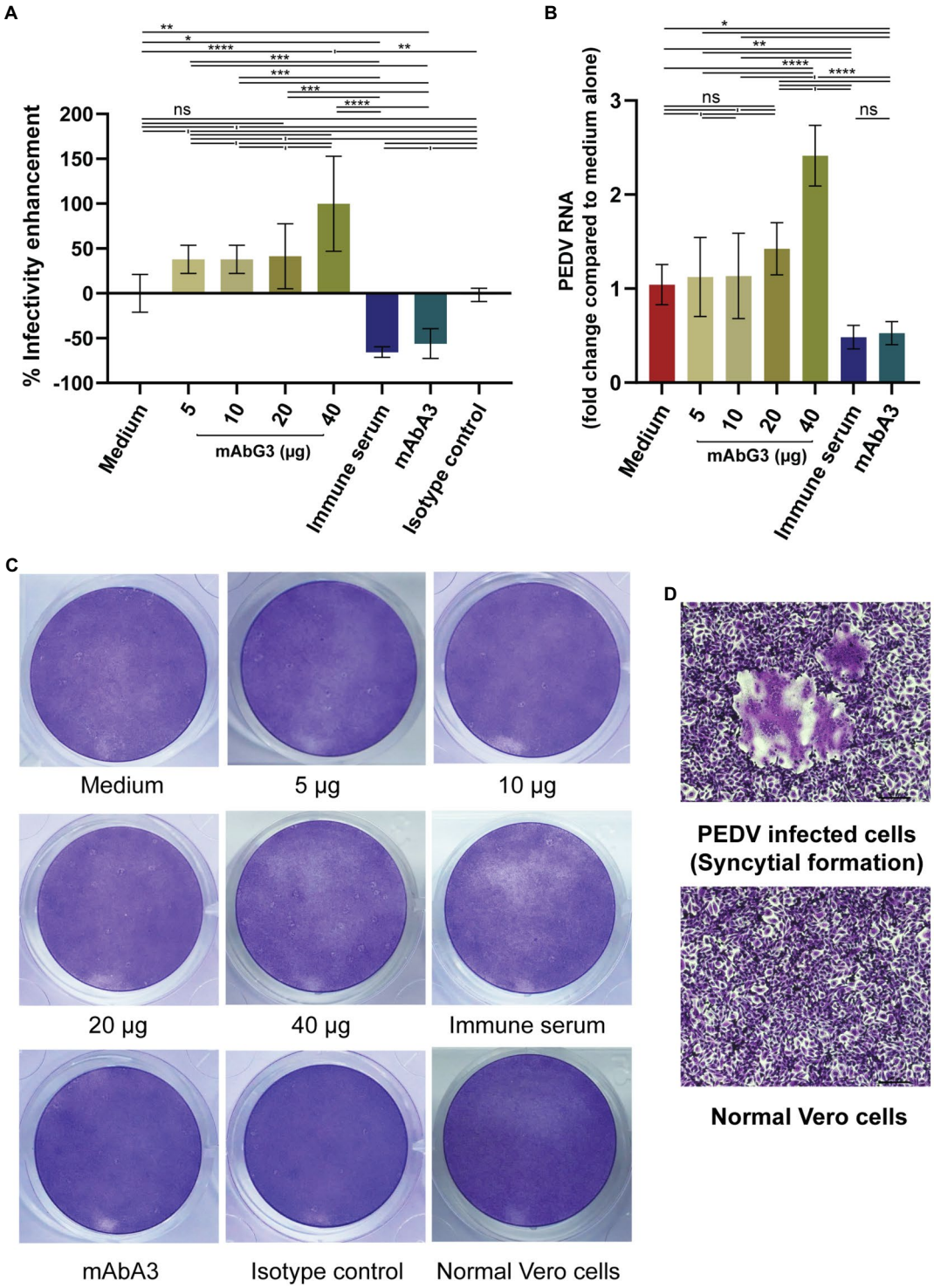


FIGURE 1
Enhancement of the PEDV infectivity mediated by the mAbG3. **(A)** The mAbG3 at the concentrations of 5, 10, 20, and 40 μg was mixed with 100 pfu of PEDV GII strain P70 and incubated for 1 h before adding to the confluent monolayers of Vero cells. Immune serum to PEDV S1 subunit (1:100) and mAbA3 (40 μg) served as positive neutralization controls, medium alone served as negative neutralization/enhancement control, and isotype-matched IgG1-kappa mAb served as irrelevant mAb control. After 1 h of incubation, the fluid in each well was removed; the cells were rinsed and covered with 2% CMC in DMEM supplemented with 2 μg/ml of TPCK-treated trypsin. Two days post-infection, cells were fixed, and the number of syncytial formations was counted and calculated to % enhancement of PEDV infectivity compared to the medium alone. Data of each bar graph are presented as the mean and standard deviation of the % mAbG3-mediated enhancement of PEDV infectivity from triplicate wells of each treatment. **(B)** Results of qRT- PCR for determination of PEDV RNA recovered from different treatment groups. Y-axis, fold changes of PEDV RNA in different treatment groups compared to medium alone; X-axis, various treatment groups. By one-way ANOVA: ns, not significantly different; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. **(C)** Examples of the PEDV-mediated Vero cell CPE of different treatments. **(D)** Appearance of syncytial formation of PEDV-infected Vero cells (upper panel) compared with normal Vero cells (lower panel) at ×100 magnification of a light microscope.

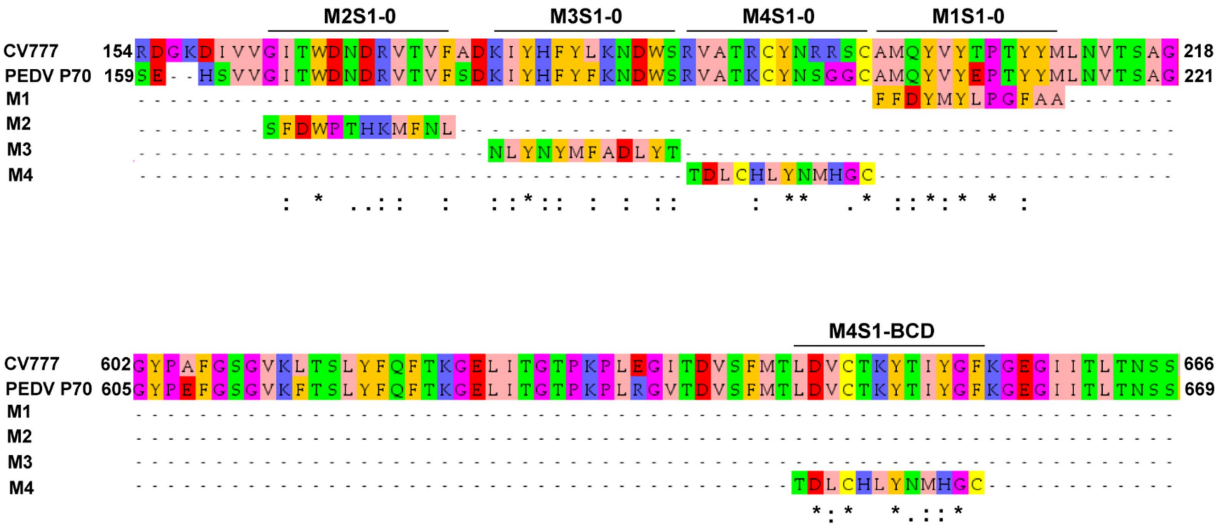


FIGURE 2
Linear amino acid sequences of S1-0 (residues 154–218/221) and S1-BCD (residues 602–666/669) sub-domains of S1 proteins of PEDV strains CV777 (GI; GenBank: AEX92968.1) and P70 (GII) showing regions matched with the phage mimotope types 1–4 (M1-M4), i.e., the S1-0 residues 165GITWDNDRVTVF176 matched with M2 (M2S1–0), 179KIYHFYFKNDWS190 matched with M3 (M3S1–0), 191RVATKCYNSGGC202 matched with M4 (M4S1–0), 203AMQYVYEPTYYM214 matched with M1 (M1S1–0), and S1-BCD sub-domain residues 646LDVCTKYTIYGF657 matched with M4 (M4S1-BCD). The amino acid residues were colored according to the Zappos scheme that is based on physicochemical properties. *, identical amino acids; :, conserved amino acids.

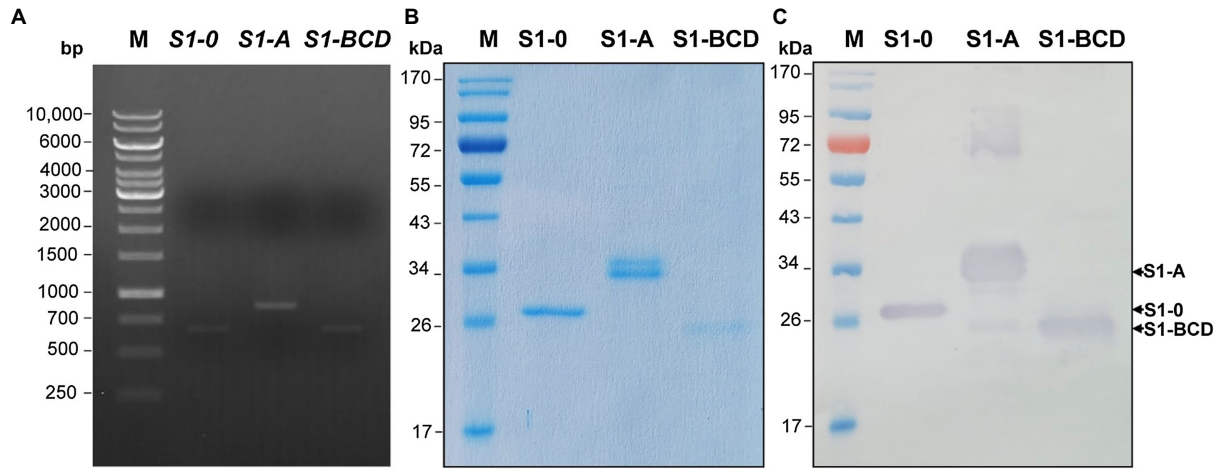


FIGURE 3
Production of recombinant polypeptides of PEDV S1 subunit (S1-0, S1-A, and S1-BCD). **(A)** PCR amplificons of the DNA sequences coding for the S1-0 (651 bp), S1-A (873 bp), and S1-BCD (660 bp). Lane M, 1 kb DNA ladder. The numbers on the left are DNA sizes in bp. **(B)** SDS-PAGE-separated patterns of the purified S1-0 (26 kDa), S1-A (32 kDa), and S1-BCD (24 kDa) after staining with CBB dye. **(C)** Western blot patterns of the purified S1-0, S1-A, and S1-BCD as detected by mouse anti-6× His antibody. Lanes M of **(B)** and **(C)** are protein molecular mass markers. The numbers on the left of **(B)** and **(C)** are protein molecular masses in kDa.

particles/virus-infected cells, opsonization of the virus particles for phagocytosis, and antibody-dependent cell-mediated cytotoxicity (ADCC) by cytotoxic T lymphocytes and NK cells (Forthal, 2014). These effective antibodies (may be vaccine-induced, disease convalescing, or passively acquired) mainly target the surface-exposed virus ligands, e.g., spike (S) proteins of coronaviruses, glycoprotein (G) protein of rabies virus, VP proteins of enteroviruses, etc. The efficacy of the antibodies depends on the affinity as well as target accessibility and occupancy (Dowd et al., 2011; Ertl, 2019; Densumite et al., 2021; Edara et al., 2021). It was during the 1960s-1970s that a dark side of the antibodies in virus infection and immunity was recognized. IgG in the avian antiserum mediated immune enhancement of infectivity in rabbitpox virus and Murray Valley encephalitis virus when assayed in the avian cell culture system (Hawkes and Lafferty, 1967). It was also observed that secondary infection with dengue viruses was associated with severe

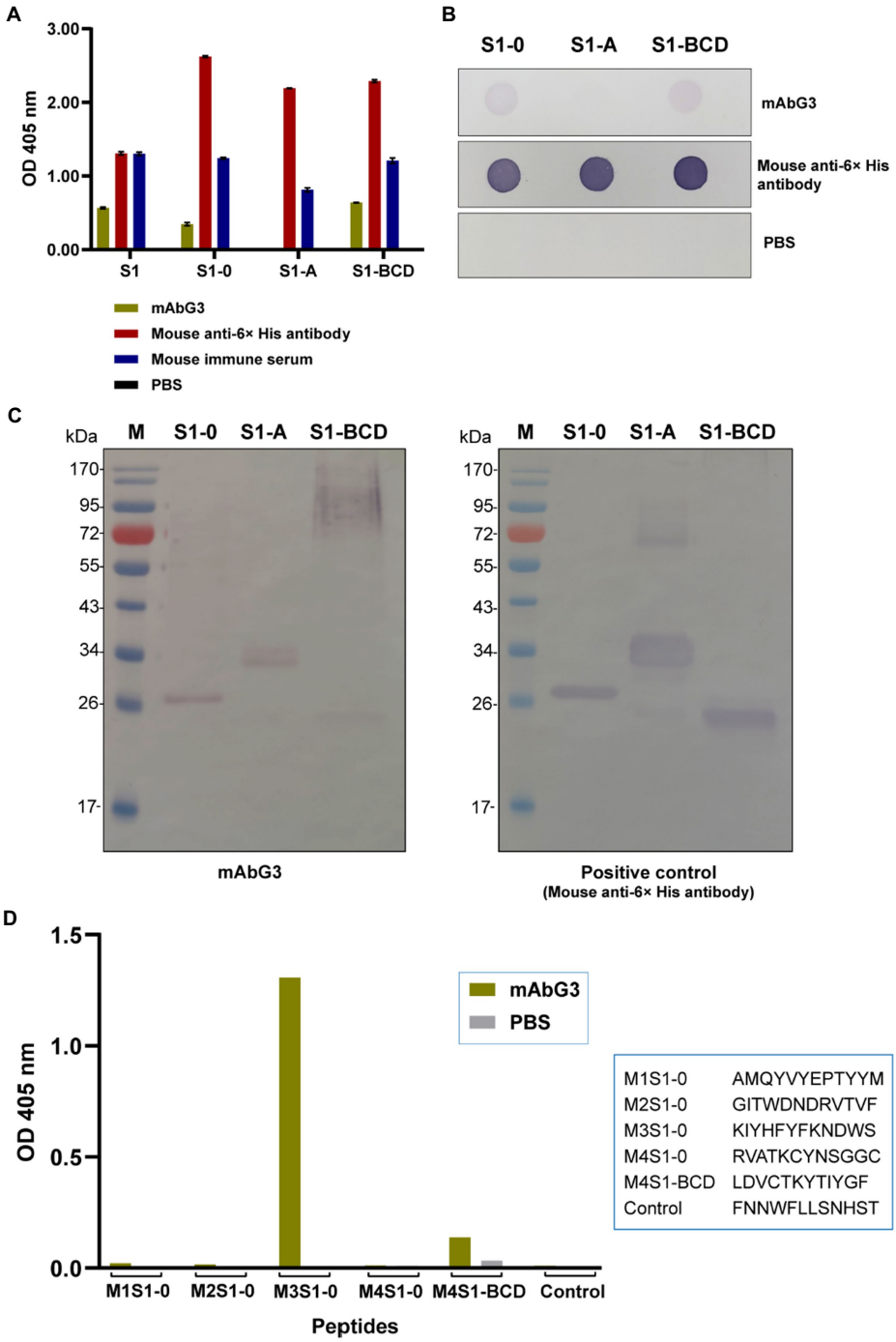


FIGURE 4
Binding of mAbG3 to recombinant polypeptides S1-0, S1-A, and S1-BCD of the PEDV S1 subunit. **(A)** Indirect ELISA (OD 405 nm) of mAbG3 binding to the full-length recombinant S1 subunit and the S1 polypeptides S1-0, S1-A, and S1-BCD. Mouse anti-6× His antibody and mouse immune serum (1:10,000) served as positive binding controls, and PBS served as a negative binding control. **(B)** Binding of mAbG3 to S1-0, S1-A, and S1-BCD by dot ELISA. Mouse anti-6× His antibody (1:1,000) and PBS served as positive and negative binding controls, respectively. **(C)** Binding of mAbG3 to SDS-PAGE-separated S1-0, S1-A, and S1-BCD by Western blotting (left panel). Mouse anti-6× His antibody (1:1,000) was used as a positive control (right panel). **(D)** Binding of mAbG3 to M1S1−0, M2S1−0, M3S1−0, M4S1−0, M4S1-BCD, and control peptides by peptide binding ELISA. Y-axis, OD 405 nm; X-axis, peptides used as the test antigens; PBS served as control.

illness along with higher viremia (Libraty et al., 2002; Halstead, 2012). The paradoxical role of antibodies in the dengue infection was found to be caused by sub- or non- neutralizing antibodies which facilitate the virus entry into Fc receptor bearing host cells, e.g., monocytes, macrophages, and dendritic cells, *via* the Fc fragments, which leads to more viral load and hence more clinical severity

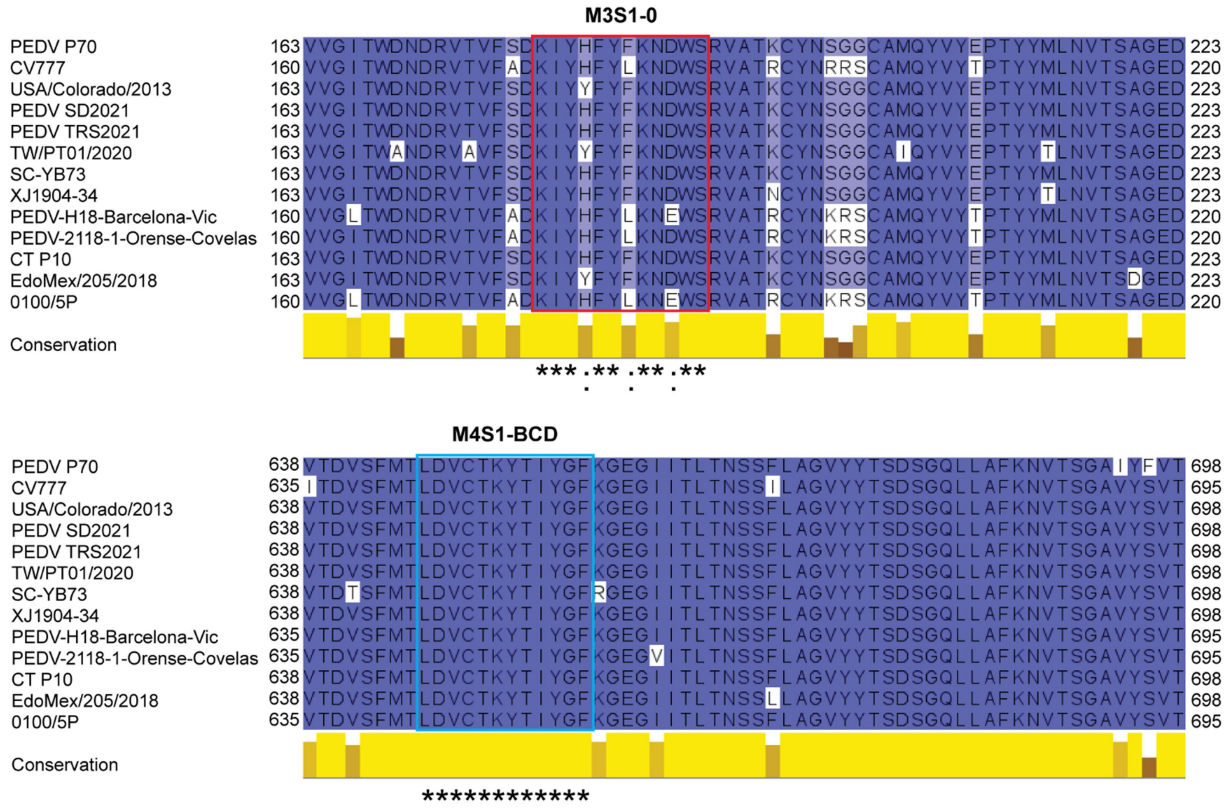


FIGURE 5
Multiple sequence alignment of amino acid sequences of spike proteins of various PEDV strains including PEDV P70 (a strain that was used in this study), CV777 (classical strain), USA/Colorado/2013 (prototype strain that was used for producing baculovirus-derived S subunit vaccine that caused severe enhancement; GenBank: KF272920.1), PEDV SD2021 (GenBank: UJZ92254.1), PEDV TRS2021 (GenBank: UJZ92268.1), TW/PT01/2020 (GenBank: UOK15705.1), SC-YB73 (GenBank: QNL15619.1), XJ1904-34 (GenBank: UGN13709.1), PEDV-H18-Barcelona-Vic (GenBank: QKV43853.1), PEDV-2118-1-Orense-Covelas (GenBank: QKV43823.1), CT P10 (GenBank: QHB92364.1), EdoMex/205/2018 (GenBank: QQK84868.1), and 0100/5P (GenBank: UAJ21031.1). The M3S1−0 sequences (red box) are highly conserved, and M4S1-BCD sequences (blue box) are identical among these PEDV strains. *, identical amino acids; :, conserved amino acids.

(Kulkarni, 2020). The phenomenon is currently known as "extrinsic" antibody-dependent enhancement (ADE; Kulkarni, 2020). The extrinsic ADE is mainly mediated by IgG, but IgM and IgA can also induce ADE (Janoff et al., 1995). Currently, several other molecular mechanisms of ADE that aggravate viral diseases have been recognized, including intrinsic ADE which results in heightened virus production by inhibition of type 1 interferon and activation of interleukin-10 biosynthesis, thereby favoring a Th2-type immune response (Narayan and Tripathi, 2020), immune enhancement ADE (Hotez et al., 2020), complement-mediated ADE (Prohászka et al., 1997), and ADE *via* host immune suppression by the intracellular virus (Chareonsirisuthigul et al., 2007). Enhancement of virus entry by non-neutralizing antibodies can be mediated by other mechanisms, e.g., antibody-mediated cross-linking of the spike proteins, which subsequently results in promoting the up-standing/open form of the receptor-binding domain (RBD) of trimeric S protein of SARS-CoV-2 (Liu et al., 2021). For the influenza viruses, the non-neutralizing antibodies promote the increment of hemagglutinin stem flexibility and virus fusion kinetics (Winarski et al., 2019).

One of the many obstacles to vaccine development against virus infections includes the ADE induced by the vaccines at new exposure to the authentic virus. For the PEDV, a baculovirus-expressed recombinant spike protein vaccine enhanced PED symptoms and PEDV replication in challenged immune piglets, although the vaccine provided a high neutralizing antibody titer (Yu et al., 2021). It was speculated that the non-neutralizing antibodies specific to the post-fusion form of the spike protein promoted the PEDV replication in the immune piglets and thus enhanced PED (Yu et al., 2021). However, the enhancing epitope of the S protein has not yet been identified.

In this study, a monoclonal antibody (mAbG3) that bound to the PEDV S1 subunit (Thavorasak et al., 2022) was found to enhance the PEDV infectivity in Vero cell culture. The enhancement was obvious at the high mAbG3 dose (40 μg), while the lower doses (5, 10, and 20 μg) only showed the enhancing trend, but the effect was not different from the infected cells in the medium alone. The higher mAbG3 doses (> 40 μg) were not tested, as the experiment was conducted only to show that the mAbG3 was an enhancing antibody and to find the enhancing epitope. The mAbG3-mediated ADE of

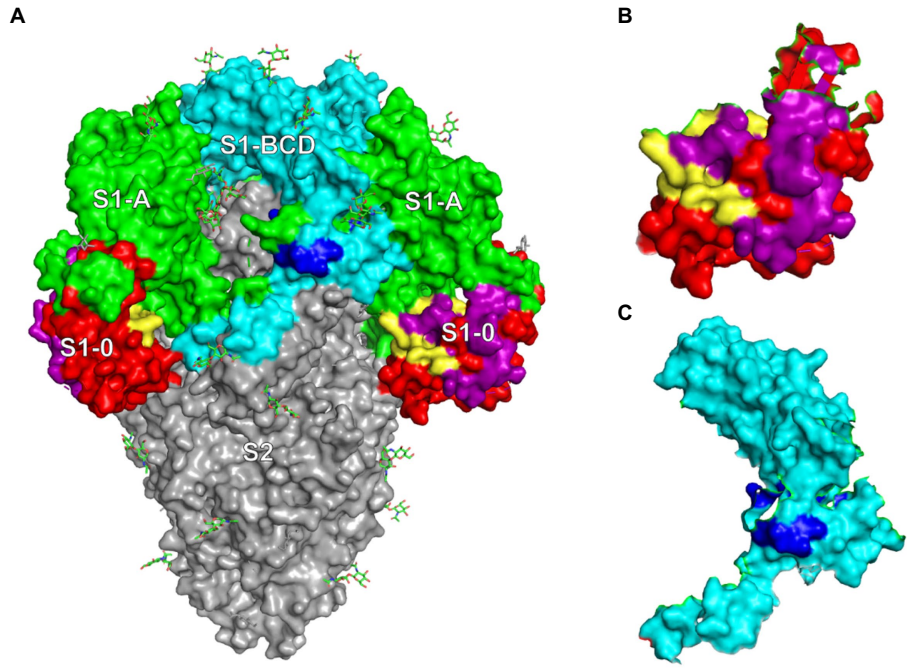


FIGURE 6
(A) PEDV trimeric spike protein (side view) showing locations of S1-0 (red, purple, and yellow), S1-A (green), S1-BCD (cyan and blue), and S2 subunits (gray). **(B)** Monomeric S1-0 showing the regions matched with phage mimotopes M1-M4 (purple and yellow); the yellow area is the M3S1–0. **(C)** S1-BCD monomer (cyan) showing the region bound by the mAbG3 (M4S1-BCD in blue).

the PEDV should not be the extrinsic ADE as the Vero cells lack the Fc receptor. Likewise, the ADE cannot be a result of immune enhancement *via* complement activation and recruitment of inflammatory and immune cells, as there were no complement proteins or inflammatory/immune cells in the *in vitro* assay performed in this study, although this mechanism may occur *in vivo*. More likely, the ADE that is observed when the PEDV-infected cells were incubated with the mAbG3 may be due to intrinsic ADE, i.e., the virus-immune complexes internalized *via* macropinocytosis (Mercer and Helenius, 2009) may modulate innate immune effectors to favor increment of virus replication and release (Narayan and Tripathi, 2020). It was also observed previously that Toll-like receptor 2 (TLR2)-mediated NF-κB activation was inhibited in Vero cells infected with herpesvirus of turkeys (HVT; Yang et al., 2013). Therefore, the PEDV-mAbG3 complexes might inhibit the danger signaling that usually occurs *via* the pathogen recognition receptors (PRRs) and favors the virus replication (Flipse et al., 2016). The mAbG3 may cross-link the spike proteins which eventually results in promoting the up-standing/open form of the receptor-binding domain (RBD) of trimeric S protein of the PEDV, as observed for SARS-CoV-2 (Liu et al., 2021).

Phage Display Peptide Library was used as a biological tool for searching the location of the enhancing epitope of the PEDV S protein. After the bio-panning of peptide phage with the immobilized mAbG3, four phage mimotope types (M1-M4) were obtained. Alignments of the four phage mimotope/peptide sequences with the linear S1 subunit sequences of both GI (classical CV777) and GII (field isolated P70) strains of PEDV found that the S1 sub-domain of both PEDV strains that matched with the phage mimotope sequences was located in the multiple regions of the S1-0 sub-domain, i.e., M1S1–0, M2S1–0, M3S1–0, and M4S1–0, as well as a linear sequence in the S1-BCD sub-domain (M4S1-BCD). The finding was verified by both indirect ELISA and dot ELISA by using the S1 truncated recombinant polypeptides as antigens in the assays and by the peptide binding ELISA. These findings indicate that the epitope of the enhancing mAbG3 is shared by the S1-0 and the S1-BCD sub-domains. In the peptide binding ELISA, the mAbG3 produced negligible ELISA signals to the M1S1–0, M2S1–0, and M4S1–0. Some residues in these three regions may participate in forming the mAbG3 conformational (structural) epitope (their residues were juxtaposed with the M3S1–0 upon the S1 folding to form the conformational epitope but could not be bound by the mAbG3 tight enough to give significant signals in their linear sequences).

There was no mAbG3-bound phage peptide that matched with the sequence of the S1-A sub-domain, although the mAbG3 bound to the S1-A polypeptide by Western blotting. In the phage panning with the immobilized mAbG3, the phage displaying peptide that matched with the S1-A peptide might not be able to compete with the phages that displayed mimotopes for S1-0 and S1-BCD and/or might be due to less population of the former than the latter in the Ph.D.™-12 Phage Display Library. The binding test of the mAbG3 with the three truncated S1 subunit polypeptides in Western blot analysis showed that the mAbG3

could bind to the S1-A polypeptide, although the antibody did not bind to the S1-A when tested by both ELISAs. These findings can be explained by the different physicochemical structures of the proteins when they were used as antigens in the assays of different principles (ELISA *versus* Western blotting). The monomeric recombinant S1-A in the free folding structure was coated or dotted onto the supported matrices (well surface of the ELISA plate or nitrocellulose membrane). In these conditions, the S1-A epitope might be hidden in the folded protein and could not be accessible to the mAbG3, thus producing negative ELISA results. In the Western blot analysis, the antigen (S1-A polypeptide) was exposed to reducing agents like 2-mercaptoethanol and SDS; the SDS was also bound to the polypeptide during the electrophoresis. Therefore, the antigen should acquire the stretched (linear) configuration which exposed the mAbG3 recognition site on the blotted membrane, thus giving rise to the antigen–antibody reactive band after adding the mAbG3. This mAbG3-bound region should not be surface-exposed on the trimeric S protein of authentic PEDV and hence should not contribute to the ADE.

The sequence of M4 matched with two sites on the PEDV S1 subunit, i.e., 191RVATKCYNSGGC202 of the S1-0 (M4S1–0) and 646LDVCTKYTIYGF657 of the S1-BCD (M4S1-BCD). When mapping these two regions on the trimeric S protein (Figure 6A), it was found that only the residues 646LDVC649 of the S1-BCD are exposed, while the 650TKYTIYGF657 are obscured underneath. The mAbG3 could access the S1-BCD epitope in the ELISA, as the recombinant S1-BCD used as the antigen was monomeric and thus exposing the epitope. Naturally, only a small portion of the S1-BCD would be exposed on the functionally active trimeric spike protein, while the S1-0191RVATKCYNSGGC202 (M4S1–0) is fully surface-exposed. Binding of the antibody specific to this epitope and the S1-BCD surface-exposed epitope portion could cause extrinsic ADE (enhancement of PEDV entry and increased replication) and immune enhancement ADE (the virus-immune complexes internalized *via* macropinocytosis and modulate innate antivirus activity to favor increment of virus replication and release; Mercer and Helenius, 2009; Narayan and Tripathi, 2020).

Taken all the data together, the enhancing epitope of the PEDV S1 subunit that may cause ADE was found to be located at the C terminus of the S1-0 sub-domain and S1-BCD. Although the exact residues that participate in the epitope formation need further elucidation (such as by co-crystallography), the finding in this study together with the information on the known S1 neutralizing epitopes identified previously (Chang et al., 2002, 2019, 2021; Cruz et al., 2008; Sun et al., 2008; Okda et al., 2017; Ho et al., 2020; Tien et al., 2021; Thavorasak et al., 2022) should be useful in designing a safe PEDV vaccine, i.e., by using a recombinant spike (S) protein whose C-terminal portion of S1-0 and S1-BCD have been deleted and contain only the neutralizing B-cell epitopes along with appropriate T-cell epitopes as immunogens in a subunit protein vaccine or by using the respective DNA counterparts as a DNA vaccine.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

WC, NS, and KT: conceptualization, project administration, and funding acquisition. WC, NS, MC, KG-a, KM, TS, RY, and DN: methodology. WC and TT: validation. WC: formal analysis and data curation. TT, MC, KG-a, KM, TS, RY, DN, and NS-l: investigation. WC, KT, and NS: resources. WC and TT original draft preparation. WC, NS, and TT: reviewing and editing. TT, MC, and KG-a: visualization. WC, NS, KT, MC, KG-a, and KM: supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Program Management Unit B (PMU.B), Ministry of Higher Education Science Research and Innovation, Thailand, and Mahidol University. TT received a Research and Researchers for Industries (RRI) scholarship grant (PHD61I0028) from the National Research Council of Thailand (NRCT), Ministry of Higher Education, Research and Innovation, Thailand.

Conflict of interest

KT was employed by the MORENA Solution Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.933249/full#supplementary-material

References

Annamalai, T., Saif, L. J., Lu, Z., and Jung, K. (2015). Age-dependent variation in innate immune responses to porcine epidemic diarrhea virus infection in suckling *versus* weaned pigs. *Vet. Immunol. Immunopathol.* 168, 193–202. doi: 10.1016/j.vetimm.2015.09.006

Chang, C. Y., Cheng, I. C., Chang, Y. C., Tsai, P. S., Lai, S. Y., Huang, Y. L., et al. (2019). Identification of neutralizing monoclonal antibodies targeting novel conformational epitopes of the porcine epidemic diarrhoea virus spike protein. *Sci. Rep.* 9, 2529. doi: 10.1038/s41598-019-39844-5

Chang, C. Y., Wang, Y.-S., Wu, J.-F., Yang, T.-J., Chang, Y.-C., Chae, C., et al. (2021). Generation and characterization of a spike glycoprotein domain A-specific neutralizing single-chain variable fragment against porcine epidemic diarrhea virus. *Vaccines.* 9, 833. doi: 10.3390/vaccines9080833

Chang, S. H., Bae, J. L., Kang, T. J., Kim, J., Chung, G. H., Lim, C. W., et al. (2002). Identification of the epitope region capable of inducing neutralizing antibodies against the porcine epidemic diarrhea virus. *Mol. Cells* 14, 295–299.

Chareonsirisuthigul, T., Kalayanarooj, S., and Ubol, S. (2007). Dengue virus (DENV) antibody-dependent enhancement of infection upregulates the production of anti-inflammatory cytokines but suppresses anti-DENV free radical and pro-inflammatory cytokine production, in THP-1 cells. *J. Gen. Virol.* 88, 365–375. doi: 10.1099/vir.0.82537-0

Cruz, D. J., Kim, C.-J., and Shin, H.-J. (2008). The GPRLQPY motif located at the carboxy-terminal of the spike protein induces antibodies that neutralize porcine epidemic diarrhea virus. *Virus Res.* 132, 192–196. doi: 10.1016/j.virusres.2007.10.015

Densumite, J., Phanthong, S., Seesuay, W., Sookrung, N., Chaisri, U., and Chaicumpa, W. (2021). Engineered human monoclonal scFv to receptor binding domain of Ebolavirus. *Vaccines* 9, 457. doi: 10.3390/vaccines9050457

Dowd, K. A., Jose, C. A., Durbin, A. P., Whitehead, S. S., and Pierson, T. C. (2011). A dynamic landscape for antibody binding modulates antibody-mediated neutralization of West Nile virus. *PLoS Pathog.* 7:e1002111. doi: 10.1371/journal.ppat.1002111

Edara, V. V., Pinsky, B. A., Suthar, M. S., Lai, L., Davis-Gardner, M. E., Floyd, K., et al. (2021). Infection and vaccine-induced neutralizing-antibody responses to the SARS-CoV-2 B.1.617 variants. *N. Engl. J. Med.* 385, 664–666. doi: 10.1056/NEJMc2107799

Ertl, H. C. J. (2019). New rabies vaccines for use in humans. *Vaccines* 7, 54. doi: 10.3390/vaccines7020054

Flipse, J., Diosa-Toro, M. A., Hoornweg, T. E., van de Pol, D. P., Urcuqui-Inchima, S., and Smit, J. M. (2016). Antibody-dependent enhancement of dengue virus infection in primary human macrophages; balancing higher fusion against antiviral responses. *Sci. Rep.* 6, 29201. doi: 10.1038/srep29201

Forthal, D. N. (2014). Functions of antibodies. *Microbiol. Spectr.* 2, 1–17. doi: 10.1128/microbiolspec.AID-0019-2014

Gerdts, V., and Zakhartchouk, A. (2017). Vaccines for porcine epidemic diarrhea virus and other swine coronaviruses. *Vet. Microbiol.* 206, 45–51. doi: 10.1016/j.vetmic.2016.11.029

Guo, J., Fang, L., Ye, X., Chen, J., Xu, S., Zhu, X., et al. (2019). Evolutionary and genotypic analyses of global porcine epidemic diarrhea virus strains. *Transbound. Emerg. Dis.* 66, 111–118. doi: 10.1111/tbed.12991

Halstead, S. B. (2012). Controversies in dengue pathogenesis. *Paediatr. Int. Child Health.* 32, 5–9. doi: 10.1179/2046904712Z.00000000045

Hawkes, R. A., and Lafferty, K. J. (1967). The enhancement of virus infectivity by antibody. *Virology* 33, 250–261. doi: 10.1016/0042-6822(67)90144-4

Ho, T. T., Nguyen, G. T., Pham, N. B., Le, V. P., Trinh, T. B., Vu, T. H., et al. (2020). Plant-derived trimeric co-26k-equivalent epitope induced neutralizing antibodies against porcine epidemic diarrhea virus. *Front. Immunol.* 11:2152. doi: 10.3389/fimmu.2020.02152

Hotez, P. J., Bottazzi, M. E., and Corry, D. B. (2020). The potential role of Th17 immune responses in coronavirus immunopathology and vaccine-induced immune enhancement. *Microbes Infect.* 22, 165–167. doi: 10.1016/j.micinf.2020.04.005

Janoff, E. N., Wahl, S. M., Thomas, K., and Smith, P. D. (1995). Modulation of human immunodeficiency virus type 1 infection of human monocytes by IgA. *J. Infect. Dis.* 172, 855–858. doi: 10.1093/infdis/172.3.855

Jung, K., Miyazaki, A., and Saif, L. J. (2018). Immunohistochemical detection of the vomiting-inducing monoamine neurotransmitter serotonin and enterochromaffin cells in the intestines of conventional or gnotobiotic (Gn) pigs infected with porcine epidemic diarrhea virus (PEDV) and serum cytokine responses of Gn pigs to acute PEDV infection. *Res. Vet. Sci.* 119, 99–108. doi: 10.1016/j.rvsc.2018.06.009

Jung, K., Saif, L. J., and Wang, Q. (2020). Porcine epidemic diarrhea virus (PEDV): An update on etiology, transmission, pathogenesis, and prevention and control. *Virus Res.* 286:198045. doi: 10.1016/j.viruses.2020.198045

Kirchdoerfer, R. N., Bhandari, M., Martini, O., Sewall, L. M., Bangaru, S., Yoon, K., et al. (2021). Structure and immune recognition of the porcine epidemic diarrhea virus spike protein. *Structure* 29, 385–392. doi: 10.1016/j.str.2020.12.003

Kulkarni, R. (2020). Antibody-dependent enhancement of viral infections. *Dyn. Immune Activation Viral Diseases.* 2019, 9–41. doi: 10.1007/978-981-15-1045-8_2

Lee, C. (2015). Porcine epidemic diarrhea virus: An emerging and re-emerging epizootic swine virus. *Virol. J.* 12, 193. doi: 10.1186/s12985-015-0421-2

Li, C., Li, W., Lucio de Esesarte, E., Guo, H., Van Den Elzen, P., Aarts, E., et al. (2017). Cell attachment domains of the porcine epidemic diarrhea virus spike protein are key targets of neutralizing antibodies. *J. Virol.* 91, e00273–e00217. doi: 10.1128/JVI.00273-17

Li, Z., Ma, Z., Li, Y., Gao, S., and Xiao, S. (2020). Porcine epidemic diarrhea virus: Molecular mechanisms of attenuation and vaccines. *Microb. Pathog.* 149:104553. doi: 10.1016/j.micpath.2020.104553

Libraty, D. H., Endy, T. P., Houng, H. S., Green, S., Kalayanarooj, S., Suntayakorn, S., et al. (2002). Differing influences of virus burden and immune activation on disease severity in secondary dengue-3 virus infections. *J. Infect. Dis.* 185, 1213–1221. doi: 10.1086/340365

Lin, C. M., Saif, L. J., Marthaler, D., and Wang, Q. (2016). Evolution, antigenicity, and pathogenicity of global porcine epidemic diarrhea virus strains. *Virus Res.* 226, 20–39. doi: 10.1016/j.virusres.2016.05.023

Liu, Y., Soh, W. T., Kishikawa, J. I., Hirose, M., Nakayama, E. E., Li, S., et al. (2021). An infectivity-enhancing site on the SARS-CoV-2 spike protein targeted by antibodies. *Cell* 184, 3452–3466.e18. doi: 10.1016/j.cell.2021.05.032

Mercer, J., and Helenius, A. (2009). Virus entry by macropinocytosis. *Nat. Cell Biol.* 11, 510–520. doi: 10.1038/ncb0509-510

Narayan, R., and Tripathi, S. (2020). Intrinsic ADE: The dark side of antibody dependent enhancement during dengue infection. *Front. Cell. Infect. Microbiol.* 10:580096. doi: 10.3389/fcimb.2020.580096

Okda, F. A., Lawson, S., Singrey, A., Nelson, J., Hain, K. S., Joshi, L. R., et al. (2017). The S2 glycoprotein subunit of porcine epidemic diarrhea virus contains immunodominant neutralizing epitopes. *Virology* 509, 185–194. doi: 10.1016/j.virol.2017.06.013

Pensaert, M. B., and de Bouck, P. (1978). A new coronavirus-like particle associated with diarrhea in swine. *Arch. Virol.* 58, 243–247. doi: 10.1007/BF01317606

Prohászka, Z., Nemes, J., Hidvégi, T., Tóth, F. D., Kerekes, K., Erdei, A., et al. (1997). Two parallel routes of the complement-mediated antibody-dependent enhancement of HIV-1 infection. *AIDS* 11, 949–958. doi: 10.1097/00002030-199708000-00002

Sun, D., Feng, L., Shi, H., Chen, J., Cui, X., Chen, H., et al. (2008). Identification of two novel B cell epitopes on porcine epidemic diarrhea virus spike protein. *Vet. Microbiol.* 131, 73–81. doi: 10.1016/j.vetmic.2008.02.022

Thavorasak, T., Chulanetra, M., Glab-ampai, K., Teeranitayatarn, K., Songserm, T., Yodsheewan, R., et al. (2022). Novel neutralizing epitope of PEDV S1 protein identified by IgM monoclonal antibody. *Viruses* 14:125. doi: 10.3390/v14010125

Tien, N.-Q.-D., Yang, M.-S., Jang, Y.-S., Kwon, T.-H., Reljic, R., and Kim, M. Y. (2021). Systemic and oral immunogenicity of porcine epidemic diarrhea virus antigen fused to poly-Fc of immunoglobulin G and expressed in ΔXT/ft Nicotiana benthamiana plants. *Front. Pharmacol.* 12:653064. doi: 10.3389/fphar.2021.653064

Wang, L., Byrum, B., and Zhang, Y. (2014). Detection and genetic characterization of deltacoronavirus in pigs, Ohio, USA, 2014. *Emerg. Infect. Dis.* 20, 1227–1230. doi: 10.3201/eid2007.140296

Wang, Q., Vlasova, A. N., Kenney, S. P., and Saif, L. J. (2019). Emerging and re-emerging coronaviruses in pigs. *Curr. Opin. Virol.* 34, 39–49. doi: 10.1016/j.coviro.2018.12.001

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview version 2- -a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Wen, J., Cheng, Y., Ling, R., Dai, Y., Huang, B., Huang, W., et al. (2020). Antibody-dependent enhancement of coronavirus. *Int. J. Infect. Dis.* 100, 483–489. doi: 10.1016/j.ijid.2020.09.015

Winarski, K. L., Tang, J., Klenow, L., Lee, J., Coyle, E. M., Manischewitz, J., et al. (2019). Antibody-dependent enhancement of influenza disease promoted by increase in hemagglutinin stem flexibility and virus fusion kinetics. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15194–15199. doi: 10.1073/pnas.1821317116

Won, H., Lim, J., Noh, Y. H., Yoon, I., and Yoo, H. S. (2020). Efficacy of porcine epidemic diarrhea vaccines: A systematic review and meta-analysis. *Vaccine* 8:642. doi: 10.3390/vaccines8040642

Xu, L., Ma, Z., Li, Y., Pang, Z., and Xiao, S. (2021). Antibody dependent enhancement: Unavoidable problems in vaccine development. *Adv. Immunol.* 151, 99–133. doi: 10.1016/bs.ai.2021.08.003

Yang, Q., Chen, H., Wei, T., and Wei, P. (2013). Inhibition of toll-like receptor 2-mediated NF-kappa B activation in Vero cells with herpesvirus of turkeys. *Avian Dis.* 57, 409–415. doi: 10.1637/10327-081712-Reg.1

Yu, J., Sreenivasan, C., Uprety, T., Gao, R., Huang, C., Lee, E. J., et al. (2021). Piglet immunization with a spike subunit vaccine enhances disease by porcine epidemic diarrhea virus. *NPJ Vaccines* 6:22. doi: 10.1038/s41541-021-00283-x

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fmicb.2022.926929

Targeting Emerging RNA Viruses by Engineered Human Superantibody to Hepatitis C Virus RNA-Dependent RNA Polymerase

*Kittirat Glab-ampai[1], Kanasap Kaewchim[1,2], Techit Thavorasak[1,2], Thanatsaran Saenlom[1], Watayagorn Thepsawat[1], Kodchakorn Mahasongkram[1], Kanyarat Thueng-In[3], Nitat Sookrung[1,4], Wanpen Chaicumpa[1] and Monrat Chulanetra[1]**

[1] Center of Research Excellence in Therapeutic Proteins and Antibody Engineering, Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, [2] Graduate Program in Immunology, Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, [3] School of Pathology, Translational Medicine Program, Institute of Medicine, Suranaree University of Technology, Nakhon Ratchasima, Thailand, [4] Biomedical Research Incubator Unit, Department of Research, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

OPEN ACCESS

Edited by:
Yoichi Takakusagi,
National Institutes for Quantum and Radiological Science and Technology, Japan

Reviewed by:
Jian Huang,
University of Electronic Science and Technology of China, China
Fumio Sugawara,
Tokyo University of Science, Japan

***Correspondence:**
Monrat Chulanetra
monrat.chl@mahidol.edu

Specialty section:
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

Received: 23 April 2022
Accepted: 15 June 2022
Published: 22 July 2022

Citation:
Glab-ampai K, Kaewchim K, Thavorasak T, Saenlom T, Thepsawat W, Mahasongkram K, Thueng-In K, Sookrung N, Chaicumpa W and Chulanetra M (2022) Targeting Emerging RNA Viruses by Engineered Human Superantibody to Hepatitis C Virus RNA-Dependent RNA Polymerase. Front. Microbiol. 13:926929. doi: 10.3389/fmicb.2022.926929

RNA-dependent RNA polymerase (RdRp) is a unique and highly conserved enzyme across all members of the RNA virus superfamilies. Besides, humans do not have a homolog of this protein. Therefore, the RdRp is an attractive target for a broadly effective therapeutic agent against RNA viruses. In this study, a formerly generated cell-penetrating human single-chain antibody variable fragment (superantibody) to a conformational epitope of hepatitis C virus (HCV) RdRp, which inhibited the polymerase activity leading to the HCV replication inhibition and the host innate immunity restoration, was tested against emerging/reemerging RNA viruses. The superantibody could inhibit the replication of the other members of the *Flaviviridae* (DENV serotypes 1−4, ZIKV, and JEV), *Picornaviridae* (genus *Enterovirus*: EV71, CVA16), and *Coronaviridae* (genus *Alphacoronavirus*: PEDV, and genus *Betacoronavirus*: SARS-CoV-2 (Wuhan wild-type and the variants of concern), in a dose-dependent manner, as demonstrated by the reduction of intracellular viral RNAs and numbers of the released infectious particles. Computerized simulation indicated that the superantibody formed contact interfaces with many residues at the back of the thumb domain (thumb II site, T2) of DENV, ZIKV, JEV, EV71, and CVA16 and fingers and thumb domains of the HCV and coronaviruses (PEDV and SARS-CoV-2). The superantibody binding may cause allosteric change in the spatial conformation of the enzyme and disrupt the catalytic activity, leading to replication inhibition. Although the speculated molecular mechanism of the superantibody needs experimental support, existing data indicate that the superantibody has high potential as a non-chemical broadly effective anti-positive sense-RNA virus agent.

Keywords: RNA viruses, RNA-dependent RNA polymerase, phage display, human single-chain antibody variable fragment, superantibody (cell penetrating antibody), computerized simulation, plaque-forming assay, focus-forming assay

INTRODUCTION

During the past two decades, several human and animal RNA viruses have emerged/reemerged to cause epidemics/epizootics or pandemics/panzootics that inflict a huge negative impact on the human and animal health as well as socioeconomics globally. Examples are influenza A viruses (IAV H5N1 and H1N1pdm2009) (Tang et al., 1998; Novel Influenza A/H1N1 Investigation Team, 2009); flaviviruses including dengue virus (DENV) (Kyle and Harris, 2008; European Centre for Disease Prevention and Control, 2020) and zika virus (ZIKV) (Noobrakhsh et al., 2019); ebola virus (EBOV) (World Health Organization Ebola Response Team, 2014); enteroviruses including EV71 and CVA16 (Schmidt et al., 1974); and coronaviruses (CoVs) including *Alphacoronavirus* (porcine epidemic diarrhea virus, PEDV), *Betacoronavirus* (severe acute respiratory syndrome virus, SARS-CoV; MERS-CoV; novel coronavirus 2019 or SARS-CoV-2), and *Deltacoronavirus* (porcine *Deltacoronavirus,* PDCoV) (Pensaert and de Bouck, 1978; Chan-Yeng et al., 2015; Hu et al., 2015; Jung et al., 2016; World Health Organization [WHO], 2019). Currently, the world population is facing the unprecedentedly scaled pandemic of coronavirus disease caused by the SARS-CoV-2, named COVID-19, that emerged in December 2019. The catastrophic COVID-19 pandemic caused by the SARS-CoV-2 mutated descendants (variants of concern, VOC) is still going on, although a large fraction of the world population has been vaccinated against the disease. As of March 10, 2022, more than 400 million people around the globe were infected by the SARS-CoV-2, and among them, more than 6 million were deceased. The world consternations frequently threatened by the emerging/re-emerging RNA viruses emphasize the need not only for effective vaccines but also for safe therapeutics to counteract the viruses, especially for those with severe morbidity.

RNA-dependent RNA polymerase (RdRp) is a highly conserved enzyme across all members of the RNA virus superfamilies (except *Retroviridae*), although the enzyme itself accounts for the rapid RNA virus mutations from the high rate of transcription errors. The RNA virus RdRps probably arose from a common ancestor (Payne, 2017). The enzyme is indispensable for the synthesis of the genomic RNA and the transcription process during the virus replication cycle (Payne, 2017). Positive-sense RNA viruses use their RNA genomes as mRNAs for protein synthesis, while the negative-sense RNA viruses use the genomic RNAs as templates of the RdRp-dependent transcriptional process in the generation of the plus-sense strand that functions as mRNAs. Some RNA viruses, including coronaviruses, use RdRp for subgenomic RNA synthesis. Although the RdRps of the RNA viruses are diverse in their amino acid sequences as well as the structural details (the RdRp molecule may be linked with other structures that perform other functions, such as methyltransferase, endonuclease, helicase, and nucleoside-triphosphatase), their catalytic modules are relatively conserved and composed of the palm, fingers, and thumb domains such that the overall architecture reminisces encircled/cupped human right hand (Jia and Gong, 2019). The catalytic motif (active site) of the RdRp is surrounded by the palm, fingers, and thumb domains with seven catalytic motifs (motifs A–G) distributed within the palm (motifs A–E) and fingers (motifs F–G) (Poch et al., 1989; Gorbalenya et al., 2002; Bruenn, 2003; te Velthuis, 2014; Wu et al., 2015; Venkataraman et al., 2018; Jia and Gong, 2019). The viral RdRp lacks a human homolog. It is the essential and most conserved protein of RNA viruses (Jia and Gong, 2019). The RdRps of *Flaviviridae*, Hepatitis C virus (HCV), DENV, ZIKV, West Nile virus, share high percentages of identity with RdRp of the *Coronaviridae* members (e.g., SARS-CoV, MERS-CoV, and SARS-CoV-2) (Picarazzi et al., 2020); it is highly plausible that drugs or therapeutics that act on the RdRp of the former virus family may as well affect the RdRp of the latter, if not also other families. This speculation is well supported by the evidence that sofosbuvir (a small molecular inhibitor of HCV RdRp/NS5B protein in combination with daclatasvir/Daklinza) showed effectiveness in reducing the mortality rate of patients with severe COVID-19 (Abbass et al., 2021; Zein et al., 2021). In this study, therefore, we tested a previously generated cell-penetrating human single-chain antibody (superantibody) to HCV RdRp that has been shown to effectively interfere with the HCV replication and rescued the virally suppressed host innate immunity (Thueng-In et al., 2014), for replication inhibition of several other positive-sense RNA viruses. The ultimate purpose is to develop the broadly effective superantibody further toward a clinical use as a pan, direct-acting anti-positive-sense RNA virus agent.

MATERIALS AND METHODS

Cells, Viruses, and Virus Propagation

Human hepatocellular carcinoma cells (Huh7), human embryonic kidney (HEK) 293T cells, African green monkey kidney epithelial (Vero) cells, and Rhabdomyosarcoma (RD) cells were obtained from American Type Culture Collection (ATCC, Manassas, VA, United States). Vero E6 cells were provided by Prasert Auewarakul, Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco, Thermo Fisher Scientific, Waltham, MA, United States) supplemented with 10% fetal bovine serum (FBS) (HyClone; GE Healthcare Life Sciences, Marlborough, MA, United States), 100 units/mL penicillin, 100 mg/mL streptomycin, and 2 mM L-glutamine (Gibco) (complete DMEM).

The viruses used in this study included HCV infectious particles, one isolate each of DENV serotypes 1-4; one isolate of ZIKV; one isolate of Japanese encephalitis virus (JEV); one isolate each of Wuhan wild-type, alpha (B.1.1.7), beta (B.1.351), delta (B.1.617.2), and omicron (B.1.1.529) of SARS-CoV-2; Enterovirus 71 (EV71, genotype A, BrCr strain, ATCCR-VR-1775TM); and Coxsackievirus A16 (CVA16) and PEDV (P70 strain, GII field isolate from a deceased infected piglet in Thailand).

The HCV infectious particles were prepared as described previously (Thueng-In et al., 2014). Full-length cDNA of pJFH-1 (Wakita et al., 2005) was linearized by digesting the plasmid with *Xba*I endonuclease (Fermentas, Burlington, ON, Canada), and 1 μg was transcribed *in vitro* by using a Megascript T7 kit (Ambion, Carlsbad, CA, United States). The RNA transcript

(10 μg) was electroporated into Huh7 cells (4.0×10^6 cells) in 0.4 mL of the serum reduced medium (Opti-MEM) (Invitrogen, Thermo Fisher Scientific) by using a single pulse at 0.27 kV and 100 milli-s. The transfected cells were immediately transferred to 40 mL of complete DMEM and seeded into wells of a 12-well culture plate (2×10^5 cells/well). The plate was incubated at 37°C in a 5% CO_2 atmosphere for 5 days. The culture supernatant containing the HCV infectious particles was concentrated by using a centrifugal device (Pall, Port Washington, NY, United States). The virus titer was determined by focus-forming assay (FFA).

DENV (serotypes 1-4), ZIKV, and JEV were propagated in Vero cells maintained in complete DMEM at 37°C in a 5% CO_2 incubator for 3−5 days. The culture supernatants containing the viruses were collected, and the virus titers (pfu/mL) were determined by the plaque-forming assay (PFA). The viruses were kept in small portions at −80°C as the stocks.

The enteroviruses (EV71 and CVA16) were propagated in RD cells grown in complete DMEM at 37°C in a 5% CO_2 atmosphere (Phanthong et al., 2020). The cytopathic effect (CPE) characterized by cell rounding, clumping, and/or detaching was observed daily. The culture was harvested (both cells and spent medium) when the CPE was at maximum and subjected to three freeze-thaw cycles, centrifuged to remove the cell debris, and the supernatant containing the virus was kept in small aliquots at –80°C as the virus stocks. The cell culture infectious dose 50 ($CCID_{50}$) of the virus stock was determined (Phanthong et al., 2020). Briefly, the virus stock was 10-fold serially diluted in complete DMEM and then added to the wells of 96-well culture plates. RD cells (2×10^4 cells) were added to each virus-containing well; the plate was incubated at 37°C in a 5% CO_2 atmosphere until the CPE was clearly observed. The Kärber formula (World Health Organization [WHO], 2004) was used to calculate the virus $CCID_{50}$ (10^x/mL) for each viral stock.

Porcine epidemic diarrhea virus (PEDV) was propagated in the permissive Vero cells as for the DENV, ZIKV, and JEV for 2 days. The amount of the PEDV in the harvested cell spent medium was determined by the plaque (syncytial)-forming assay (Thavorasak et al., 2022). The virus was kept at −80°C in small portions until use.

Preparation of Cell-Penetrating Human Superantibody to Hepatitis C Virus RNA-Dependent RNA Polymerase

HB2151 *E. coli* clone 34 that carried pCANTAB5E phagemid with inserted gene sequence coding for human single-chain antibody variable fragment (*huscfv*) specific to HCV RdRp (HuscFv34) was generated previously by using phage display technology (Thueng-In et al., 2014). Recombinant HCV NS5BΔ55 (RdRp) protein was used as an antigen in the phage bio-panning to select out the antigen-bound phage clones from the HuscFv phage display library (Kulkeaw et al., 2009). One of the HB2151 *E. coli* clones (clone 34) infected with the antigen-bound phage produced HuscFv (HuscFv34) that inhibited the HCV RdRp activity *in vitro* (Thueng-In et al., 2014) and when the HuscFv34 was linked to a cell-penetrating peptide (CPP), i.e., penetratin (PEN), the PEN-HuscFv34 could enter the Huh7 cells (being superantibody). The superantibody not only inhibited HCV replication *ex vivo* but also rescued the host's innate immunity from the HCV suppression (Thueng-In et al., 2014).

In this study, the *huscfv* from the recombinant *huscfv*-phagemid of the HB2151 *E. coli* clone 34 was subcloned to recombinant pET23b+ plasmid backbone carrying a DNA insert coding for a protein transduction domain/cell-penetrating peptide, penetratin (PEN) (Poungpair et al., 2010), and the DNA construct was introduced to BL21(DE3) *E. coli.* Non-chromatographic purification of the *E. coli* inclusion body (IB) was used to isolate the PEN-HuscFv34 from the bacterial cells grown under IPTG induction conditions. Four grams of the bacterial pellet was resuspended with 20 mL of 1 × BugBuster® protein extraction reagent (Millipore, Merck KGaA, Darmstadt, Germany) dissolved in 50 mM tris(hydroxymethyl)aminomethane (Tris; Millipore, Merck KGaA), pH 8.0. After the bacterial pellet was completely resuspended, Lysonase™ bioprocessing reagent (Millipore, Merck, KGaA) was added at 10 μL per gram of the wet bacterial pellet. After 20 min at room temperature (25 ± 2°C) on a slow setting shaking platform, the soluble fraction was removed from the preparation by centrifugation at 8000 × *g* for 30 min. The insoluble contents was washed with wash-100 reagent [50 mM phosphate buffer, pH 8.0, 500 mM sodium chloride (Kemaus, CherryBrook, NSW, Australia), 5 mM ethylenediaminetetraacetic acid (Kemaus), 8% (*v/v*) glycerol (Kemaus), and 1% (*v/v*) Triton X-100 (USB, Affymetrix, Thermo Fisher Scientific)] at 25°C and wash-114 reagent [50 mM Tris-HCl pH 8.0, 500 mM sodium chloride, 1% (*v/v*) Triton X-114 (Sigma Aldrich, St. Louis, MO, United States)] at 4°C. The preparation was spun down at 8000 × *g* for 30 min and the supernatant was removed. The inclusion body was washed with deionized distilled water and collected by centrifugation at 8000 × *g* for 30 min. The PEN-HuscFv34 was retrieved from the inclusion body by solubilization and refolding. The inclusion body solubilization was performed by dissolving the inclusion body in 50 mM *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS) (Sigma Aldrich, St. Louis, MO, United States) buffer, pH 10.8 supplemented with 0.3% (*w/v*) sodium lauroyl sarcosinate (Sigma Aldrich, St. Louis, MO, United States) and 1 mM dithiothreitol (DTT; USB, Affymetrix) at a protein concentration of 1 mg/mL. Solvation of the inclusion body was performed at room temperature for 15 min followed by keeping at 4°C for 16 h. The non-solubilized part was removed by centrifugation at 10,000 × *g* for 10 min. The preparation was immediately refolded by buffer exchange against 20 mM imidazole, pH 8.5 with and without 0.1 mM DTT. The refolded PEN-HuscFv34 was subsequently verified by SDS-PAGE and Coomassie Brilliant Blue G-250 (CBB) staining.

Verification of the Cell-Penetrating Ability of the Penetratin-HuscFv34

Human hepatocellular carcinoma cells (1×10^5 cells) in complete DMEM were seeded onto a cover glass placed in a well of 24-well

cell culture plate and incubated at 37°C in a 5% CO_2 atmosphere overnight. The established cell monolayer was added with the PEN-HuscFv34 prepared from the *E. coli* inclusion body and kept at 37°C in a 5% CO_2 atmosphere for 24 h. The cells were washed with PBS, fixed with 4% paraformaldehyde in PBS, and permeated with 0.1% Triton X-100 (USB, Affymetrix) in PBS. The cells were blocked with 5% BSA in PBS at room temperature for 20 min. After the excess BSA was removed by washing with PBS, rabbit anti-HuscFv34 was added to the cell monolayer and incubated for 1 h. Goat anti-rabbit Ig-AlexaFlour488 (1: 200; Thermo Fisher Scientific) was used as the secondary antibody, and DAPI was used to locate nuclei. After washing, the cells were mounted and observed under a confocal microscope (Nikon, Melville, NY, United States) for intracellular PEN-HuscFv34.

Biocompatibility of the Penetratin-HuscFv34/Superantibody to Mammalian Cells

Mammalian cells including A549, Huh7, Vero, and Vero E6 cells (4×10^4) were seeded separately in a 96-well white plate (Corning, Thermo Fisher Scientific) and incubated at 37°C in the CO_2 incubator overnight. The fluids were discarded; the cells were replenished with a culture medium containing PEN-HuscFv34 (0.25, 0.5, 1.0, 1.5, and 2.0 μM) and kept at 37°C in the CO_2 incubator overnight. Cytotoxicity of the superantibody was determined by using Cytotox-Glo™ Cytotoxicity Assay (Promega, Madison, WI, United States). The assay buffer provided with the kit was added to each well (50 μL/well), and the plate was kept at room temperature for 15 min. Experimental dead cell luminescence was detected by using Multidetection Microplate Reader Synergy H1 (Biotek, Agilent Technology, Santa Clara, CA, United States). Lysate reagent of the test kit was then added to all wells (50 μL/well), and the plate was placed on an orbital shaker (100 rpm) for 15 min. Total dead cell luminescence was detected, also by the microplate reader. Viable cell luminescence (Test luminescence) was calculated: Test luminescence = Total dead cell luminescence – Experimental dead cell luminescence. Percent cell viability was calculated: (Test luminescence ÷ Normal cell luminescence) × 100.

RNA Virus Replication Inhibition Mediated by Superantibody

Ten micrograms of HCV-JFH1 RNA was transfected into Huh7 cells by electroporation. The transfected cells were immediately seeded to 12-well cell culture plate (2×10^5 cells/well) and incubated at 37°C, 5% CO_2 for 6 h. After washing the cells, the complete DMEM containing various concentrations of superantibody (0.25, 0.5, 1.0, 1.5, and 2.0 μM) or medium alone was added. The treated cells were cultured at 37°C in a 5% CO_2 atmosphere for 5 days. The RNAs were extracted from the treated cells for viral RNA quantification by real-time PCR; the HCV infectious particles in the culture supernatants were enumerated by focus-forming assay (FFA).

Vero cells (3×10^5 cells) were seeded to 12-well cell culture plates and incubated at 37°C in a 5% CO_2 incubator overnight. DENV (serotypes 1-4), ZIKV, and JEV at MOI 0.1 and PEDV at MOI 0.0005 were added individually to the Vero cells in different wells and incubated further for 1 h. After washing with plain DMEM, complete DMEM containing the PEN-HuscFv34/superantibody (0.25, 0.5, 1.0, 1.5, and 2.0 μM) or medium alone (negative control) were added to appropriate wells containing the infected cells and incubated for 24 h for the PEDV and 48 h for the other viruses. The culture supernatants and cells were collected, and RNA was extracted from each cell sample and subjected to real-time RT-PCR for viral RNA quantification. Infectious virus particles in the culture supernatant samples were enumerated by the plaque-forming assay (PFA).

The RD cells were transfected with 100 $CCID_{50}$/mL of EV71 strain BrCr or MOI 0.1 of CVA16. After 1 h incubation at 37°C in a 5% CO_2 incubator, cells were washed one time with plain DMEM and replaced with complete DMEM containing superantibody (0.25, 0.5, 1.0, 1.5, and 2.0 μM) or medium alone as a negative control. Cells were incubated further for 24 h. Then, RNA was extracted from the collection and the viral RNA was quantified by real-time RT-PCR. Infectious virus particles in the culture supernatants were enumerated by PFA.

For SARS-CoV-2 replication inhibition, 1.5×10^5 cells of Vero E6 cells were seeded to wells of 24-well cell culture plates and incubated at 37°C, 5% CO_2 for 24 h. The plates were moved to the BSL-3 room to perform all the subsequent processes of the experiment. The seeded cells were infected with SARS-CoV-2 [Wuhan wild type, alpha (B.1.1.7), beta (B.1.351), delta (B.1.617.2), and omicron (B.1.1.529)] at 50 PFU/well. After 1 h incubation, the supernatants were removed and replenished with the superantibody (0.25, 0.5, 1.0, 1.5, and 2.0 μM) containing DMEM supplemented with the 2% FBS. The treated cells were incubated at 37°C, 5% CO_2 for 18 h. The RNAs were extracted from the cells for the real-time RT-PCR, and the culture supernatants were collected to detect the infectious particles by PFA for the Wuhan wild type and α, β, and δ variants and by FFA for the omicron variant (their plaques in the PFA were too tiny to be counted accurately).

Real-Time RT-PCR

The RNAs from the superantibody/medium-treated infected cells were extracted using TRIzol® reagent (Invitrogen). The amounts of viral RNA were quantified by real-time RT-PCR using a 1-step brilliant III SYBR green RT-qPCR master mix (Agilent Technologies). The real-time RT-PCR primers for each virus and house-keeping gene control are listed in **Supplementary Table 1**. The copy numbers of viruses were calculated from the Cq value using a comparative method.

Plaque-Forming Assay

The Vero or Vero E6 cells were seeded into wells of 24-well-culture plates (1.5×10^5 cells per well) and kept in humidified 5% CO_2 incubator at 37°C overnight. The virus-containing samples were diluted 10-fold serially, and 250 μL aliquots were added to the wells containing the cell monolayer. Experiments involving SARS-CoV-2 were performed in BSL-3. The plates were incubated further for 1 h; the fluids were discarded; the infected cells were rinsed with sterile PBS before adding with 1.5% carboxymethyl cellulose (CMC) (Sigma Aldrich, St. Louis, MO, United States) in complete DMEM and the plates were incubated further for 3 days (SARS-CoV-2) or 7 days (DENV, ZIKV, and

JEV). After incubations, the infected cells were fixed with 10% formaldehyde at room temperature for 1 h (2 h for SARS-CoV-2). The cells were washed with distilled water five times to get rid of the CMC and stained with 1% crystal violet in 10% ethanol at room temperature for 15 min. After washing with distilled water, the plates were dried, and plaques were counted visually. The amount of the virus in the original sample was calculated: PFU/mL = plaque number/(infection volume × dilution factor).

The RD cells were seeded on a 24-well culture plate (1.5×10^5 cells per well) and incubated at 37°C in a 5% CO_2 incubator overnight. After discarding the supernatant, the 10-fold serially diluted supernatants of the superantibody-mediated virus replication inhibition experiments were added into appropriate RD cell-containing wells, and the plates were incubated further for 1 h. The fluids were removed and 1.5% CMC in complete DMEM was added to each well and incubated further for 72 h. The cells were fixed with formalin and stained with crystal violet dye as described earlier. Plaque number were counted by eyes, and the number of viruses in the original sample was calculated.

For PEDV, after incubating with 10-fold diluted samples, the extracellular fluids were discarded; the cells were rinsed with sterile PBS, added with 1.5% CMC (Sigma Aldrich, St. Louis, MO, United States) in DMEM containing N-tosyl-L-phenylalanyl chloromethyl ketone (TPCK) trypsin, and incubated at 37°C in a 5% CO_2 atmosphere for 2 days. The cells were fixed with formalin and stained with crystal violet dye as described earlier. The CPE (syncytial formation) was enumerated under a microscope (40 × magnification), and the number of viruses in the original preparation was calculated.

Focus-Forming Assay

Vero E6 cells (4×10^4 cells/well) were seeded to wells of a 96-well cell culture plate and incubated at 37°C, 5% CO_2 for 24 h. The cell-seeded plates were moved to the BSL-3 room. The samples containing viruses (SARS-CoV-2 omicron variant) were 10-fold serially diluted. The fluids were removed from the cell-containing wells and replaced with 50 μL of the diluted virus samples. After 1 h incubation, the fluids were removed, replaced with 1.5% CMC in complete DMEM, and incubated further at 37°C, 5% CO_2 for 3 days. The CMC was removed from wells; the cells were fixed with 10% formaldehyde at room temperature for 2 h, washed with PBS three times, and permeated with 0.1% Triton X-100 in PBS at room temperature for 20 min. After washing two times with PBS, the cells were blocked with 5% BSA in PBS and stained with mouse anti-SARS-CoV-2 nucleoprotein antibody (1:5000) at room temperature for 1 h followed by incubating with goat anti-mouse IgG-HRP conjugate (SouthernBiotech, Birmingham, AL, United States). After the 1-h incubation, the cells were washed with PBS and the foci were developed by adding TMB sure blue substrate (SeraCare Life Sciences, Milford, MA, United States). The focal numbers were counted under an inverted microscope (40 × magnification). The numbers of foci (infectious virus particles) were calculated: FFU/mL = foci number/(infection volume × dilution factor).

For enumeration of HCV infectious particles, Huh7 cells (4×10^4 cells/well) were seeded to wells of a 96-well cell culture plate and incubated at 37°C, 5% CO_2 for 24 h. The fluids in all wells were discarded, and the cells were added with virus samples (culture supernatants from the superantibody-mediated inhibition of virus replication experiments) for 1 h; the fluids were discarded, the complete DMEM were replenished, and the infected cells were incubated for 3 days. The cells were fixed with 4% paraformaldehyde at room temperature for 20 min, washed, permeated with 0.1% Triton X-100, and blocked with 5% BSA in PBS. After blocking, the cells were probed with mouse anti-NS5A (Glab-ampai et al., 2017) for 1 h, washed, and added with goat anti-mouse Ig-alkaline phosphatase conjugate (Southern Biotech) for 1 h. After washing, the BCIP/NBT substrate (SeraCare Life Science) was added for color development. Numbers of foci were counted under an inverted light microscope, and the FFU/mL was calculated as mentioned earlier.

Computerized Simulation to Determine Presumptive Interaction Between the Viral RNA-Dependent RNA Polymerase and the Human Single-Chain Antibody Variable Fragment

Amino acid sequences of the HuscFv34 and three-dimensional (3D) structures of RdRp of DENV serotypes 1 and 4 and of PEDV were submitted for protein modeling using AlphaFold2 (Jumper et al., 2021) available in ColabFold's online notebook (Mirdita et al., 2022). Modeled 3D structure of the HuscFv34 was docked against existing crystal structures of RdRp of different viruses, [HCV (PDB ID: 1QUV), DENV serotype 2 (PDB ID: 6IZY), DENV serotype 3 (PDB ID: 2J7U), ZIKV (PDB ID: 6LD1), JEV (PDB ID: 4MTP), EV71 (PDB ID: 3N6L), CVA16 (PDB ID: 5Y6Z), and SARS-CoV-2 (PDB ID: 6M71)], and the predicted 3D structures of RdRp of DENV serotypes 1 and 4 and PEDV, *via* HADDOCK server version 2.4 (van Zundert et al., 2016). The parameters from the HADDOCK (HADDOCK scores, van der Waals energy, electrostatic energy, desolvation energy, restraint violation energy, buried surface area, and Z-Score) were collected. The intermolecular docking that showed the best HADDOCK score was selected. Pymol software (The PyMOL Molecular Graphics System, Version 2.5.2, Schrodinger, LLC, NY, United States) was used for building the molecular interactive protein structure models. The docked structures were further submitted to the PRODIGY server to predict the binding energy [ΔG (kcal per mol)] and K_d (M) at 25°C (Honorato et al., 2021).

Statistical Analysis

GraphPad Prism version 9 software (GraphPad Software, San Diego, CA, United States[1]) was used for the calculation of the half-maximal effective dose (EC_{50}) of the superantibody. Mean values and standard deviations (SD) of each treatment group from three independent experiments were compared using a one-way ANOVA. *P*-values of 0.05 or lower were considered statistically different: $p > 0.05$ (ns, not significant); $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***), and $p \leq 0.0001$ (****).

[1]www.graphpad.com

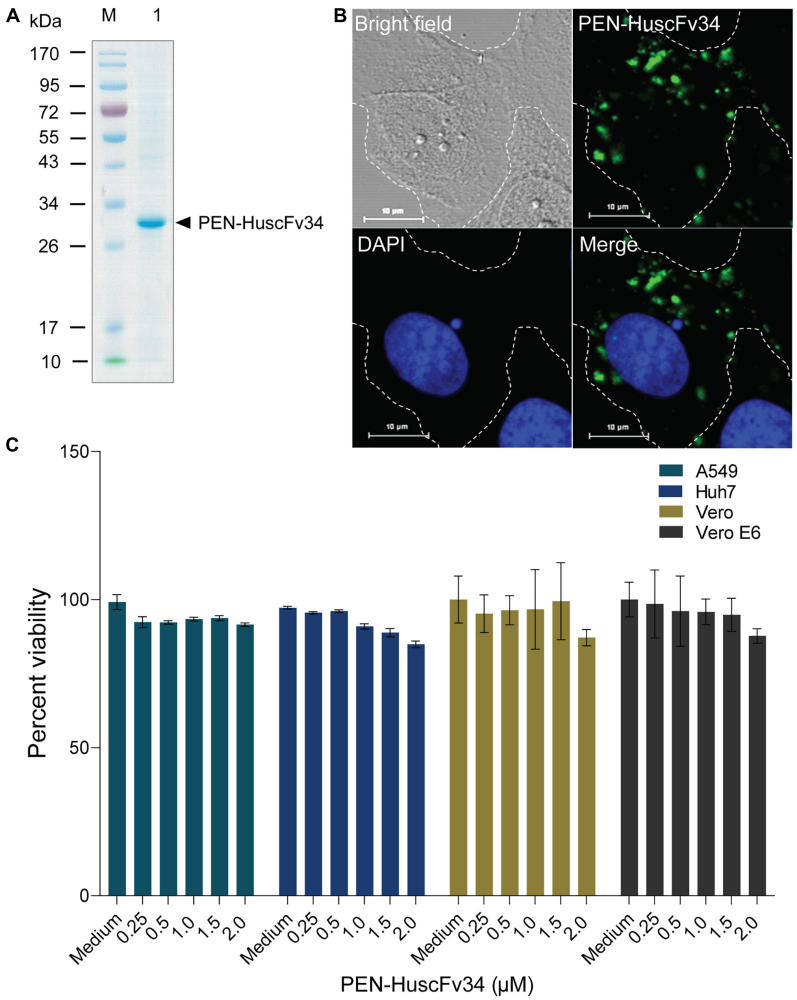


FIGURE 1 | Characteristics of the penetratin (PEN)-HuscFv34 to HCV RdRp. **(A)** Purified PEN-HuscFv34 after SDS-PAGE and CBB staining. Lane M, Protein molecular mass standard; Lane 1, SDS-PAGE-separated purified PEN-HuscFv34 stained by CBB dye (~30 kDa; arrowhead). **(B)** Cell-penetrating ability of the PEN-HuscFv34. The intracellular PEN-HuscFv34 stained green while nuclei are blue. **(C)** Biocompatibility of the PEN-HuscFv34 with mammalian cells including A549, Huh7, Vero, and Vero E6 cells. The PEN-HuscFv34 (superantibody) at the concentrations that were tested (0.25–2.0 μM) did not cause cytotoxicity to the cells. Percent viability of the cells (mean ± standard deviation) was not different from each other ($p > 0.05$).

RESULTS

Penetratin-Linked HuscFv34 to Hepatitis C Virus RNA-Dependent RNA Polymerase

Penetratin-linked human single-chain antibody variable fragments (PEN-HuscFv34) was produced from transformed BL21(DE3) *E. coli* grown under an IPTG-induced condition. The yield of the bacterial inclusion body obtained from the *E. coli* homogenate was 7.029 g/L of the bacterial culture. After the solubilization and refolding, the total protein obtained from each mg of the inclusion body was 680 μg. The CBB-stained SDS-PAGE-separated purified preparation revealed a single protein band with a relative mass of 30 kDa (**Figure 1A**), the correct molecular mass for PEN-HuscFv, in which each molecule consists of a VH domain linked to the VL domain *via* the $(Gly_4Ser)_3$ linker plus 16 amino acids (RQIKIWFQNRRMKWKK) of penetratin (Poungpair et al., 2010). As shown in **Figure 1B**, the PEN-HuscFv34 (green) could enter the mammalian cells (being cell-penetrable antibody/superantibody). The superantibody did not cause cytotoxicity to the mammalian cells that were tested, as determined by using the CytoTox-Glo™ Cytotoxicity Assay, based on the Practical Guide to ISO 109903-5 (Wallin, 1998; **Figure 1C**). The superantibody-treated cells appeared unchanged in their morphology under the light microscope (data not shown).

Inhibition of RNA Virus Replication by Superantibody

The ability of the superantibody specific to HCV RdRp (PEN-HuscFv34) in inhibiting replication of the homologous virus and

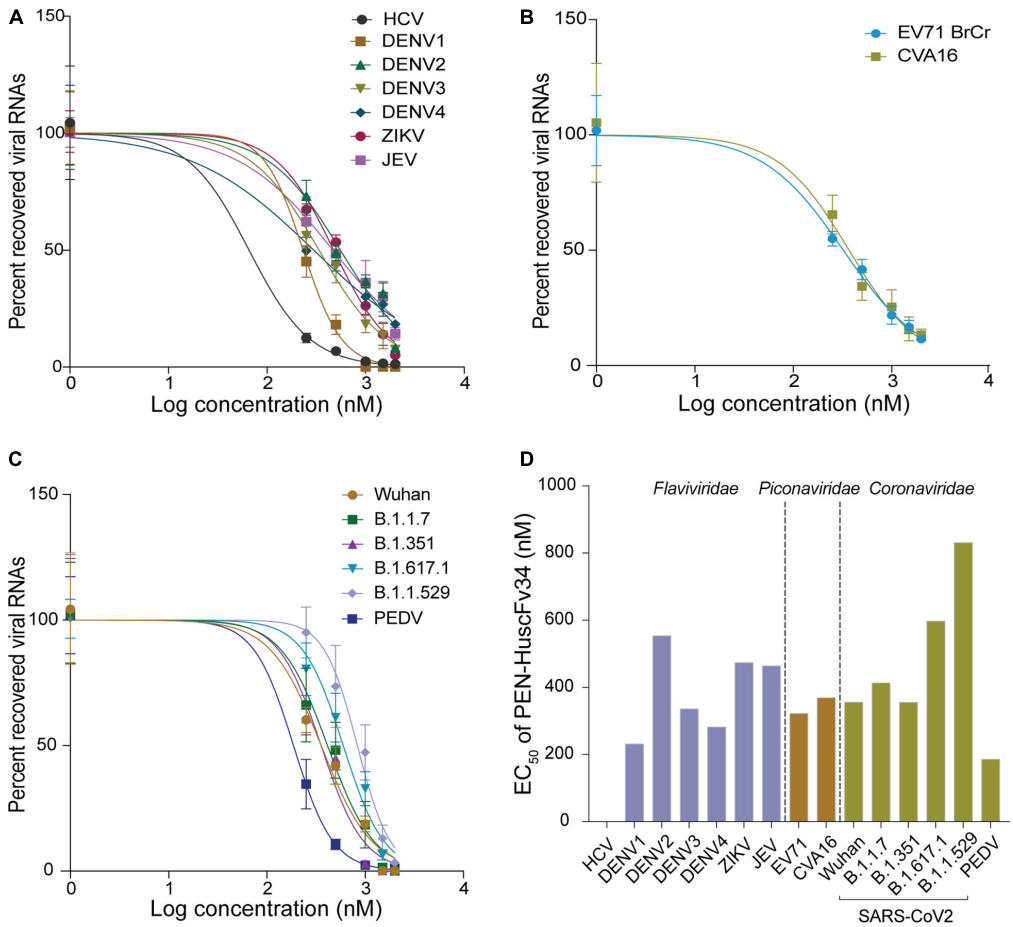


FIGURE 2 | Inhibition of the plus-sense RNA virus replication by the superantibody shown as percent recovered viral RNAs inside the superantibody-treated infected cells, when compared with the infected cells in the medium alone. **(A)** Viruses of the family *Flaviviridae* (HCV, DENV1–4, ZIKV, and JEV); **(B)** Enteroviruses of the family *Picornaviridae* (EV71 and CVA16). **(C)** Members of the family *Coronaviridae* (genus *Betacoronavirus*: SARS-CoV-2 Wuhan wild-type and variants of concerns: α, β, δ, and omicron; and genus *Alphacoronavirus*: PEDV). **(D)** Half-maximal effective dose (EC_{50}) of the superantibody against the tested viruses.

other plus-sense RNA viruses in the family *Flaviviridae* (DENV1-4, ZIKV, and JEV), *Picornaviridae* (EV71 and CVA16), and *Coronaviridae* (PEDV and SARS-CoV-2) was determined. Cells infected with the respective viruses were treated with different concentrations (0.25−2.0 μM) of the HCV-RdRp-specific superantibody or medium alone (negative control); the treated cells were subjected to real-time RT-PCR for quantification of the intracellular viral mRNAs, and their respective culture fluids were tested by PFA/FFA for enumeration of the released infectious particles. The superantibody could inhibit replication of the viruses that were tested in a dose-dependent manner as indicated by the percent reduction of the viral RNAs inside the infected cells compared to negative replication inhibition (medium) (**Figures 2A–C**). The superantibody also mediated the reduction of the numbers of the infectious particles released into the cell culture supernatants (**Figures 3A–C**). The EC_{50} of the superantibody on individual studied viruses is summarized in **Figure 2D** and **Table 1**. In the experiments, positive inhibition controls for individual viruses were not included as there are no approved direct-acting drugs/agents for certain viruses: DENV, ZIKV, JEV, EV71, CVA16, and PEDV. Details of the fold reduction of virus RNA from infected cells treated with the medium containing different concentrations of superantibody when compared with infected cells treated with the medium alone are shown in **Supplementary Figures 1–3**, and details of the reduction of released infectious viral particles (PFU/mL or FFU/mL) from the virus-infected cells treated with the medium-containing different concentrations of superantibody to RdRp when compared with infected cells treated with the medium alone are shown in **Supplementary Figures 4–6**.

Computerized Simulation to Determine Presumptive Interaction Between the Viral RNA-Dependent RNA Polymerase and the HuscFv34

In this study, the *in-silico* interactions of the homology modeled HuscFv34 3D structure with RdRp of the HCV and other plus-sense RNA viruses were determined by using the available crystal structures of HCV, DENV serotypes 2 and 3, ZIKV, JEV,

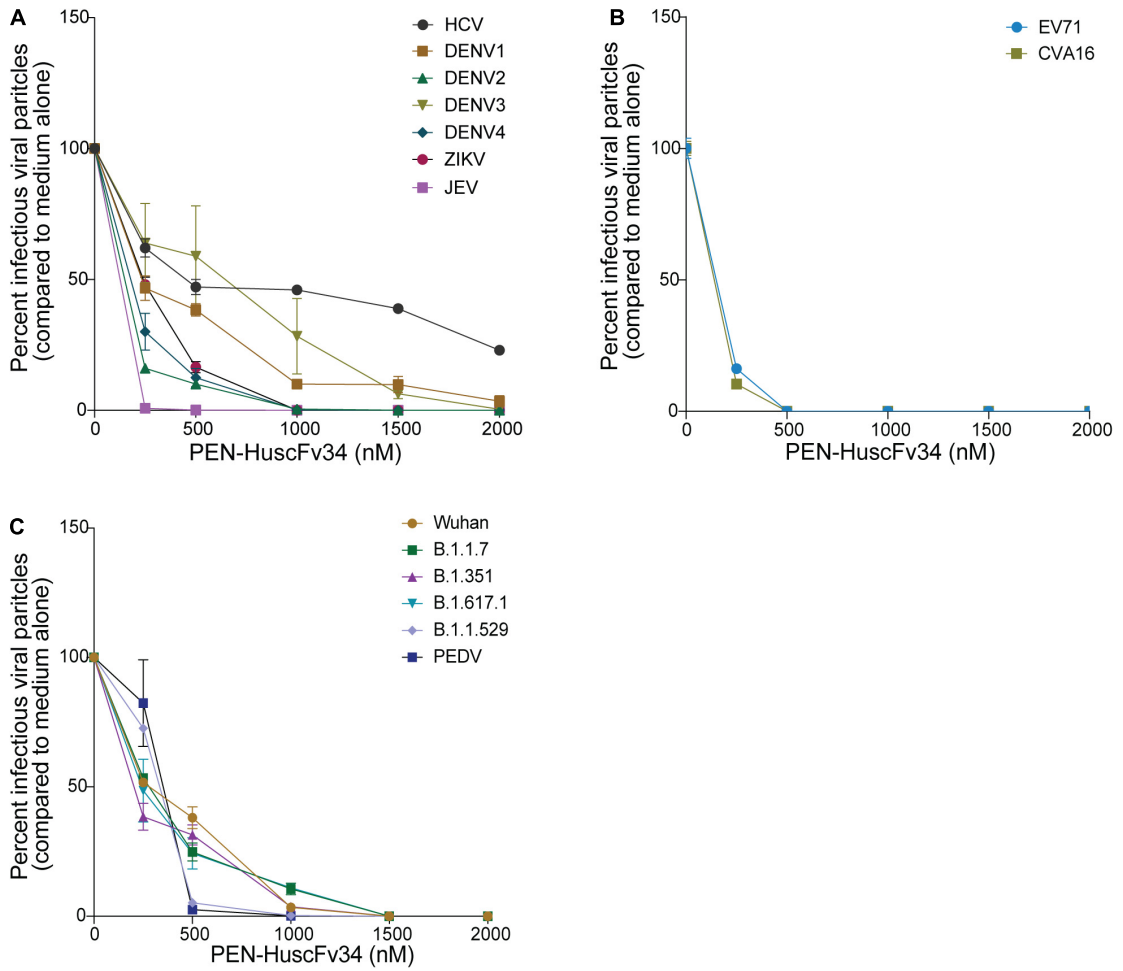


FIGURE 3 | Reduction of the infectious viral particles released from the infected cells after treatment with various concentrations of the superantibody when compared with infected cells in the medium alone. **(A)** Viruses of the family *Flaviviridae*; **(B)** Enteroviruses of the family *Picornaviridae*; **(C)** Viruses of the family *Coronaviridae*.

TABLE 1 | EC_{50} (nM) of penetratin (PEN)-HuscFv34 in replication inhibition of the tested viruses.

Flaviviridae							
Virus name	HCV	DENV1	DENV2	DENV3	DENV4	ZIKV	JEV
EC_{50}	65.6	232	553.6	336.3	282.5	473.9	464.4
Picornaviridae							
Virus name	EV71	CVA16					
EC_{50}	322.4	369.6					
Coronaviridae							
Betacoronavirus						*Alphacoronavirus*	
SARS-CoV-2						PEDV	
Variant	Wuhan	α (B.1.1.7)	β (B.1.351)	δ (B.1.617.1)	omicron (B.1.1.529)	GII	
EC_{50}	356.4	413.4	355.7	597.7	831.6	186.3	

EV71, CVA16, and SARS-CoV-2 and modeled 3D structures of DENV serotypes 1 and 4 and PEDV, for which the crystal structures were not yet available. The data for computerized prediction of HuscFv34 and RdRp models and their interaction are summarized in **Supplementary Table 2**. The computerized models of interaction between the HuscFv34 and the RdRp of the studied viruses are shown in **Figure 4**. The details on the residues and domains of the RdRp of the viruses that formed contact interface with residues in the CDRs of the HuscFv34 are given in **Table 2**.

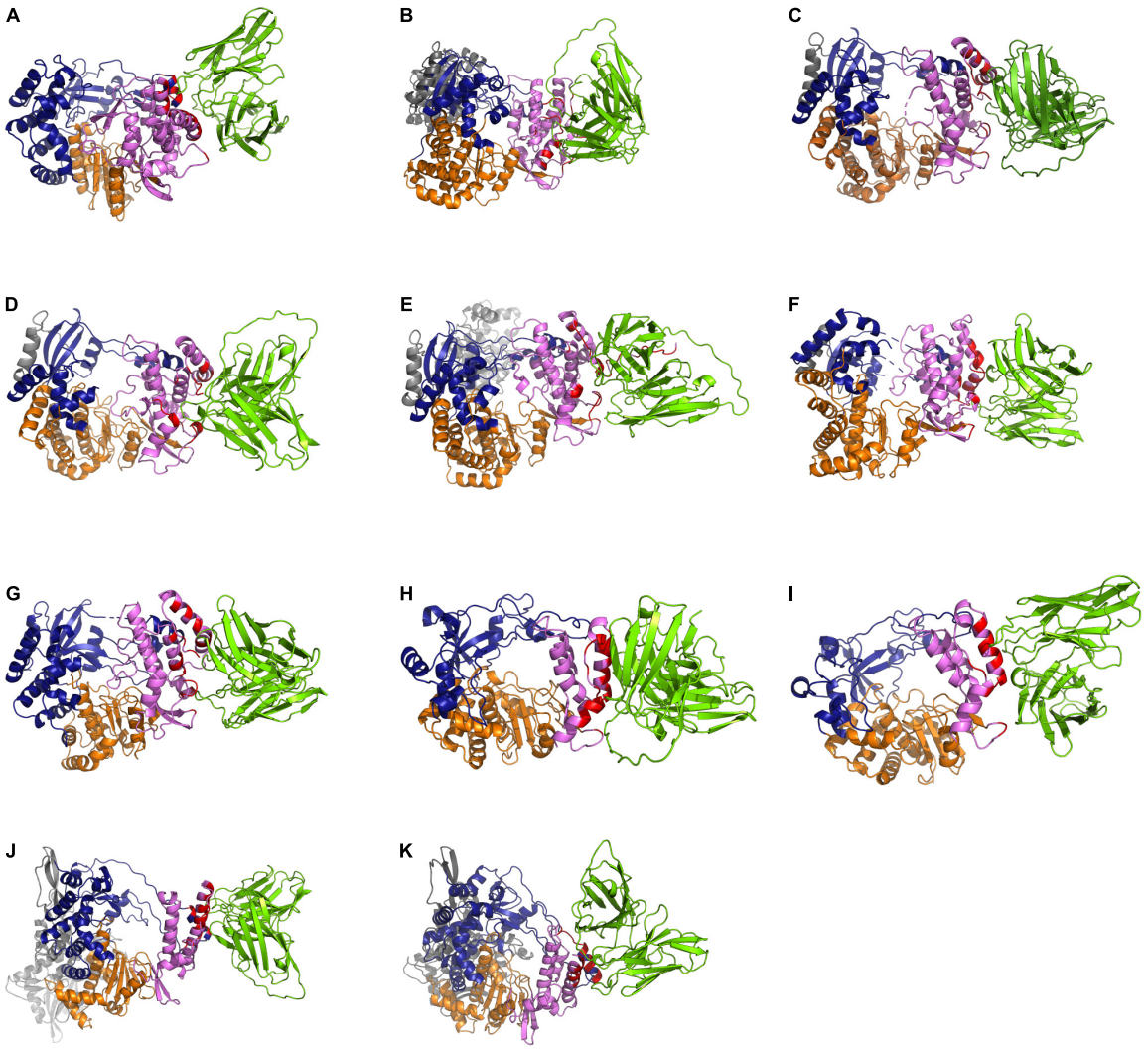


FIGURE 4 | Computerized models of interaction between HuscFv34 and viral RdRp. **(A–G)** Viruses of the family *Flaviviridae* (HCV, DENV1–4, ZIKV, and JEV); **(H,I)** viruses of the family *Picornaviridae* (EV71 and CVA16); and **(J,K)** viruses of the family *Coronaviridae* (SARS-CoV-2 and PEDV). The RdRp are shown as cartoons: fingers domains (deep blue), palm domains (orange), and thumb domains (pink). The cartoon colored in red represents the contact interface between the HuscFv34 (green cartoon structure) and the target RdRp. Gray cartoons in DENV, ZIKV, and JEV are N-terminal *S*-adenosyl methionine methyltransferases (MTases). Gray cartoons in SARS-CoV-2 are the beta-hairpin that sandwiches with the palm domain, the Nidovirus-specific extension domain (NIRAN) domain, and the interface subdomain of the viral nsp12.

DISCUSSION

RNA-dependent RNA polymerase (RdRp) is an inscribed protein of RNA viruses that is indispensable for the virus replication cycle. The protein is a principal component of the replicase/transcriptase complex that generates new genomic RNA and virus proteins which assemble to form virus progeny for further spread. RdRp is structurally conserved among the RNA viruses with no human homolog. Therefore, it is a potential target for a pan-anti-RNA virus agent. Currently, several small chemical inhibitors (both nucleoside and non-nucleoside inhibitors) that target the RdRp have been developed and tested for the treatment of the RNA virus infections; some of them have been approved and launched for clinical use while the others are at various phases of clinical trials or were discontinued (Tian et al., 2021). Examples of the nucleoside inhibitors that target RdRp are sofosbuvir (Sovaldi/PSI-7977/GS-7977) for the treatment of hepatitis B and C, favipiravir (T-705/Avigan/Favipiravir, Favilavir) for the treatment of influenza (repurposed for COVID-19 treatment), ribavirin (ICN-1229/Tribavirin) for the treatment of influenza, hepatitis C, and respiratory syncytial virus (RSV) infection (repurposed for the treatment of SARS in 2003 and COVID-19), and remdesivir for COVID-19 and other infections. More recently, a few non-nucleoside inhibitors of RdRp have been launched for the treatment of hepatitis C including dasabuvir (Exviera/Viekira Pak/Viekira XR/ABT-333) and lomibuvir (VX-222/VCH-222) (Tian et al., 2021). Limitations of the chemical inhibitors besides their off-target

and adverse side effects such as teratogenicity, hemolytic anemia, gastro-intestinal disturbance, and others that preclude patients' compliance are their susceptibility to virus mutation; thus, often they must be used in combined medication among themselves, with other drugs or an interferon, for the treatment of viruses of the drug-resistant phenotypes, such as genotype I HCV.

Antibodies have been used for the treatment of human diseases including infectious, non-infectious, and toxin/venom-mediated maladies. For safety issues, the therapeutic antibodies or antibodies for passive immunization should have negligible immunogenicity in the recipients, implying that the fully human isotype is the safest antibody format. Although the penetratin (PEN) that has been linked to the HuscFv34 is derived from the third helix of *Drosophila* Antennapedia homeodomain protein (Derossi et al., 1994), it has been shown that dendritic cells (DCs) pulsed with this peptide could not activate autologous T cells, implying that the peptide is not immunogenic (Brooks et al., 2015). Currently, the PEN has been used in several vaccine studies to deliver tumor-associated antigens into antigen-presenting cells (APCs), and as a non-viral gene delivery vehicle in DNA vaccines, as well as carrying therapeutic substances into cellular compartments (reviewed by Brooks et al., 2010; Yang et al., 2019). However, in preclinical and clinical trials, the immunogenicity and biocompatibility of the PEN-HuscFvs must be investigated.

Antibody uses several residues in multiple CDRs in synergistic binding to the target, causing difficulty for the pathogens to create an antibody-escape target mutant that retains the inherent functional activity, particularly the proteins that require high conservation. The main concern in using therapeutic antibodies in the treatment of the virus infection is the antibody-dependent enhancement (ADE) (Kulkarni, 2020) that often aggravates the morbidity. Conventional antibodies elicit ADE by different mechanisms. For *Flavivirus* infection, the Fc fragments of the virus-antibody complexes bind to the Fc-receptors and enhance the virus entry to myeloid cells, leading to increment of the virus replication and viral load (extrinsic ADE) (Khandia et al., 2018). The intracellular virus may inhibit type 1 interferon response and activates the production of interleukin-10 that causes a type 2 (Th2) immune response bias, which heightens virus production and release (intrinsic ADE); the intrinsic ADE enhanced more DENV replication than the extrinsic ADE (Narayan and Tripathi, 2020). For other viruses, including respiratory viruses such as RSV, influenza virus, and coronavirus, the bi-/multi-valent antibodies may form large immune complexes that activate complement, causing the release/formation of anaphylatoxins, chemotaxis, and membrane attack complexes (MAC) that recruit immune and inflammatory cells to the infected areas and exacerbate the tissue inflammation, cytokine storm, cellular apoptosis, and multi-organ damage, i.e., the so-called immune enhancement ADE (Sánchez-Zuno et al., 2021). The antibody may promote virus entry to host cells by other mechanisms besides the Fc-mediated; for SARS-CoV-2, non-neutralizing antibodies to an epitope in the N-terminal domain (NTD) of the S1 subunit of the spike protein promote an upstanding/open form of the RBD by cross-linking two adjacent spike trimers, which then

TABLE 2 | Residues and domains of the RdRp of the viruses that formed contact interface with the residues in CDRs of the HuscFv34.

HCV RdRp		HuscFv34		Interactive bond
Residue	**Region**	**Residue**	**Region**	
A25	Finger	V167	VL-CDR1	Alkyl
N28	Finger	Q164	VL-CDR1	Hydrogen
N28	Finger	G165	VL-CDR1	Hydrogen
S29	Finger	H168	VL-CDR1	Hydrogen
S29	Finger	H169	VL-CDR1	Hydrogen
R32	Finger	Q235	VL-CDR3	Hydrogen
R32	Finger	S137	VH-CDR3	Hydrogen
R32	Finger	P237	VL-CDR3	Hydrogen
R32	Finger	N138	VH-CDR3	Hydrogen
S431	Thumb	H169	VL-CDR1	Hydrogen
R490	Thumb	Q62	VH-CDR2	Hydrogen
R498	Thumb	N57	VH-CDR2	Hydrogen
R498	Thumb	T58	VH-CDR2	Hydrogen
V499	Thumb	F236	VL-CDR3	Pi-Alkyl
H502	Thumb	D33	VH-CDR1	Hydrogen
H502	Thumb	W50	VH-CDR2	Pi-Pi
R503	Thumb	D103	VH-CDR3	Hydrogen
R503	Thumb	H169	VL-CDR1	Pi-Alkyl
R503	Thumb	T234	VL-CDR3	Hydrogen
K531	Thumb	N54	VH-CDR2	Hydrogen
DENV1 RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
H800*	Thumb	G165	VL-CDR1	Hydrogen
T805*	Thumb	F236	VL-CDR3	Pi-Sigma
E806*	Thumb	Q164	VL-CDR1	Hydrogen
D807*	Thumb	H169	VL-CDR1	Hydrogen
D807*	Thumb	H168	VL-CDR1	Hydrogen
L809*	Thumb	H169	VL-CDR1	Hydrogen
S810	Thumb	V167	VL-CDR1	Hydrogen
S810	Thumb	H169	VL-CDR1	Hydrogen
S810	Thumb	H168	VL-CDR1	Hydrogen
R814	Thumb	G165	VL-CDR1	Hydrogen
V829	Thumb	H169	VL-CDR1	Pi-Anion
S830	Thumb	H169	VL-CDR1	Pi-Anion
S892	Thumb	N54	VH-CDR2	Electrostatic
D893	Thumb	S55	VH-CDR2	Electrostatic
L898	Thumb	D103	VH-CDR3	Hydrogen
W899	Thumb	D103	VH-CDR3	Hydrogen
DENV2 RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
K719	Thumb	Q235	VL-CDR3	Hydrogen
R770	Thumb	H169	VL-CDR1	Hydrogen
E834	Thumb	H168	VL-CDR1	Salt bridge,Electrostatic
E834	Thumb	H169	VL-CDR1	Hydrogen
E834	Thumb	R238	VL-CDR3	Electrostatic
Y838	Thumb	H169	VL-CDR1	Electrostatic
R856	Thumb	H169	VL-CDR1	Hydrogen
R856	Thumb	Y102	VH-CDR3	Pi-Alkyl
A860	Thumb	Y102	VH-CDR3	Hydrogen
K861	Thumb	G105	VH-CDR3	Hydrogen
K861	Thumb	D106	VH-CDR3	Salt bridge,Electrostatic
N868	Thumb	N54	VH-CDR2	Hydrogen
D881	Thumb	N54	VH-CDR2	Hydrogen
D881	Thumb	S55	VH-CDR2	Hydrogen
D881	Thumb	N57	VH-CDR2	Hydrogen

(Continued)

TABLE 2 | (Continued)

DENV3 RdRp		HuscFv34		Interactive bond
Residue	**Region**	**Residue**	**Region**	
T806*	Thumb	H169	VL-CDR1	Hydrogen
E807*	Thumb	H169	VL-CDR1	Hydrogen
D808*	Thumb	H169	VL-CDR1	Hydrogen,Electrostatic
D808*	Thumb	Y102	VH-CDR3	Hydrogen
T832	Thumb	Y104	VH-CDR3	Hydrogen
W833	Thumb	Y104	VH-CDR3	Hydrogen
E834	Thumb	Y104	VH-CDR3	Hydrogen,Pi-Anion
E834	Thumb	S31	VH-CDR1	Hydrogen
E834	Thumb	H32	VH-CDR1	Electrostatic, Hydrogen
A860	Thumb	S55	VH-CDR2	Hydrogen
Q861	Thumb	R72	VH-CDR2	Hydrogen
L864	Thumb	N57	VH-CDR2	Hydrogen
E878	Thumb	Q164	VL-CDR1	Hydrogen
E878	Thumb	Q235	VL-CDR3	Hydrogen
L880	Thumb	H168	VL-CDR1	Pi-Sigma
D881	Thumb	F236	VL-CDR3	Pi-Anion,Pi-Sigma
Y882	Thumb	H169	VL-CDR1	Hydrogen
M883	Thumb	N52	VH-CDR2	Hydrogen
DENV4 RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
K812	Thumb	D106	VH-CDR3	Hydrogen,Electrostatic
K812	Thumb	E108	VH-CDR3	Hydrogen,Electrostatic
P830	Thumb	T28	VH-CDR1	Hydrogen
H832	Thumb	H32	VH-CDR1	Hydrogen
H832	Thumb	G26	Electrostatic	
H832	Thumb	T28	VH-CDR1	Hydrogen,Electrostatic
E835	Thumb	Y104	VH-CDR3	Hydrogen
D836	Thumb	T30	VH-CDR1	Hydrogen
R872	Thumb	N170	VL-CDR1	Hydrogen
Y880	Thumb	N172	VL-CDR1	Hydrogen
D882	Thumb	N172	VL-CDR1	Hydrogen
P885	Thumb	Y102	VH-CDR3	Pi-Alkyl
R888	Thumb	N52	VH-CDR2	Hydrogen
E895	Thumb	H168	VL-CDR1	Hydrogen
E895	Thumb	Q235	VL-CDR3	Hydrogen
E895	Thumb	F236	VL-CDR3	Hydrogen
Y890	Thumb	Y102	VH-CDR3	Hydrogen
A892	Thumb	Y102	VH-CDR3	Pi-Alkyl
ZIKV RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
K721	Thumb	H169	VL-CDR1	Hydrogen
L776	Thumb	G171	VL-CDR1	Hydrogen
K843	Thumb	N170	VL-CDR1	Pi-Cation
K843	Thumb	Q192	VL-CDR2	Hydrogen
G854	Thumb	Y174	VL-CDR1	Hydrogen
A862	Thumb	Y102	VH-CDR3	Hydrogen
A862, E863	Thumb	Y102	VH-CDR3	Pi-Alkyl
E863	Thumb	F236	VL-CDR3	Hydrogen
E863	Thumb	V167	VL-CDR1	Pi-Alkyl
E863	Thumb	S137	VH-CDR3	Hydrogen
I865	Thumb	D103	VH-CDR3	Amide-Pi Stacked
K866	Thumb	D103	VH-CDR3	Hydrogen,Alkyl
K866	Thumb	Y104	VH-CDR3	Hydrogen,Pi-Alkyl
K866	Thumb	G105	VH-CDR3	Hydrogen
K866	Thumb	Y107	VH-CDR3	Hydrogen
K866	Thumb	D33	VH-CDR1	Hydrogen
D884	Thumb	S55	VH-CDR2	Salt bridge,Electrostatic
D884	Thumb	N57	VH-CDR2	Electrostatic

(Continued)

TABLE 2 | (Continued)

JEV RdRp		HuscFv34		Interactive bond
Residue	**Region**	**Residue**	**Region**	
K724	Thumb	Q235	VL-CDR3	Hydrogen
K724	Thumb	G165	VL-CDR1	Hydrogen
K724	Thumb	V167	VL-CDR1	Hydrogen
R775	Thumb	N170	VL-CDR1	Hydrogen
T839	Thumb	H168	VL-CDR1	Hydrogen
T839	Thumb	H169	VL-CDR1	Hydrogen
D840	Thumb	H169	VL-CDR1	Hydrogen
Y843	Thumb	H169	VL-CDR1	Hydrogen
K846	Thumb	N170	VL-CDR1	Hydrogen
K846	Thumb	G171	VL-CDR1	Hydrogen
Y869	Thumb	Y104	VH-CDR3	Hydrogen
R876	Thumb	N54	VH-CDR2	Hydrogen
D886	Thumb	N54	VH-CDR2	Hydrogen
T889	Thumb	N57	VH-CDR2	Hydrogen
T889	Thumb	D103	VH-CDR3	Hydrogen
T889	Thumb	D33	VH-CDR1	Hydrogen
EV71 RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
K427	Thumb	F236	VL-CDR3	Hydrogen
Q428	Thumb	Q164	VL-CDR1	Hydrogen
Q428	Thumb	G165	VL-CDR1	Hydrogen
Q428	Thumb	V167	VL-CDR1	Hydrogen
E431	Thumb	F236	VL-CDR3	Hydrogen
S435	Thumb	H168	VL-CDR1	Hydrogen
T436	Thumb	H169	VL-CDR1	Hydrogen
R438	Thumb	T58	VH-CDR2	Hydrogen
R444	Thumb	Y104	VH-CDR3	Pi-Cation
R444	Thumb	D33	VH-CDR1	Electrostatic
R444	Thumb	N57	VH-CDR2	Hydrogen
R444	Thumb	D103	VH-CDR3	Hydrogen,Electrostatic
L446	Thumb	N57	VH-CDR2	Hydrogen
CVA16 RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
H383	Thumb	N57	VH-CDR2	Pi-Sigma
H383	Thumb	T58	VH-CDR2	Hydrogen
H383	Thumb	G59	VH-CDR2	Hydrogen
Q384	Thumb	N54	VH-CDR2	Hydrogen
K427	Thumb	D33	VH-CDR1	Electrostatic
K427	Thumb	F236	VL-CDR3	Pi-Anion
E428	Thumb	P237	VL-CDR3	Hydrogen
E428	Thumb	Y102	VH-CDR3	Hydrogen
E428	Thumb	R238	VL-CDR3	Electrostatic
E431	Thumb	H169	VL-CDR1	Salt bridge,Electrostatic
E431	Thumb	H169	VL-CDR1	Pi-Alkyl
K432	Thumb	Q164	VL-CDR1	Hydrogen
V434	Thumb	H169	VL-CDR1	Hydrogen
S435	Thumb	F236	VL-CDR3	Pi-Alkyl
R438	Thumb	H169	VL-CDR1	Hydrogen
R438	Thumb	H168	VL-CDR1	Salt bridge,Electrostatic
N450	Thumb	D103	VH-CDR3	Salt bridge,Electrostatic
N450	Thumb	Q235	VL-CDR3	Hydrogen
PEDV RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
K412	Fingers	H169	VL-CDR1	Hydrogen
E413	Fingers	H169	VL-CDR1	Electrostatic,Hydrogen
E413	Fingers	Y102	VH-CDR3	Hydrogen

(Continued)

TABLE 2 | (Continued)

PEDV RdRp		HuscFv34		Interactive bond
Residue	**Region**	**Residue**	**Region**	
E420	Fingers	H169	VL-CDR1	Electrostatic,Hydrogen
E420	Fingers	Q235	VL-CDR3	Hydrogen
E420	Fingers	H168	VL-CDR1	Salt bridge,Electrostatic,Hydrogen
H885	Thumb	D103	VH-CDR3	Pi-Anion,Hydrogen
K888	Thumb	D103	VH-CDR3	Salt bridge,Electrostatic
K888	Thumb	D33	VH-CDR1	Electrostatic
K888	Thumb	W50	Pi-Cation	
K888	Thumb	Y104	VH-CDR3	Pi-Alkyl
N891	Thumb	Y104	VH-CDR3	Hydrogen
A892	Thumb	Y104	VH-CDR3	Pi-Alkyl
E896	Thumb	D106	VH-CDR3	Hydrogen
E896	Thumb	H32	VH-CDR1	Electrostatic
SARS-CoV-2 RdRp		**HuscFv34**		**Interactive bond**
Residue	**Region**	**Residue**	**Region**	
F415	Fingers	N54	VH-CDR2	Hydrogen
N416	Fingers	S55	VH-CDR2	Hydrogen
D418	Fingers	S55	VH-CDR2	Hydrogen
D418	Fingers	N52	VH-CDR2	Hydrogen
D418	Fingers	Y104	VH-CDR3	Hydrogen
K426	Fingers	Y102	VH-CDR3	Hydrogen
K849	Thumb	S55	VH-CDR2	Hydrogen
Q886	Thumb	H169	VL-CDR1	Hydrogen
R889	Thumb	H169	VL-CDR1	Hydrogen,Pi-Alkyl
K890	Thumb	D103	VH-CDR3	Salt bridge,Electrostatic
K890	Thumb	H169	VL-CDR1	Pi-Alkyl
D893	Thumb	H168	VL-CDR1	Salt bridge,Electrostatic
E894	Thumb	N57	VH-CDR2	Hydrogen
E894	Thumb	F236	VL-CDR3	Pi-Sigma

**Priming loop of the thumb domain.*

enhances the virus entry (Liu et al., 2021). For the influenza virus, the non-neutralizing antibody promotes the virus entry by increasing hemagglutinin stem flexibility and virus fusion to the cell membrane (Winarski et al., 2019). The antibodies may enhance entry of SARS-CoV-2 into monocytes/macrophages *via* the Fc receptors; nevertheless, the infection is abortive; instead, the virus induces a specific M2 macrophage transcriptional program and causes host immune paralysis for the benefit of COVID-19 progression and pathogenesis (Boumaza et al., 2021). In this study, the superantibody (PEN-HuscFv34) specific to intracellular RdRp that works inside the cells cannot bind to the Fc receptors on cells and cannot form large immune complexes (cannot activate complement) but inhibited the replication of RNA viruses across families, is offered for testing further as a safe and broadly effective anti-RNA virus agent. Usually, the superantibodies (the term coined by Charles Morgan, president of InNexus Biotechnology, Vancouver, WA, Canada) enter cells; if there is no target, they leave the cells and enter new cells. The superantibodies bind intracellular targets and eventually the antibody-bound substances are eliminated by the normal cell physiological process, including the ubiquitin-proteasome and/or autophagy. "The beauty of the sole human antibodies is that they have minimal, if there were any, immunogenicity; thus, they should be less or not toxic. Besides, they are highly discriminating, i.e., far more specific than small-molecule drugs" (Coghlan, 2022). They are more tolerable to target mutation than the small chemical drugs as they bind to several target sites by using many residues in multiple CDRs.

The HCV RdRp epitope bound by the HuscFv34 was identified previously (by phage mimotope search using 12-mer peptide phage display library and competitive peptide ELISA) as a conformational epitope that is composed of residues in the finger's tip of the finger domain and helix O of the thumb domain of the HCV RdRp (NS5B protein), which were juxtaposed upon the protein folding to form the roof of the active enzymatic groove (closed catalytic tunnel) (Thueng-In et al., 2014). There were three phage mimotopic peptides (mimotopes 1−3; M1-M3) derived from the mimotope search that matched with the stretched sequence of the HCV RdRp, including M1: ALPFMGYHNSVY matched with 22PISPLSNSLLRHHNLVY40 of the Δ1 loop of finger domain; M2: NYPATNTHRYTP matched with residues 470GLSAFTLHSYFT481; and M3: IPVKSWPIRPSS matched with residues 495PPLRAWRHRARA506 of the thumb domain (based on the identical, conserved, and semiconserved amino acid residues upon the pairwise alignment) (Thueng-In et al., 2014). In this study, the computerized simulation of the HuscFv34-HCV RdRp interaction was performed to verify the results of the previous finding. We did not model the interaction of the superantibody (PEN-HuscFv) with the target RdRps because the penetratin (PEN) was linked to the HuscFv by a flexible linker and another end of the PEN was free. Besides, the PEN itself is not structured. Therefore, it should be inappropriate to fix the PEN in the rigid model for modeling and intermolecular docking as in reality the PEN would move freely while the HuscFv would be the principal part involved in target binding. By the *in-silico* analysis, the HuscFv34 interacted with residues of the HCV RdRp fingers domain, i.e., A25, N28, S29, and R32, located at the finger's Δ1 extension loop [residues I11-S46)] that usually packs against the thumb domain to form active closing (form 1) of the HCV RdRp channel (Bressanelli et al., 1999). Binding of the HuscFv34 at the finger's Δ1 extension loop could disturb the conformation and rigidity of the enzymatic groove (Biswal et al., 2005). Besides the fingertip, the HuscFv34 also formed a contact interface with many residues at the back of the thumb domain. Previous evidence has shown that interaction of the HCV NS5B (RdRp) with a host component, nucleolin, is indispensable for HCV replication (Shimakami et al., 2006). Residue W500 and three arginines (R498, R501, and R503) at the armadillo-like arm repeats of the thumb domain (Bressanelli et al., 1999) are important for the nucleolin binding and the HCV replication (Kusakawa et al., 2007). The HuscFv34 interaction with several residues in this region of the thumb domain (shown in **Figure 4A** and **Table 2**) may interfere with the RdRp-host nucleolin interaction, hence HCV replication inhibition.

For dengue viruses, the RdRp is located at the C-terminal residues 270 to 900 of the bifunctional NS5 protein that contains 900 amino acids (the N-terminal residues of the NS5 form the enzyme S-adenosyl methionine transferase) (Yap et al., 2007). The thumb domain (residues 706−900) of the RdRp contains a motif (motif E/primer grip) that lies between the palm domain and α-helices of the thumb domain (Yap et al., 2007). There is a loop that spans amino acids 782 to 809 of the thumb

domain, called a priming loop. The priming loop together with another loop of the finger domain form the roof of the tunnel that regulates RNA entry and exit from the RdRp active site (Yap et al., 2007). Several residues of the priming loop protrude into the RdRp active groove and stabilize the NTPs on the RNA template at the initial stage of the *de novo* RNA synthesis; these residues also pad alongside the RNA template during the process of the RNA synthesis (Yap et al., 2007; Gong and Peersen, 2010). The thumb domain is also involved in the motility of the newly synthesized RNA. Various interactive bonds (hydrogen, salt bridge, stacking interaction) between amino acid residues of the priming loop, including Thr794 and Ser796, Glu807 and Arg815, and Arg749 and Trp787, contribute to maintaining the orientation of the RdRp protein (Yap et al., 2007). From the *in-silico* prediction, the HuscFv34 interacted with several residues in the priming loops of DENV1 and DENV3 of the thumb domain (asterisks in **Table 2**) that may interfere with their functional activity and/or cause a structural change of the protein, leading to impairment of the RdRp activity, hence the DENV replication inhibition.

From the *in-silico* analysis, the HuscFv34 is also predicted to form interaction with many residues at the back surface of the thumb domains of DENV1, DENV3, PEDV, and SARS-CoV-2 and interacted solely with the C-terminal helices of thumb domains of DENV2, DENV4, ZIKV, JEV, EV71, and CVA16, which could be the site of the polymerase interaction with other viral/host cellular proteins during the formation of the replication/transcriptase complex and replication initiation (Bressanelli et al., 1999). Several non-nucleoside chemical inhibitors have been shown to bind to allosteric sites on the outer surface of the thumb subdomain (Thumb II or T2) and cause changes in the spatial conformation of the enzyme, rendering it inactive and reducing the viral load (Le Pogam et al., 2006; De Clercq, 2013; Li et al., 2016; Lim et al., 2016; Tian et al., 2021).

The EC_{50} of the superantibody (PEN-HuscFv34) was found in the nanomolar range for all of the tested RNA viruses, ranging from 65.6 nM for the homologous HCV to 831.6 for SARS-CoV-2 omicron variant, which was comparable to the chemical nucleoside and non-nucleoside inhibitors: favipiravir EC_{50} for SARS-CoV-2 was 61.88 μM (Wang et al., 2020); cytosine analog (NHC, EIDD-1931) EC_{50} values for SARS-CoV-2 and MERS-CoV were 0.3 and 0.56 μM, respectively (Sheahan et al., 2020); EC_{50} values of remdesivir (GS-5734) in inhibiting SARS-CoV and MERS-CoV in human airway epithelial cells (HAE) were 0.069 and 0.07/0.074 μM, respectively (Sheahan et al., 2017; Agostini et al., 2018), and SARS-CoV-2 in Vero E6 cells were 0.77 μM (Wang et al., 2020) and 23.15 μM (Choy et al., 2020); EC_{50} value of ribavirin in inhibiting SARS-CoV-2 in Vero E6 cells was 109.5 μM (Wang et al., 2020; Frediansyah et al., 2021).

CONCLUSION

The cell-penetrating human single-chain antibody variable fragments (superantibody) specific to NS5B (RdRp) of HCV that were found previously to inhibit the HCV replication and that rescued the HCV suppressed host innate (anti-viral) immunity were tested against other RNA viruses of the families *Flaviviridae* (DENV1-4, ZIKV, JEV), *Picornaviridae* (EV71 and CVA16), and *Coronaviridae* (genus *Alphacoronavirus*: PEDV and genus *Betacoronavirus*: SARS-CoV-2 including Wuhan wild type and variants of concerns including alpha, beta, delta, and omicron). The superantibody inhibited replication of all RNA viruses that were tested in a dose-dependent manner. *In-silico* analysis indicated that the superantibody interacted mainly with the armadillo-like arm repeats at the back of the RdRp thumb domain, which may cause allosterical changes in the spatial conformation of the RdRp, rendering the enzyme inactive, hence virus replication inhibition. Although the molecular mechanisms of the superantibody against the viruses await experimental elucidation, data of this study persuade testing the superantibody further toward clinical application as a pan-direct acting anti-RNA virus agent.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

WC, MC, and KG-A contributed to the conceptualization, funding acquisition, resources, and project administration. MC, NS, and WC contributed to the methodology, data curation, formal analysis, supervision, visualization, and writing and editing the manuscript. KG-A, TT, TS, WT, KM, and MC contributed to the investigation, methodology, and visualization. KK and MC did the computerization. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the Program Management Unit-Brain Power (PMUB), Office of National Higher Education Science Research and Innovation Policy Council (NXPO) (Grant Number B05F640123), and Mahidol University (Grant Number MRC-IM 03/2565).

ACKNOWLEDGMENTS

We thank Takaji Wakita of the Department of Microbiology, Tokyo Metropolitan Institute for Neuroscience, Tokyo, Japan, and Ralf Bartenschlager, Department of Molecular Virology, University of Heidelberg, Germany, for providing the pJFH-1 replicon. Thanks to Yong Poovorawan, Chulalongkorn University, Bangkok, and Passanesh Sukpohpetch, Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, for providing

the DENV serotypes 1-4; Duncan Smith, Institute of Molecular Biosciences, Mahidol University, for providing ZIKV and JEV; Jeeraphong Thanongsaksrikul, Faculty of Allied Health Sciences, Thammasat University, Thailand, for providing the EV71 and RD cells; Thaweesak Songserm, Department of Veterinary Pathology, Faculty of Veterinary Medicine, Kasetsart University, Kampaengsaen Campus, Nakhon Pathom, Thailand, for providing the PEDV; Prasert Auewarakul, Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, for providing the Vero E6 cells and SARS-CoV-2; Ratigorn Guntapong, Department of Medical Sciences, Ministry of Public Health, Thailand, for providing CVA16; and Thawornchai Limjindaporn, Faculty of Medicine Siriraj Hospital, Mahidol University, for providing the Huh7 cells.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.926929/full#supplementary-material

REFERENCES

Abbass, S., Kamal, E., Salama, M., Salman, T., Sabry, A., and Abdel-Razek, W. (2021). Efficacy and safety of sofosbuvir plus daclatasvir or ravidasvir in patients with COVID-19: a randomized controlled trial. *J. Med. Virol.* 93, 6750–6759. doi: 10.1002/jmv.27264

Agostini, M. L., Andres, E. L., Sims, A. C., Graham, R. L., Sheahan, T. P., Lu, X., et al. (2018). Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *mBio* 9, e00221–e00218. doi: 10.1128/mBio.00221-18

Biswal, B. K., Cherney, M. M., Wang, M., Chan, L., Yannopoulos, C. G., and Bilimoria, D. (2005). Crystal structures of the RNA-dependent RNA polymerase genotype 2a of hepatitis C virus reveal two conformations and suggest mechanisms of inhibition by non-nucleoside inhibitors. *J. Biol. Chem.* 280, 18202–18210. doi: 10.1074/jbc.M413410200

Boumaza, A., Gay, L., Mezouar, S., Bestion, E., Diallo, A. B., Michel, M., et al. (2021). Monocytes and macrophages, targets of severe acute respiratory syndrome coronavirus 2: the clue for coronavirus disease 2019 immunoparalysis. *J. Infect. Dis.* 224, 395–406. doi: 10.1093/infdis/jiab044

Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R. L., and Mathieu, M. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13034–13039. doi: 10.1073/pnas.96.23.13034

Brooks, N. A., Esparon, S., Pouniotis, D., and Pietersz, G. A. (2015). Comparative immunogenicity of a cytotoxic T cell epitope delivered by penetratin and TAT cell penetrating peptides. *Molecules* 20, 14033–14050. doi: 10.3390/molecules200814033

Brooks, N. A., Pouniotis, D. S., Tang, C. K., Apostolopoulos, V., and Pietersz, G. A. (2010). Cell-penetrating peptides: application in vaccine delivery. *Biochim. Biophys. Acta* 1805, 25–34. doi: 10.1016/j.bbcan.2009.09.004

Bruenn, J. A. (2003). A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.* 31, 1821–1829. doi: 10.1093/nar/gkg277

Chan-Yeng, M., Xu, R. H., and Shi, Z. (2015). Bat origin of human coronaviruses. *Virol. J.* 12:21.

Choy, K. T., Wong, A. Y., Kaewpreedee, P., Sia, S. F., Chen, D., and Hui, K. P. Y. (2020). Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. *Antiviral Res.* 178:104786. doi: 10.1016/j.antiviral.2020.104786

Coghlan, A. (2022). *Super Antibodies Break the Cell Barrier*. London, UK: NewScientist.

De Clercq, E. (2013). Antivirals: past, present and future. *Biochem. Pharmacol.* 85, 727–744. doi: 10.1016/j.bcp.2012.12.011

Derossi, D., Joliot, A. H., Chassaing, G., and Prochiantz, A. (1994). The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* 269, 10444–10450.

European Centre for Disease Prevention and Control (2020). *Dengue Worldwide Overview*. Solna: European Centre for Disease Prevention and Control.

Frediansyah, A., Nainu, F., Dhama, K., Mudatsir, M., and Harapan, H. (2021). Remdesivir and its antiviral activity against COVID-19: a systematic review. *Clin. Epidemiol. Glob. Health* 9, 123–127. doi: 10.1016/j.cegh.2020.07.011

Glab-ampai, K., Chulanetra, M., Malik, A. A., Juntadech, T., Thanongsaksrikul, J., and Srimanote, P. (2017). Human single chain-transbodies that bound to domain-I of non-structural protein 5A (NS5A) of hepatitis C virus. *Sci. Rep.* 7:15042. doi: 10.1038/s41598-017-14886-9

Gong, P., and Peersen, O. B. (2010). Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* 107, 22505–22510. doi: 10.1073/pnas.1007626107

Gorbalenya, A. E., Pringle, F. M., Zeddam, J. L., Luke, B. T., Cameron, C. E., Kalmakoff, J., et al. (2002). The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J. Mol. Biol.* 324, 47–62. doi: 10.1016/s0022-2836(02)01033-1

Honorato, R. V., Koukos, P. I., Jimenez-Garcia, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., et al. (2021). Structural biology in the clouds: the WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* 8:729513. doi: 10.3389/fmolb.2021.729513

Hu, B., Ge, X., Wang, L. F., and Shi, Z. (2015). Bat origin of human coronaviruses. *Virol. J.* 12:221.

Jia, H., and Gong, P. (2019). A structure-function diversity survey of the RNA-dependent RNA polymerases from the positive-strand RNA viruses. *Front. Microbiol.* 10:1945. doi: 10.3389/fmicb.2019.01945

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Jung, K., Hu, H., and Saif, L. J. (2016). Porcine deltacoronavirus infection: etiology, cell culture for virus isolation and propagation, molecular epidemiology and pathogenesis. *Virus Res.* 226, 50–59. doi: 10.1016/j.virusres.2016.04.009

Khandia, R., Munjal, A., Dhama, K., Karthik, K., Tiwari, R., Malik, Y. S., et al. (2018). Modulation of Dengue/Zika virus pathogenicity by antibody-dependent enhancement and strategies to protect against enhancement in Zika virus infection. *Front. Immunol.* 9:597. doi: 10.3389/fimmu.2018.00597

Kulkarni, R. (2020). "Antibody-dependent enhancement of viral infections," in *Dynamics of Immune Activation in Viral Diseases*, ed. P. V. Bramhachari (Singapore: Springer Singapore), 9–41.

Kulkeaw, K., Sakolvaree, Y., Srimanote, P., Tongtawe, P., Maneewatch, S., and Sookrung, N. (2009). Human monoclonal ScFv neutralize lethal Thai cobra, *Naja kaouthia*, neurotoxin. *J. Proteomics* 72, 270–282. doi: 10.1016/j.jprot.2008.12.007

Kusakawa, T., Shimakami, T., Kaneko, S., Yoshioka, K., and Murakami, S. (2007). Functional interaction of hepatitis C virus NS5B with nucleolin GAR domain. *J. Biochem.* 141, 917–927. doi: 10.1093/jb/mvm102

Kyle, J. L., and Harris, E. (2008). Global spread and persistence of dengue. *Ann. Rev. Microbiol.* 62, 71–92.

Le Pogam, S., Kang, H., Harris, S. F., Leveque, V., Giannetti, A. M., Ali, S., et al. (2006). Selection and characterization of replicon variants dually resistant to thumb- and palm-binding nonnucleoside polymerase inhibitors of the hepatitis C virus. *J. Virol.* 80, 6146–6154. doi: 10.1128/JVI.02628-05

Li, J., and Johnson, K. A. (2016). Thumb site 2 inhibitors of hepatitis C viral RNA-dependent RNA polymerase allosterically block the transition from initiation to elongation. *J. Biol. Chem.* 291, 10067–10077. doi: 10.1074/jbc.M115.708354

Lim, S. P., Noble, C. G., She, C. C., Soh, T. S., El Sahili, A., Chan, G. K., et al. (2016). Potent allosteric dengue virus NS5 polymerase inhibitors: mechanism of action and resistance profiling. *PLoS Pathog.* 12:e1005737. doi: 10.1371/journal.ppat.1005737

Liu, Y., Soh, W. T., Kishikawa, J. I., Hirose, M., Nakayama, E. E., Li, S., et al. (2021). An infectivity-enhancing site on the SARS-CoV-2 spike protein targeted by antibodies. *Cell* 184, 3452–3466.e18. doi: 10.1016/j.cell.2021.05.032

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi: 10.1038/s41592-022-01488-1

Narayan, R., and Tripathi, S. (2020). Intrinsic ADE: the dark side of antibody dependent enhancement during dengue infection. *Front. Cell. Infect. Microbiol.* 10:580096. doi: 10.3389/fcimb.2020.580096

Noobrakhsh, F., Abdolmohammadi, K., Fatahi, Y., Dalili, H., Rasoolinejad, M., Rezaei, F., et al. (2019). Zika virus infection, basic and clinical aspects: a review article. *Iran. J. Public Health* 48, 20–31.

Novel Influenza A/H1N1 Investigation Team (2009). Description of the early stage of pandemic (H1N1) 2009 in Germany, 27 April-16 June 2009. *Euro. Surveill.* 14:19295. doi: 10.2807/ese.14.31.19295-en

Payne, S. (2017). "Chapter 10—Introduction to RNA viruses," in *Viruses*, ed. S. Payne (Cambridge, MA: Academic Press), 97–105.

Pensaert, M. B., and de Bouck, P. (1978). A new coronavirus-like particle associated with diarrhea in swine. *Arch. Virol.* 58, 243–247.

Phanthong, S., Densumite, J., Seesuay, W., Thanongsaksrikul, J., Teimoori, S., Sookrung, N., et al. (2020). Human antibodies to VP4 inhibit replication of enteroviruses across subgenotypes and serotypes and enhance host innate immunity. *Front. Microbiol.* 11:562768. doi: 10.3389/fmicb.2020.562768

Picarazzi, F., Vicenti, I., Saladini, F., Zazzi, M., and Mori, M. (2020). Targeting the RdRp of emerging RNA viruses: the structure-based drug design challenge. *Molecules* 25:5695. doi: 10.3390/molecules25235695

Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 8, 3867–74. doi: 10.1002/j.1460-2075.1989.tb08565.x

Poungpair, O., Pootong, A., Maneewatch, S., Srimanote, P., Tongtawe, P., Songserm, T., et al. (2010). A human single chain transbody specific to matrix protein (M1) interferes with the replication of influenza A virus. *Bioconjug. Chem.* 21, 1134–1141. doi: 10.1021/bc900251u

Sánchez-Zuno, G. A., Matuz-Flores, M. G., González-Estevez, G., Nicoletti, F., Turrubiates-Hernández, F. J., Mangano, K., et al. (2021). A review: antibody-dependent enhancement in COVID-19: the not so friendly side of antibodies. *Int. J. Immunopathol. Pharmacol.* 35:20587384211050199. doi: 10.1177/20587384211050199

Schmidt, N. J., Lennette, E. H., and Ho, H. H. (1974). An apparently new enterovirus isolated from patients with disease of the central nervous system. *J. Infect. Dis.* 129, 304–309.

Sheahan, T. P., Sims, A. C., Graham, R. L., Menachery, V. D., Gralinski, L. E., Case, J. B., et al. (2017). Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci. Transl. Med.* 9:eaal3653. doi: 10.1126/scitranslmed.aal3653

Sheahan, T. P., Sims, A. C., Zhou, S., Graham, R. L., Pruijssers, A. J., Agostini, M. L., et al. (2020). An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci. Transl. Med.* 2:eabb5883. doi: 10.1126/scitranslmed.abb5883

Shimakami, T., Honda, M., Kusakawa, T., Murata, T., Shimotohno, K., Kaneko, S., et al. (2006). Effect of hepatitis C virus (HCV) NS5B-nucleolin interaction on HCV replication with HCV subgenomic replicon. *J. Virol.* 80, 3332–3340. doi: 10.1128/JVI.80.7.3332-3340.2006

Tang, X., Tian, G., Zhao, J., and Zhou, K. Y. (1998). Isolation and characterization of prevalent strains of avian influenza viruses in China. *Chin. J. Anim. Poult. Infect. Dis.* 20, 1–5.

te Velthuis, A. J. (2014). Common and unique features of viral RNA-dependent polymerases. *Cell. Mol. Life. Sci.* 71, 4403–4420. doi: 10.1007/s00018-014-1695-z

Thavorasak, T., Chulanetra, M., Glab-ampai, K., Teeranitayatarn, K., Songserm, T., and Yodsheewan, R. (2022). Novel neutralizing epitope of PEDV S1 protein identified by IgM monoclonal antibody. *Viruses* 14:125. doi: 10.3390/v14010125

Thueng-In, K., Thanongsaksrikul, J., Jittavisutthikul, S., Seesuay, W., Chulanetra, M., Sakolvaree, Y., et al. (2014). Interference of HCV replication by cell penetrable human monoclonal scFv specific to NS5B polymerase. *mAbs* 6, 1327–1339. doi: 10.4161/mabs.29978

Tian, L., Qiang, T., Liang, C., Ren, X., Jia, M., Zhang, J., et al. (2021). RNA-dependent RNA polymerase (RdRp) inhibitors: the current landscape and repurposing for the COVID-19 pandemic. *Eur. J. Med. Chem.* 213:113201. doi: 10.1016/j.ejmech.2021.113201

van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 Web Server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* 428, 720–725. doi: 10.1016/j.jmb.2015.09.014

Venkataraman, S., Prasad, B. V. L. S., and Selvarajan, R. (2018). RNA dependent RNA polymerases: insights from structure, function and evolution. *Viruses* 10:76. doi: 10.3390/v10020076

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., and Zhao, Z. (2005). Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat. Med.* 11, 791–796. doi: 10.1038/nm1268

Wallin, R. F. (1998). *A Practical Guide to ISO 109903-5: Cytotoxicity. Medical Devices and Diagnostic Industry*. Available online at: https://www.mddionline.com/testing/practical-guide-iso-10993-5-cytotoxicity (accessed July 9, 2021).

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., and Xu, M. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* 30, 269–271. doi: 10.1038/s41422-020-0282-0

Winarski, K. L., Tang, J., Klenow, L., Lee, J., Coyle, E. M., and Manischewitz, J. (2019). Antibody-dependent enhancement of influenza disease promoted by increase in hemagglutinin stem flexibility and virus fusion kinetics. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15194–15199. doi: 10.1073/pnas.1821317116

World Health Organization [WHO] (2004). *The 4th Edition of the Manual for the Virological Investigation of Polio*. Geneva: WHO.

World Health Organization [WHO] (2019). *Middle East Respiratory Syndrome Coronavirus (MERS-CoV)*. Geneva: WHO.

World Health Organization Ebola Response Team (2014). Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* 371, 1481–1495. doi: 10.1056/NEJMoa1411100

Wu, J., Liu, W., and Gong, P. (2015). A structural overview of RNA-dependent RNA polymerases from the Flaviviridae family. *Int. J. Mol. Sci.* 16, 12943–12957. doi: 10.3390/ijms160612943

Yang, J., Luo, Y., Shibu, M. A., Toth, I., and Skwarczynskia, M. (2019). Cell-penetrating peptides: efficient vectors for vaccine delivery. *Curr. Drug. Deliv.* 16, 430–443. doi: 10.2174/1567201816666190123120915

Yap, T. L., Xu, T., Chen, Y. L., Malet, H., Egloff, M. P., Canard, B., et al. (2007). Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J. Virol.* 81, 4753–4765. doi: 10.1128/JVI.02283-06

Zein, A. F. M. Z., Sulistiyana, C. S., Raffaello, W. M., Wibowo, A., and Pranata, R. (2021). Sofosbuvir with daclatasvir and the outcomes of patients with COVID-19: a systematic review and meta-analysis with GRADE assessment. *Postgrad. Med. J.* 21, 1–6. doi: 10.1136/postgradmedj-2021-140287

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 03 August 2022
DOI 10.3389/fmicb.2022.931307

OPEN ACCESS

EDITED BY
Beibei Ru,
National Institutes of Health (NIH),
United States

REVIEWED BY
Sabine Flicker,
Medical University of Vienna, Austria
Arjun Mishra,
University of Maryland,
College Park,
United States

*CORRESPONDENCE
Clarence Ronald Geyer
ron.geyer@usask.ca

†PRESENT ADDRESS
Bharathikumar Vellalore Maruthachalam,
Janssen Pharmaceuticals,
Spring House, PA, United States

SPECIALTY SECTION
This article was submitted to
Phage Biology,
a section of the journal
Frontiers in Microbiology

RECEIVED 28 April 2022
ACCEPTED 12 July 2022
PUBLISHED 03 August 2022

CITATION
Maruthachalam BV, Barreto K, Hogan D, Kusalik A and Geyer CR (2022) Generation of synthetic antibody fragments with optimal complementarity determining region lengths for Notch-1 recognition. *Front. Microbiol.* 13:931307. doi: 10.3389/fmicb.2022.931307

Generation of synthetic antibody fragments with optimal complementarity determining region lengths for Notch-1 recognition

Bharathikumar Vellalore Maruthachalam[1†], Kris Barreto[1], Daniel Hogan[2], Anthony Kusalik[2] and Clarence Ronald Geyer[3*]

[1]Department of Biochemistry, University of Saskatchewan, Saskatoon, SK, Canada, [2]Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada, [3]Department of Pathology, University of Saskatchewan, Saskatoon, SK, Canada

Synthetic antibodies have been engineered against a wide variety of antigens with desirable biophysical, biochemical, and pharmacological properties. Here, we describe the generation and characterization of synthetic antigen-binding fragments (Fabs) against Notch-1. Three single-framework synthetic Fab libraries, named S, F, and modified-F, were screened against the recombinant human Notch-1 extracellular domain using phage display. These libraries were built on a modified trastuzumab framework, containing two or four diversified complementarity-determining regions (CDRs) and different CDR diversity designs. In total, 12 Notch-1 Fabs were generated with 10 different CDRH3 lengths. These Fabs possessed a high affinity for Notch-1 (sub-nM to mid-nM K_{Dapp} values) and exhibited different binding profiles (mono-, bi-or tri-specific) toward Notch/Jagged receptors. Importantly, we showed that screening focused diversity libraries, implementing next-generation sequencing approaches, and fine-tuning the CDR length diversity provided improved binding solutions for Notch-1 recognition. These findings have implications for antibody library design and antibody phage display.

KEYWORDS

phage display, synthetic antibody, focused library, long CDR library, next-generation sequencing, Notch signaling, Notch1

Introduction

Notch signaling is a highly conserved pathway that influences multiple cell fate decisions, including proliferation, apoptosis, differentiation, migration, and angiogenesis in developing and adult metazoan organisms. In mammals, Notch signaling is initiated by four Notch receptors (Notch1-4) and five Delta/Serrate/LAG-2 (DSL) ligands (Jagged-1/2 and DLL-1/3/4), all of which are modular, type-I, single-pass, transmembrane proteins. The extracellular region of Notch contains a series of epidermal growth factor (EGF)-like

repeats that are required for ligand binding, followed by a negative regulatory region, and a hetero-dimerization domain. The extracellular region of DSL ligands contains an MNNL domain (module at the N-terminus of Notch ligands), followed by a DSL domain, and a series of EGF-like repeats similar to Notch receptors. The intracellular region of Notch contains numerous well-defined domains, whereas the intracellular region is not conserved in DSL ligands (reviewed by Katoh and Katoh, 2020).

Canonical, short-range Notch signaling happens between two adjacent cells, the signal-sending cell, which expresses the DSL ligand, and the signal-receiving cell, which expresses the Notch receptor. A bi-molecular interaction between the Notch receptor and the DSL ligand at the cell surface initiates a process called regulated intramembrane proteolysis where the Notch receptor is cleaved within the extracellular hetero-dimerization domain by a metalloprotease of the ADAM family. This DSL-dependent metalloprotease processing renders Notch sensitive to the γ-secretase protein complex. γ-Secretase cleaves the Notch receptor within the transmembrane region, releasing the Notch intracellular domain. The intracellular domain translocates into the nucleus where it forms a transcriptional activation complex and regulates Notch-responsive genes (reviewed by Katoh and Katoh, 2020).

In addition to its physiological roles, deregulation of Notch signaling is associated with developmental disorders, neurological diseases, solid cancers, and hematologic malignancies. These pathological conditions result from overexpression or gain/loss-of-function mutations within Notch and Jagged receptors (Ranganathan et al., 2011). Small molecules have been developed that block Notch signaling by inhibiting the γ-secretase complex and monoclonal antibodies have been developed that that target Notch or Jagged ectodomains (Zhou et al., 2022). Since γ-secretase inhibitors block the proteolysis of multiple transmembrane proteins, including all four Notch receptors, they cause side effects in clinical trials. Similar side effects are observed with pan-Notch-specific monoclonal antibodies (Takebe et al., 2014). Synthetic antibodies have been constructed that target the negative-regulatory region of Notch-1 and the receptor-binding region of Jagged-1 and bind specifically to Notch and Jagged receptors, respectively. They exhibit potent and selective inhibition of Notch-1 or Jagged-1 signaling in tumor models and show promising results in pre-clinical studies (Wu et al., 2010; Lafkas et al., 2015). Thus, there is interest in developing paralogue-specific, synthetic antibodies against Notch receptors and DSL ligands.

Previously, we designed, constructed, and validated a phage-displayed, single-framework, synthetic antigen-binding fragment (Fab) library named library-S (Maruthachalam et al., 2017). We used Library-S and a similar synthetic Fab library, named Library-F (Persson et al., 2013) for generating high-affinity Fabs against Notch-2/3 and Jagged-2 (Barreto et al., 2019). In this work, we describe the generation and characterization of Notch-1 Fabs from three synthetic Fab libraries S, F, and modified-F. These Fab libraries were built on a modified trastuzumab framework, containing two or four diversified complementarity determining regions (CDRs) and different CDR diversity designs. During the course of the Fab generation, we showed that screening focused diversity libraries, implementing next-generation sequencing (NGS) approaches, and fine-tuning the library diversity, can improve binding solutions for Notch-1 recognition. We generated 12 Notch-1 Fabs with high affinity (sub-nM to mid-nM K_{Dapp} values) and different binding specificities toward Notch/Jagged receptors. The highest-affinity Fab bound specifically to Notch-1 with a K_{Dapp} value of 0.16 ± 0.1 nM. Notch-1 Fabs described in this work should permit a more precise modulation of the Notch signaling pathway.

Results

Notch-1 Fabs from Library-S

Library-S contains four-fixed CDRs (L1, L2, H1, and H2) and two-diversified CDRs (L3 and H3) on a modified Hu4D5 trastuzumab framework (Maruthachalam et al., 2017). Length diversity is included within CDRH3 and amino acid diversity is included within CDRH3 and CDRL3 (Figures 1A,B). To generate Notch-1 Fabs, we conducted solid-phase, phage display panning of Library-S against the human Notch-1 extracellular domain. Random clone picking and Sanger sequencing of 20 phagemids from the round 4 phage pool gave rise to two Fab clones (Figure 1C) N1/S/1 and N1/S/2 with CDRH3 lengths of 13 and 16 residues, respectively. Phage-ELISA indicated that both Fabs bound to Notch-1 but not to BSA or the Fc protein (Supplementary Table 2).

Since length diversity was restricted to CDRH3, we hypothesized that CDRH3 length would play a central role in Library-S binding. To develop a selection strategy for obtaining Notch-1 Fabs with different CDRH3 lengths, we split Library-S into four sub-libraries each containing a set of CDRH3 lengths, and panned them against Notch-1. After rounds 4 and 5, we calculated the number of phages eluted from target-coated wells relative to BSA-coated wells. A positive enrichment in target-specific phage number was only observed with two sub-libraries SL2 (CDRH3 range of 11–13 aa) and SL3 (CDRH3 range of 14–16 aa; Figure 1D), which contained CDRH3 lengths of Fabs (N1/S/1 and N1/S/2) isolated in the Library-S selection. Through random clone picking and Sanger sequencing we isolated Fab clones N1/SL2/1 and N1/SL3/1 from SL2 and SL3 round 4 phage selection pools, respectively (Figure 1D). N1/SL2/1 had a CDRH3 length of 13 amino acids, which was the same length as N1/S/1, and N1/SL3/1 and N1/S/2 had the same CDRH3 length of 16 residues (Figure 1E). Phage-ELISAs indicated that N1/SL2/1 and N1/SL3/1 bound to Notch-1 but not to BSA or the Fc protein (Supplementary Table 2). Random clone picking and Sanger sequencing of 20 phagemids from SL1 and SL4 round 4 selection pools gave rise to six Fab clones; however, they bound to all test and control antigens in phage-ELISA.

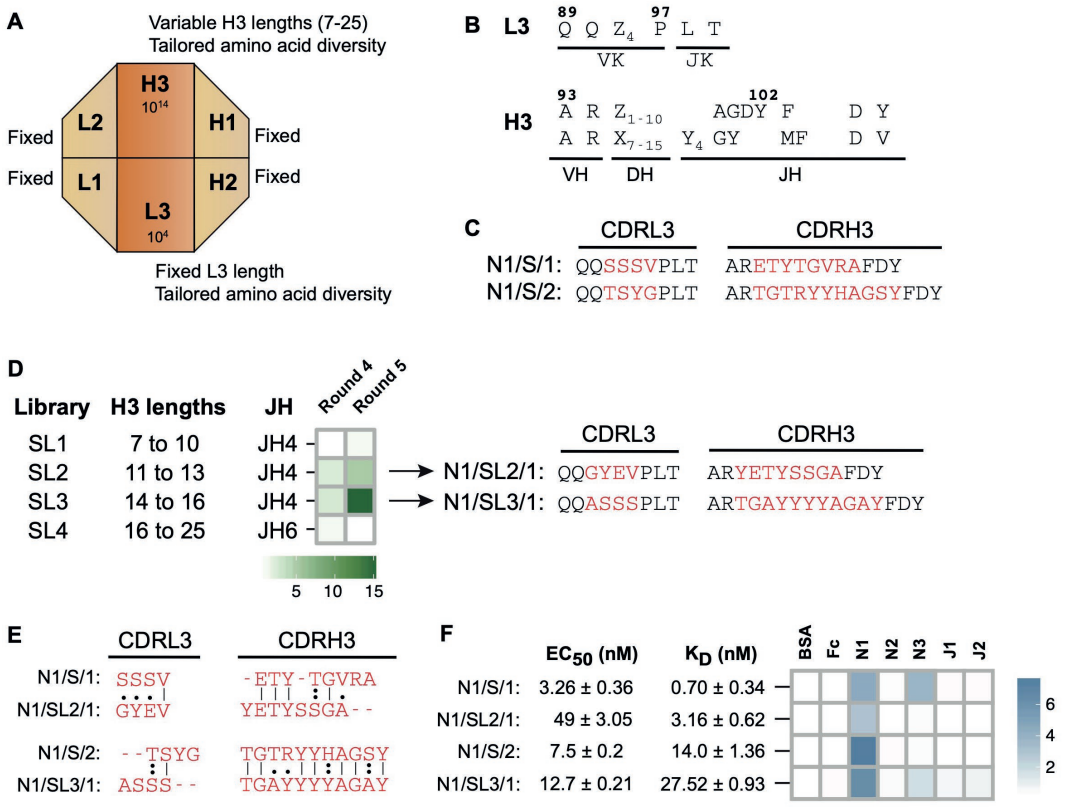


FIGURE 1
Notch-1 Fabs from Library-S. **(A)** Schematic representation of Library-S CDR diversity. **(B)** CDR diversity designs for L3 and H3. Z denotes any of the following 13 amino acids introduced at different proportions: Y (20%), S (20%), G (20%), T (6.5%), A (6.5%), P (6.5%), H (3.5%), R (3.5%), E (3.5%), F (2.5%), W (2.5%), V (2.5%), or L (2.5%). X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%), or V (5%). CDRH3 length is varied by altering the number of Z and X. **(C)** CDRL3 and CDRH3 sequences of two Notch-1 Fabs isolated from Library-S. Fixed anchor residues are in black and diversified CDR positions are in red. **(D)** Left panel: panning of Library-S sub-libraries (SL1–SL4) against Notch-1. Heatmap showing the enrichment of Notch-1 binding phages after phage display selection rounds 4 and 5. Fold enrichment is the ratio of number of phages eluted from target-coated wells to number of phages eluted from BSA-coated wells. Right panel: CDRL3 and CDRH3 sequences of Fab N1/SL2/1 (isolated from SL2) and Fab N1/SL3/1 (isolated from SL3). **(E)** Pairwise sequence alignments between master-and sub-library Fab CDR sequences. Only diversified CDR positions within L3 and H3 were included for the alignment. Sequences were aligned using the Needleman–Wunsch algorithm in the EMBOSS Needle server. **(F)** Affinity and specificity of four Notch-1 Fabs isolated from Library-S master-and sub-libraries. EC_{50} and K_D values were determined using multi-point Fab-ELISA and bio-layer interferometry, respectively. Fab specificity was determined using single-point Fab-ELISA at 1μM Fab concentration. Abs_{450} values are shown as a heatmap. Notch-1, Notch-2, Notch-3, Jagged-1, and Jagged-2 are indicated as N1, N2, N3, J1, and J2, respectively.

Next, we purified four Notch-1 Fabs and measured their affinity and specificity for Notch-1. First, we determined EC_{50} values for Fabs binding to Notch-1 using Fab-ELISA (Figure 1F; Supplementary Figure 1). The EC_{50} values for the 13 aa CDRH3 Fabs N1/S/1 and N1/SL2/1 were 3.3 nM and 49 nM, respectively. The EC_{50} of the 16 aa CDRH3 Fabs N1/S/2 and N1/SL3/1 were 7.5 nM and 12.7 nM, respectively. Second, we measured kinetics of Fab binding to Notch-1 using bio-layer interferometry. Consistent with Fab-ELISA, N1/S/1 was the highest-affinity binder ($K_D = 0.7$ nM). N1/SL2/1 bound to Notch-1 with a K_D value of 3.2 nM. N1/S/2 and N1/SL3/1 possessed mid-nM K_D values, 14 nM and 27.5 nM, respectively (Figure 1F; Supplementary Table 4). These values represent apparent K_D values (K_{Dapp}), as the Fc-Notch proteins may dimerize through the Fc domain resulting in avidity, which may explain the lower K_D values observed for N1/S/1 and N1/SL2/1 compared to the EC_{50} values. In contrast, 16 aa CDRH3 Fabs had slightly higher K_{Dapp} than EC_{50} values. Third, to assess Fab specificity, we tested the binding of Fabs to Notch and Jagged receptor ectodomains using a Fab-ELISA. At 1 μM Fab concentration, N1/S/1 and N1/SL3/1 cross-reacted with Notch-3, whereas N1/SL2/1 and N1/S/2 exhibited Notch-1 specific binding (Figure 1F; Supplementary Table 3). Kinetic analysis indicated that the highest-affinity Fab N1/S/1 also bound to Notch-3 with a K_{Dapp} of 14.1 ± 0.53 nM.

Notch-1 Fabs from Library-F

In an attempt to obtain Notch-1 Fabs with higher affinity and specificity, we used Library-F, which contains four-diversified CDRs and two-fixed CDRs on a modified trastuzumab framework

(Persson et al., 2013). Length diversity is included within CDRs L3 and H3 and amino acid diversity is included within CDRs L3 and H3 and solvent-accessible positions of CDRs H1 and H2 (Figures 2A,B). We hypothesized that the additional diversity in Library-F (length diversity in CDRL3 and amino acid diversity in CDRs H1/H2) would generate higher affinity and more selective Fabs against Notch-1.

Solid-phase panning of Library-F was conducted against the Notch-1 extracellular domain. To identify Fabs with different CDRH3 lengths, we sequenced the CDRH3 region of the Fab-phage selection outputs using Ion Torrent sequencing (Rothberg et al., 2011). We monitored changes in CDRH3 length distribution over five rounds of selection (Figure 2C). We observed a prominent enrichment in three short CDRH3 lengths (8, 9, and 10 residues). Four other CDRH3 lengths (14, 15, 17, and 19 residues) were also enriched albeit at lower frequencies, ranging between 2% and 7%. We ranked round 3 CDRH3 sequences based on their relative frequencies. Nineteen sequences were present above 0.1%, and all of them had the above-mentioned CDRH3 lengths. While 15 out of 19 sequences had short CDRH3 lengths (Figure 2D), each of the other four CDRH3 lengths had one sequence above 0.1% (Figure 2E). The longest CDRH3 was 19 residues and was the anti-maltose binding protein (MBP) CDRH3 sequence used in template phagemid for library F construction (Persson et al., 2013). We analyzed the frequency of the most abundant Fabs from six selected lengths over five rounds of selection (Figure 2F). The three short CDRH3 sequences showed a significant enrichment throughout the selection process. The three larger CDRH3 sequences showed a lower enrichment in round 3, and their frequency decreased in later rounds.

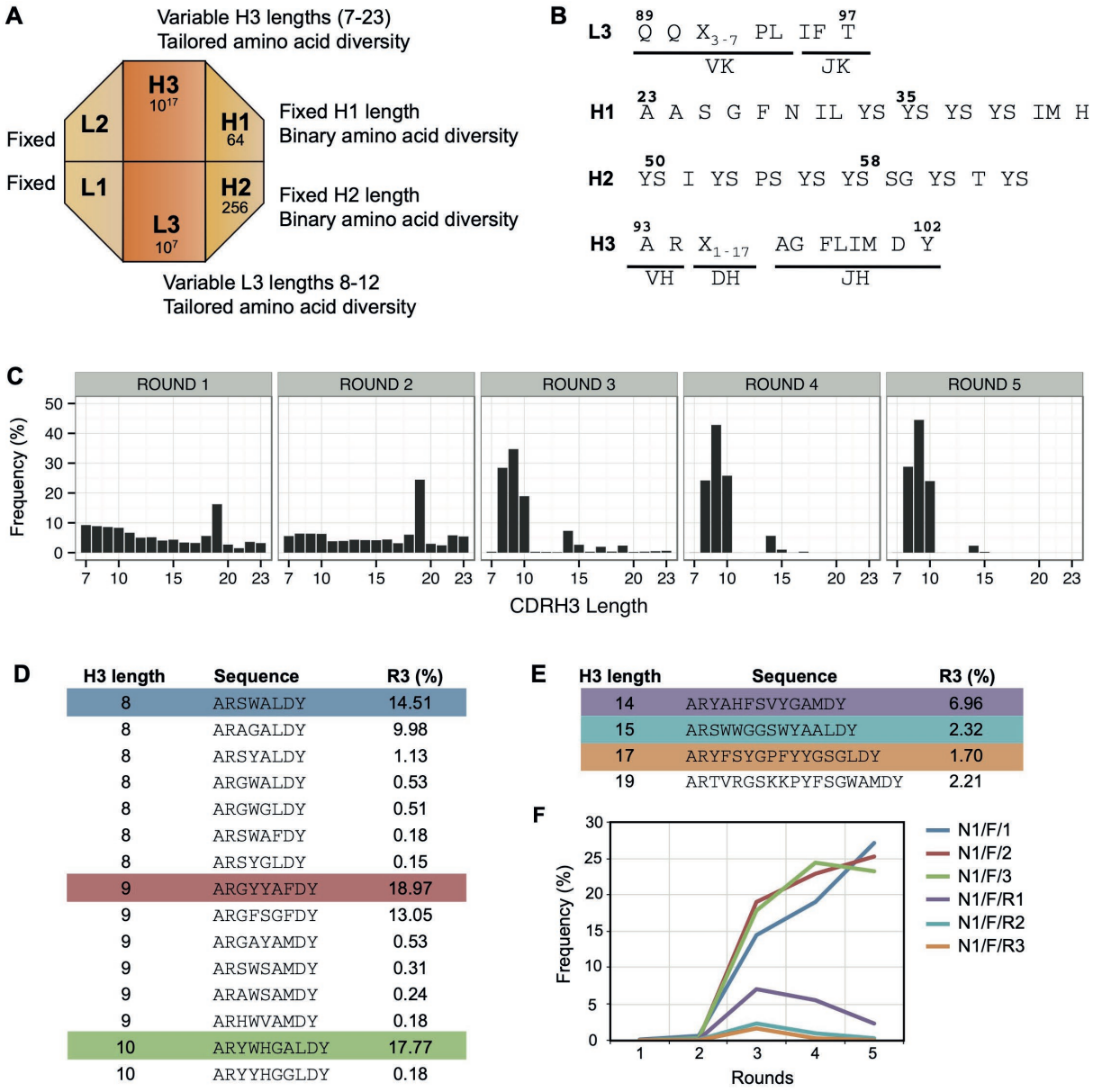
Random clone picking and Sanger sequencing of 20 clones from rounds 3 and 4 phage pools only recovered the three most-frequent clones with short CDRH3 lengths (referred as high-frequency clones). Therefore, we used our NGS-assisted Fab reconstruction method to recover the three less-frequent clones with longer CDRH3 lengths (referred as low-frequency clones). To identify the diversified CDR sequences (L3, H1, and H2) that paired with the CDRH3 sequences of interest, we removed the intervening sequence between the CDRs using our CDR strip methodology (Barreto et al., 2019). This generated a PCR amplicon of ~200 bp that contained CDRs L3-H1-H2-H3 closer in linear sequence space. We used this strategy to sequence a CDR strip from round 3 on a 400-bp chip. As the maximum read length offered by Ion Torrent was 400 bases, the framework deletion step was necessary to reduce the amplicon length from 1,100 bases to 200 bases. Once CDR-combinations were identified, we reconstructed desired Fab clones by cloning CDR-encoding oligonucleotides into the MBP template phagemid (Persson et al., 2013) by Kunkel mutagenesis (Kunkel et al., 1987). Diversified CDR sequences of six Fab clones are shown in Figure 3A.

In phage-ELISA, all six Fabs tested positive for binding to Notch-1 (Figure 3A; Supplementary Table 2). Compared to high-frequency Fabs, ELISA signals were ~4-fold weaker for N1/F/R2 and ~ 10-fold weaker for N1/F/R3. Next, we converted six phage-Fab clones into soluble Fabs and measured their EC_{50} values against Notch-1 using Fab-ELISA (Figure 3B; Supplementary Figure 1). Except for N1/F/R3, the other five Fabs had EC_{50} values of ~4 nM. N1/F/R3 bound ~10-fold weaker than the other Fabs (EC_{50} = 37 nM). Next, we analyzed Fab-binding kinetics using bio-layer interferometry and Fab specificity using Fab-ELISA (Figure 3B; Supplementary Tables 3–4). Even though the two highest frequency Fabs, N1/F/1 and N1/F/2, exhibited sub-nM K_{Dapp} values for Notch-1, they also possessed mid-nM K_{Dapp} values for Notch-2 and low-nM values for Notch-3. N1/F/1 bound to Notch-1, Notch-2, and Notch-3 with K_{Dapp} values of 0.45 ± 0.05 nM, 48.2 ± 1.42 nM and 5.7 ± 0.23 nM, respectively. N1/F/1 affinity was 107-fold lower for Notch-2 and 13-fold lower for Notch-3. N1/F/2 bound to Notch-1, Notch-2, and Notch-3 with K_{Dapp} values of 0.45 ± 0.05 nM, 31 ± 1 nM and 4.1 ± 0.18 nM, respectively. N1/F/2 affinity was 69-fold lower for Notch-2 and 9-fold lower for Notch-3. The third highest frequency Fab N1/F/3 had a K_{Dapp} value of 1.88 nM for Notch-1 and cross-reacted with Notch-3 and Jagged-2 in Fab-ELISA. The three low-frequency Fabs with medium CDRH3 lengths possessed very similar binding profiles. They bound to Notch-1 with low-nM K_D values and showed weak to moderate binding with Jagged-2 in Fab-ELISA. N1/F/R1 and N1/F/R2 bound to Notch-1 with K_{Dapp} values of 1.98 nM and 1.38 nM, respectively. In agreement with phage-ELISA and Fab-ELISA, N1/F/R3 was the lowest-affinity Notch-1 binder from Library-F (K_{Dapp} = 6.6 nM). Kinetic analysis confirmed dose-dependent binding of N1/F/R2 with Jagged-2 with a K_{Dapp} value of 70.4 ± 2.3 nM.

Most notably, low-frequency Fabs recovered using the NGS approach had low-nM affinities for Notch-1 and were more specific than high-frequency Fabs isolated by Sanger sequencing. In particular, N1/F/R2 was less frequent in the round-3 phage pool (~2%), did not show enrichment in round-4 and round-5, and had weaker phage-ELISA output; however the K_{Dapp} value for purified Fab N1/F/R2 was very similar to N1/F/3 and N1/F/R1 that were more frequent in phage pools.

Notch-1 Fabs from the Modified-F Library

In addition to facilitating low-frequency Fab reconstruction, sequencing CDR-combinations in a selection output allows the comparison of CDRs that are paired with each other. To check the influence of CDRL3 length diversity on CDRH3 length solutions for Notch-1, we monitored the pairing between CDRL3 lengths and CDRH3 lengths in the round 3 phage selection output. We observed a preferential pairing of certain CDRL3 and CDRH3 lengths (Figure 4A). For example, CDRH3 length of eight amino acids is most-frequently paired with CDRL3 lengths of 8 and 9 amino acids. Also, the longest CDRL3 length preferred longer CDRH3 length solutions for Notch-1 recognition. In addition to slower amplification rates in *E. coli*, the poor enrichment of long CDR clones could result from the under representation of long



H3 length	Sequence	R3 (%)
8	ARSWALDY	14.51
8	ARAGALDY	9.98
8	ARSYALDY	1.13
8	ARGWALDY	0.53
8	ARGWGLDY	0.51
8	ARSWAFDY	0.18
8	ARSYGLDY	0.15
9	ARGYYAFDY	18.97
9	ARGFSGFDY	13.05
9	ARGAYAMDY	0.53
9	ARSWSAMDY	0.31
9	ARAWSAMDY	0.24
9	ARHWVAMDY	0.18
10	ARYWHGALDY	17.77
10	ARYYHGGLDY	0.18

H3 length	Sequence	R3 (%)
14	ARYAHFSVYGAMDY	6.96
15	ARSWWGGSWYAALDY	2.32
17	ARYFSYGPFYYGSGLDY	1.70
19	ARTVRGSKKPYFSGWAMDY	2.21

FIGURE 2
Panning Library-F against Notch-1. **(A)** Schematic representation of Library-F CDR diversity. **(B)** CDR diversity designs for L3, H1, H2, and H3. X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%), or V (5%). CDRL3 and CDRH3 lengths are varied by altering the number of X. **(C)** Changes in CDRH3 length distribution over five subsequent phage display rounds for the Notch-1 selection. The peak in rounds 1 and 2 (CDRH3 length of 19 residues) is due to the presence of anti-MBP Fab CDRH3 sequence from the template phagemid. **(D)** CDRH3 sequences >0.1% in round 3 with short loop lengths (8, 9, and 10 amino acids). **(E)** CDRH3 sequences >0.1% in round 3 with loop lengths of 14, 15, 17, and 19 amino acids. In panels **D,E**, the most abundant sequence from each CDRH3 length is highlighted. **(F)** Propagation behavior of highlighted CDRH3 sequences over five rounds of selection.

CDR sequences in the library. To confirm this, we amplified the CDRL3 and CDRH3 regions from the naïve Library-F phage pool and sequenced them on a 100-bp Ion chip. Length analysis confirmed the presence of a bias in the library toward short sequences in both CDRs (Figure 4B). Also, we noticed high retention of template CDRL3 (~24%) and CDRH3 (~13%) sequences in the library, which has a CDRL3 length of 9 residues and CDRH3 length of 19 residues.

To isolate Notch-1 Fabs with long CDRs, we decided to modify the length bias in Library-F toward long CDRs and use the modified library for selections. The Modified-F library had the following design features: (1) CDRL3, CDRH1, and CDRH2 diversity designs were unaltered from Library-F; (2) in addition to 17 JH4-CDRH3 lengths (7–23 residues) used in Library-F, 10 JH6-CDRH3 lengths (16–25 residues) used in Library-S were included; (3) to bias the length diversity toward long CDR lengths,

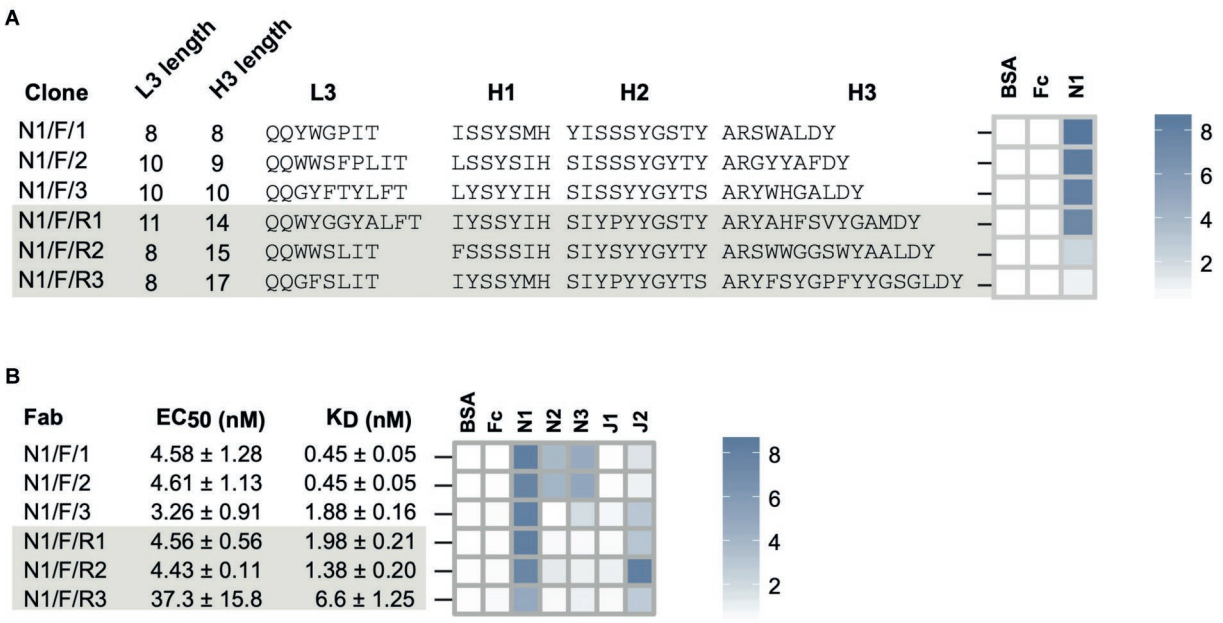


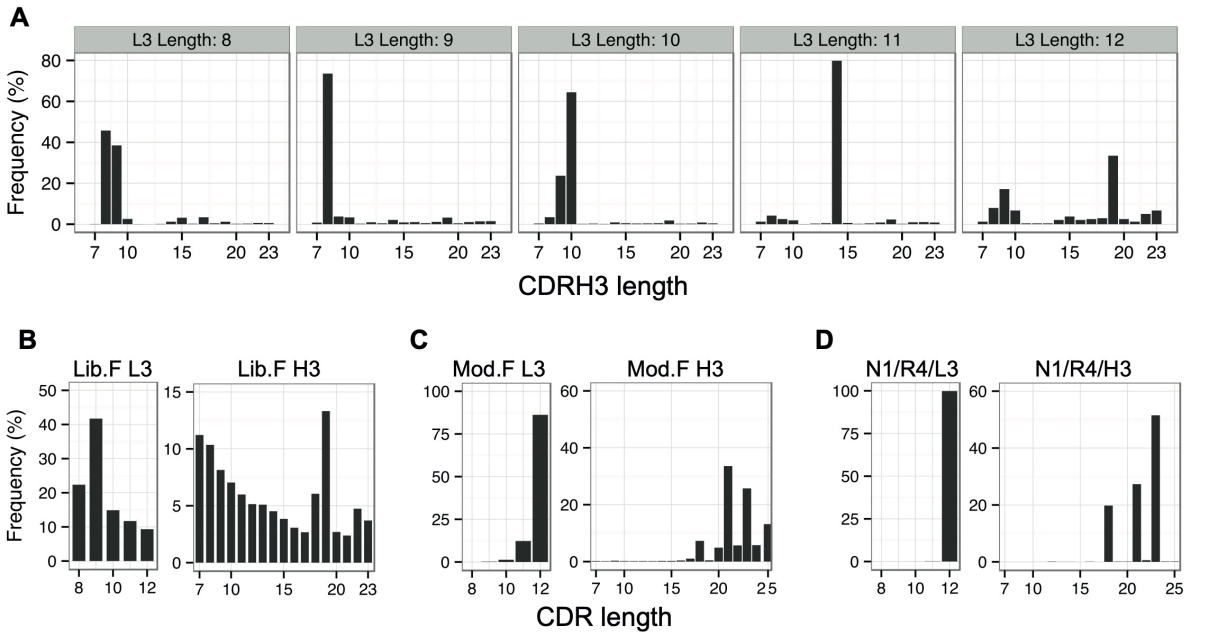
FIGURE 3
Notch-1 Fabs from Library-F. **(A)** Sequence characteristics and phage-displayed Fab binding characteristics of 3 top Fabs (N1/F/1, N1/F/2, and N1/F/3) isolated by Sanger sequencing and 3 rare Fabs (N1/F/R1, N1/F/R2, and N1/F/R3) reconstructed from L3-H1-H2-H3 NGS information. **(B)** Affinity and specificity of purified Notch-1 Fabs. Fab specificity was determined at 1μM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps. Fab nomenclature is listed as Target/Library/Clone number. Notch-1, Notch-2, Notch-3, Jagged-1, and Jagged-2 are indicated as N1, N2, N3, J1, and J2, respectively. S and F indicate library S, Library F, respectively. The clone number is listed in the third position and if they are a "rare" clone they have an "R" in front of the number.

the amount of mutagenic oligonucleotides used in mutagenesis reactions was normalized according to their theoretical diversity; and (4) to prevent the expression of Fabs with unmutated template CDRs, the anti-MBP diversified CDR sequences (Persson et al., 2013) were replaced with TAA stop codons in the template phagemid. The Modified-F library was constructed using an M13 bacteriophage system that allows bivalent Fab display (Lee et al., 2004). Kunkel mutagenesis was used to repair stop codons in the MBP-Fab-encoding template phagemid and replace CDR positions with fixed or degenerate codons encoding the designed amino acid composition. Following mutagenesis, the library DNA was electroporated into an *E. coli* strain suitable for high-efficiency transformation and phage production. NGS analysis of the CDRL3 and CDRH3 regions of the naïve Modified-F library confirmed proper CDR diversification and the length bias toward long CDRs (Figure 4C).

We panned the Modified-F library against Notch-1 and sequenced CDRL3 and CDRH3 regions from the round 4 phage selection output on a 100-bp chip. Length analysis indicated a significant enrichment in CDRL3 sequences of 12 residues and CDRH3 sequences of 18, 21, and 23 residues (Figure 4D). Random clone picking and Sanger sequencing of 20 phagemids recovered seven unique clones from the round 4 phage pool. Three clones tested positive for binding to Notch-1 in phage-ELISA (Figure 4E and Supplementary Table 2). Each ELISA-positive clone had a different CDRH3 length and showed a high CDRH3 sequence enrichment in the round 4 phage pool (19.15% for N1/ModF/1, 25.96% for N1/ModF/2, and 43.5% for N1/ModF/5). We chose to pursue two clones that had longer CDRH3 lengths (N1/ModF/2 and N1/ModF/5). We converted phage-Fab clones into soluble Fab proteins and assayed them using Fab-ELISA (Figure 4F; Supplementary Figure 1; Supplementary Table 3). In Fab ELISA, both Fabs bound stronger to Notch-1 than all other Library-S or Library-F Fabs. The EC_{50} values for N1/ModF/2 and N1/ModF/5 were 1.7 nM and 2.4 nM, respectively. Further, N1/ModF/2 showed strict specificity for Notch-1, whereas N1/ModF/5 showed dual specificity for Notch-1 and Notch-3. During kinetic studies, N1/ModF/2 exhibited the highest affinity for Notch-1 (K_{Dapp} = 0.16 nM; Figure 4F; Supplementary Table 4). N1/ModF/2 possessed ~15-fold higher affinity than the Notch-1 specific antibody anti-NRR1 reported previously (K_D = 2.5 nM; Wu et al., 2010). With a sub-nM K_D and strict specificity, N1/ModF/2 turned out to be the best Fab clone for Notch-1. N1/ModF/5 bound to Notch-1 and Notch-3 with K_{Dapp} values of 0.68 nM and 7.1 nM, respectively.

Discussion

Given that Library-S delivered Fabs against Jagged-1/2 and Notch-2/3 (Barreto et al., 2019), we tested whether Library-S could give rise to Fabs with high affinity and specificity against Notch-1. Library-S gave rise to a sub-nM bi-specific binder (N1/S/1) and a mid-nM mono-specific binder (N1/S/2). Next, we split Library-S into four sub-libraries each containing a set of



E

Clone	L3	H1	H2	H3
N1/ModF/1	QQPSSHSAALIT	LYSSYMH	YISPSYSYTY	ARWGHSPSGVGWYGALDY
N1/ModF/2	QQVSAGYGPPIT	LYSSYIH	YISPYSGSTY	ARYGYAYHPYHSGYWYGALDY
N1/ModF/3	QQYWWWGWYPIT	IYYYYIH	SIYPSSGSTS	ARHYAPWSYASAHFHWWALDY
N1/ModF/4	QQSGYAWWYPIT	LYSSYMH	SIYPSYGYTY	ARYYAWWYGYYSWYHAHGFDY
N1/ModF/5	QQGPYHGYSLIT	ISSSYMH	SIYSYSGSTY	ARSGYSGVSGSVYSYWPFYGLDY
N1/ModF/6	QQGGGYSFGLFT	LYSYSMH	SIYSSSGYTY	ARHGGHHWSGYWWSWYGSYGMDY
N1/ModF/7	QQSYYWSAHPIT	LYSSYMH	SISPYYSYTS	ARWGGYVYAYVWIPYWAVYGLDY

F

Clone	L3 length	H3 length	EC_{50} (nM)	K_D (nM)
N1/ModF/2	12	21	1.68 ± 0.22	0.16 ± 0.10
N1/ModF/5	12	23	2.44 ± 0.60	0.68 ± 0.13

FIGURE 4
Notch-1 Fabs from the Modified-F library. **(A)** Pairing between CDRL3 lengths and CDRH3 lengths in the Notch-1 selection output (round 3) from Library-F selections. L3-H3 pairs were obtained from L3-H1-H2-H3 NGS information, and for each possible L3 length, H3 length distribution was generated. **(B)** CDRL3 and CDRH3 length distribution of the naïve Library-F. **(C)** CDRL3 and CDRH3 length distribution of the Modified-F library. **(D)** CDRL3 and CDRH3 length distribution of the Notch-1 selection output after four rounds of selection with the Modified-F library. **(E)** Diversified CDR sequences and phage-displayed Fab binding characteristics of 7 Notch-1 Fabs isolated from Modified-F selections using Sanger sequencing. **(F)** CDRL3 length, CDRH3 length, affinity, and specificity of Fabs N1/ModF/2 and N1/ModF/5. EC_{50} and K_D values were determined by multi-point Fab-ELISA and bio-layer interferometry, respectively. Fab specificity was determined at 1μM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps. Notch-1, Notch-2, Notch-3, Jagged-1, and Jagged-2 are indicated as N1, N2, N3, J1, and J2, respectively.

CDRH3 lengths and panned them against Notch-1. Through sub-library screens, we expected to obtain Fabs with different CDRH3 lengths and Fabs with higher affinities compared to the master-library Fabs. Although results were contrary to our expectations, sub-library panning gave rise to a low-nM mono-specific binder (N1/SL2/1). Its corresponding master-library Fab showed ~5-fold higher affinity for Notch-1 (sub-nM K_D), however, also had a mid-nM K_D for Notch-3. In going from Library-S to Library-F, we anticipated that the additional diversity in Library-F (length diversity in CDRL3 and amino acid diversity in CDRs H1/H2) would generate more potent and selective Fabs for Notch-1. Through Library-F selections, we isolated three high-frequency clones with short CDRH3 lengths by Sanger sequencing and reconstructed three low-frequency clones with medium CDRH3 lengths from NGS information. The high-frequency clones possessed sub-nM to low-nM K_D values for Notch-1 and cross-reacted with two other receptors. Low-frequency clones possessed low-nM to mid-nM K_D values for Notch-1 and cross-reacted with Jagged-2, making them bi-specific binders.

Even though CDRH3 in Library-S had higher length, amino acid, and JH segment diversities than in Library-F, Library-S selections provided only two optimal CDRH3 length solutions for

Notch-1. In Library-F, the addition of length and conformational diversity to CDRs adjacent to CDRH3 increased the number of optimal CDRH3 length solutions for Notch-1. Further, CDRH3 length solutions allowed by Library-F Fabs were distinct from Library-S Fabs, indicating that length/conformational diversity in CDRs L3, H1, and H2 could also alter and determine the type of CDRH3 loop responsible for the interaction. Not only CDR lengths and amino acids, but also binding profiles of Library-F Fabs were different from Library-S Fabs, confirming previous observations that the increase in overall interface diversity fundamentally changes the nature of Fab-binding solutions rather than optimizing a common binding solution (Fellouse et al., 2007).

Since CDR length diversity in Library-F was biased toward short sequences, we designed and constructed the Modified-F library with the length bias toward long CDRs. We used the Modified-F library to isolate two Notch-1 Fabs (N1/ModF/2 and N1/ModF/5) with long CDRL3 and CDRH3 sequences. N1/ModF/2 exhibited the highest affinity for Notch-1 (K_{Dapp} = 0.16 nM) and showed strict specificity for Notch-1. N1/ModF/5 showed low-nM binding to Notch-1 and cross-reacted with Notch-3. Strikingly, binding profiles of Modified-F Fabs were very similar to Library-S Fabs. Only mono-specific Fabs (Notch-1) and bi-specific Fabs (Notch-1 and Notch-3) were isolated from both the libraries. N1/ModF/2 bound ~20-fold stronger than the best Library-S-based mono-specific Fab N1/SL2/1. By possessing the same affinity for Notch-1 (K_{Dapp} = 0.7 nM) but two-fold higher affinity for Notch-3, N1/ModF/5 performed poorly than the Library-S-based bi-specific Fab N1/S/1. While long CDR lengths contributed to the specificity of Modified-F Fabs, we speculate that the additional amino acid diversity contributed to the specificity of Library-S Fabs.

In this work, both single-framework synthetic Fab libraries (S and F) yielded sub-nM to mid-nM Notch-1 binders with different specificity profiles. However, to obtain high-affinity Notch-1 specific Fabs, screening the master library alone was not sufficient. In the case of Library-S, screening sub-libraries identified a low-nM Notch-1 specific binder. In the case of Library-F, fine-tuning the library toward long CDR lengths gave rise to a sub-nM Notch-1 specific binder. In addition to delivering Notch-1 Fabs, our work highlighted the importance of sampling a focused diversity within an antibody library. The master library only has a sparse coverage of a large diversity space, due to limitations in attainable library size. Its sub-libraries provide a denser coverage of a smaller diversity space. In contrast, affinity maturation libraries offer optimization of an existing binding solution and are constructed by random or rational diversification of an existing antibody (Marvin and Lowman, 2015).

Previous studies have shown that NGS has four main applications in antibody phage display: library quality control, analysis of selection outputs, reconstruction of low-frequency clones, and rationale design based on natural repertoires (Ravn et al., 2010; Zhang et al., 2011; Mahon et al., 2013; D'Angelo et al., 2014; Glanville et al., 2015; Yang et al., 2017; Jian et al., 2019; Valadon et al., 2019; Azevedo Reis Teixeira et al., 2021; Wang et al., 2021), which have resulted in some FDA approved antibodies (Alfaleh et al., 2020). In this work, we used NGS information for identifying and reconstructing low-frequency clones with less-frequent CDRH3 lengths, for studying the pairing between CDRL3 and CDRH3 lengths in the selection output, and for fine-tuning the library length diversity. These advances in the use of NGS information gave rise to Fab clones with better binding properties. We used CDRH3 length as a parameter for designing and screening libraries and for isolating Fabs from selection outputs. This approach proved to be successful for obtaining Fabs with different and superior binding profiles that may prove useful in clinical applications.

The isolation of antibodies with new epitopes and binding profiles against Notch pathway members are of interest for designing new antibodies for clinical studies. Antibodies targeting the Notch pathway first entered clinical trials in 2010 and had limited success. One reason for their failure is likely due to the complex nature of this signaling pathway, which can be both pro-and anti-tumor depending on the tumor microenvironment (Zhou et al., 2022). Antibodies targeting Notch1 (brontictuzumab; NCT01703572, NCT01778439, NCT02662608, NCT03031691) or Notch2/3 (tarextumab; NCT01277146, NCT01859741; NCT01647828) receptors have failed in phase I and phase II clinical trials due to lack of efficacy (Ferrarotto et al., 2018; Hu et al., 2019). Clinical trials against Notch receptor ligands DLL3 (rovalpituzumab tesirine; Uprety et al., 2021) and DLL4 (enoticumab (Chiorean et al., 2015) and demcizumab (Coleman et al., 2020)) also showed poor efficacy. Current Notch pathway antibodies in clinical trials, are bi-specific or tri-specific. These antibodies include bispecific DLL4/VEGFA targeting antibodies navicixizumab (NCT05043402) and dilpacimab (NCT01946074), bi-specific targeting DLL3/CD3ε antibody tarlatamab (NCT05060016, NCT05361395, NCT04702737, and NCT04885998) and the tri-specific targeting DLL3/CD3ε/albumin antibody HPN328 (NCT04471727). Multi-specific antibodies should improve their tumor targeting and the targeting of T-cells to tumors.

In summary, we used synthetic antibody technology for generating selective Fabs against the extracellular domain of Notch-1. Twelve Fabs with 10 different CDRH3 lengths were identified from single-framework synthetic Fab libraries using phage display. Upon testing, Fabs showed high affinity for Notch receptors (sub-nM to mid-nM K_{Dapp} values) and exhibited different binding profiles (mono-or bi-or tri-specific) toward Notch/Jagged receptors. Most likely, these Fabs recognize different epitopes on Notch-1 and could be used for modulating the Notch signaling pathway using different mechanisms of action. Two Fabs exhibited strict specificity for Notch-1 with low nanomolar K_{Dapp} values. In contrast to gene knockout approaches, γ-secretase inhibitors, and pan-Notch antibodies (Espinoza and Miele, 2013; Takebe et al., 2014), our mono-specific Fabs may permit a more precise control of Notch-1 inhibition. Over the course of Fab generation, we also showed that implementing NGS approaches, screening focused diversity libraries, and fine-tuning the library

diversity can improve the success rate of single-framework synthetic Fab libraries. These findings have valuable implications for antibody library design, antibody phage display, and combinatorial antibody engineering.

Materials and methods

Construction of single-framework synthetic Fab libraries

Library-S is previously described and contains four-fixed CDRs (L1, L2, H1, and H2) and two-diversified CDRs (L3 and H3) on a modified trastuzumab framework (Maruthachalam et al., 2017).

The modified library (Modified-F) was constructed as follows. The template phagemid used for constructing the Modified-F Fab library was derived from the phagemid encoding the anti-maltose binding protein (MBP) Fab from Library-F (Persson et al., 2013). Kunkel mutagenesis (Kunkel et al., 1987) was used to replace four CDRs (L3, H1, H2, and H3) in the MBP phagemid with TAA stop codons. The resulting phagemid was sequence verified and used as the template phagemid. The Modified-F library was constructed and stored using previously established protocols (Fellouse and Sidhu, 2006; Rajan and Sidhu, 2012). Kunkel mutagenesis was used to simultaneously diversify four CDR regions (L3, H1, H2, and H3) within the template phagemid. CDRs L3, H1, and H2 were diversified using Library-F mutagenic oligonucleotides. CDRH3 was diversified using 17 J_H4-CDRH3 Library-F mutagenic oligonucleotides and 10 J_H6-CDRH3 Library-S mutagenic oligonucleotides (Oligonucleotide sequences given in Supplementary Table 1). Eight different mutagenesis reactions were required for constructing the Modified-F library, each reaction representing a set of CDRH3 lengths. DNA from the mutagenesis reaction was purified using the PCR cleanup kit and 10 µg of purified DNA was electroporated into M13KO7-infected SR320 *E. coli* cells for phage production. Phages were purified from the culture supernatant using PEG/NaCl precipitation, resuspended in PBS and stored at-80°C in the presence of protease inhibitors (2%) and sterile glycerol (25%). Phages from eight sub-libraries were rescued separately and an equal number of phages from each sub-library (~5×10^{13} PFU) was mixed together to create the Modified-F library.

Phage display selections

The recombinant Fc-tag-fused human Notch-1 extracellular domain was purchased from R&D Systems. Solid-phase panning of phage-displayed Fab libraries (S, F, and modified-F) was conducted according to previously described protocols (Fellouse and Sidhu, 2006; Rajan and Sidhu, 2012). Briefly, phages from the frozen master library were precipitated, deselected for binding to the Fc protein and cycled through rounds of binding selection with Notch-1 coated on 96-well MaxiSorp plates and amplification of Notch-1-bound phage in XL1-Blue *E. coli* cells. After four rounds of selections, phage clones were plated as individual colonies for isolation, sequencing, and manipulation of phagemid DNA.

Ion Torrent sequencing and data analysis

Ion Torrent sequencing of one diversified CDR (L3 or H3) was accomplished in three steps: PCR amplification of CDR, emulsion PCR on Ion sphere particles (ISPs), and sequencing enriched ISPs on an Ion semiconductor chip. To PCR amplify CDRs from phage pools, we designed primers that hybridize to the fixed framework regions of the phagemid that flank the CDR region. Primers contain barcodes for multiplexing purposes and adapter sequences to facilitate emulsion PCR. We PCR-amplified the CDR of interest from phage samples, checked the purity, concentration, and length of PCR products using a 2,100 bio-analyzer (Agilent Technologies), prepared the template for emulsion PCR by pooling multiple PCR products, performed emulsion amplification of the amplicon library on the Ion OneTouch 2 instrument (Life Technologies) using the Ion OneTouch template kit, loaded the enriched ISPs into an Ion 314 Semiconductor chip, and sequenced the loaded ISPs on the V2 Ion Personal Genome Machine (Thermo Scientific) using the Ion PGM supplies kit.

Ion Torrent sequencing of the L3-H1-H2-H3 CDR strip was accomplished in six steps. (1) ssDNA was extracted from amplified phage selection outputs (10^{13} PFU) using the Spin M13 kit (Qiagen). (2) 500 ng of ssDNA was subjected to Kunkel mutagenesis for deleting the framework regions between four diversified CDRs. In the mutagenesis reaction, three oligonucleotides, L3-H1 Seq, H1-H2 Seq, and H2-H3 Seq, were used to link the L3-H1-H2-H3 regions together. Phosphorylation of oligonucleotides, annealing of oligonucleotides to the ssDNA template, and *in vitro* synthesis of CCC-dsDNA were carried out as described previously (Tonikian et al., 2007; Nelson and Sidhu, 2012). (3) DNA from the mutagenesis reaction was run on an agarose gel and the right-sized product (CCC-dsDNA) was excised and purified using a gel extraction kit. (4) The L3-H1-H2-H3 CDR strip was PCR amplified from 50 ng of purified CCC-dsDNA using barcoded L3-Fwd and H3-Rev primers. (5) PCR amplicons were purified, quantified, multiplexed, and subjected to emulsion PCR using the Ion PGM Template OT2 400 kit. (6) Enriched ISPs were loaded on an Ion 314 Chip and sequenced using the Ion PGM Sequencing 400 kit.

We built a custom workflow for NGS data processing and analysis (Barreto et al., 2019). Sequences were base called and separated by the barcode on the Ion PGM Torrent Server and exported in FASTQ format. Sequences were imported into the Galaxy server (Blankenberg et al., 2010; Goecks et al., 2010), where they were trimmed based on quality score (>17), converted to FASTA, and then run on a custom R script (R Core Team, 2013) to parse the CDR, translate, and perform sequence counts. CDR

sequences were processed using the Biostrings package (Pagès et al., 2016) and length distribution plots were generated using the ggplot2 package (Wickham, 2009).

Reconstruction of rare Notch-1 Fabs

To reconstruct rare Notch-1 Fab clones from the CDR strip sequencing information, we cloned desired CDR combinations into the MBP-encoding phagemid by Kunkel mutagenesis (Kunkel et al., 1987). The MBP phagemid CDRs were replaced with *NotI* sites and used as a template to reconstruct Fab clones. Oligonucleotides were designed to encode for a desired CDR sequence and to hybridize to either side of the CDR. Four oligonucleotides were used in the Kunkel mutagenesis reaction, one for each diversified CDR. Phosphorylation of oligonucleotides, annealing of oligonucleotides to the uracil-inserted ssDNA template, and *in vitro* synthesis of CCC-dsDNA were carried out as described previously (Tonikian et al., 2007; Nelson and Sidhu, 2012). Following mutagenesis, the reaction was transformed into *dut*$^+$/*ung*$^+$ *E. coli* to eliminate the wild-type template strand. Positive Fab clones (N1/F/R1, N1/F/R2, and N1/F/R3) were screened by *NotI* restriction digestion analysis.

Sub-cloning, expression, and purification of Notch-1 Fabs

Notch-1 Fabs were sub-cloned from the phagemid vector into a modified pCW-LIC Fab expression vector using standard molecular biology procedures. Briefly, Fab sequences were amplified from phagemids by PCR, and ligated into the *SacI*/*XhoI*-digested pCW-LIC vector using Gibson assembly (Gibson et al., 2009). Gibson assembly reactions were electroporated into BL21 *E. coli* cells, and three colonies from each reaction were screened for Fab expression using bio-layer interferometry in 96-well format. Briefly, single colonies were transferred to 1 ml of Overnight Express Terrific Broth (TB) auto-induction medium (EMD Millipore) supplemented with 100 μg/ml carbenicillin in 96-well deep-well boxes and incubated for 18 h at 25°C and 200 rpm. Cells were pelleted by centrifugation and lysed with 200 μl of B-PER bacterial protein extraction reagent (Pierce). Cells were centrifuged again, and 50 μl of the clarified supernatant was transferred to 384-well plates. Fab expression was detected using anti-Fab C_H1 biosensors and anti-HIS biosensors in the ForteBio Octet RED384 system (Pall Corporation) according to the manufacturer's instructions. Positive clones were transferred into 30 ml of TB auto-induction medium supplemented with 100 μg/ml carbenicillin and incubated for 18 h at 25°C with shaking at 200 rpm. Cells were pelleted by centrifugation, and lysed in protein-L binding buffer (20 mM Na_2HPO_4, 0.15 M NaCl, pH 8) containing 1:100 dilution of protease inhibitor cocktail (Sigma) using a cell disruptor (Constant Systems) Clarified supernatant was incubated with 200 μl of Protein-L resin (GenScript) for 1 h at 4°C. The Protein-L resin was collected by centrifugation and washed 5X with Protein-L binding buffer. Fabs were eluted with IgG elution buffer (Thermo-Scientific) and neutralized with 1 M Tris–HCl (pH 9). Eluted Fabs were dialyzed against PBS and stored at −20°C. Fab purity was verified using a 2,100 bio-analyzer and Fab concentration was determined by UV–visible spectrometry.

Enzyme-linked immunosorbent assays

Phage-ELISA was performed to check the binding of phage-displayed Fabs to immobilized target and control proteins. The Fab-encoding phagemid was electroporated into M13KO7-infected electro-competent SR320 *E. coli* cells for phage production. The cells were rescued with pre-warmed SOC media and incubated for 30 min at 37°C. The culture was transferred to 30 ml of 2YT media supplemented with 100 μg/ml carbenicillin and 25 μg/ml kanamycin and incubated overnight at 37°C with shaking at 200 rpm. Phages were precipitated from the culture supernatant using 6 ml of ice-cold PEG/NaCl solution, resuspended in PBT buffer (PBS containing 0.5% BSA and 0.05% Tween), and quantified using UV spectrometry.

To conduct phage-ELISA, Notch-1 and control proteins were immobilized at 5 μg/ml on MaxiSorp plates (Nunc) by overnight incubation at 4°C. The wells were subsequently blocked with PB buffer (PBS containing 0.5% BSA) for 90 min at RT before washing four times with PT buffer (PBS containing 0.05% Tween). Wells were then exposed to PBT-diluted phage solution (10^{12} PFU/ml) for 30 min, washed 8X with PT buffer, and then incubated with a 1:3,000 dilution of HRP-conjugated anti-M13 antibody (GE healthcare) for 30 min at RT. Plates were washed again 6X with PT buffer and 2X with PBS. Wells were developed with 3,3′,5,5′-tetramethylbenzidine (TMB) substrate for 5 min and quenched with an equal volume of 1 M H_3PO_4. The plates were read at 450 nm using a SpectraMax 340PC plate reader (Molecular devices). Phage-ELISA values are provided in Supplementary Table 2.

Fab-ELISA was performed to check the binding of purified Fabs to immobilized target proteins. Proteins were immobilized at 5 μg/ml on MaxiSorp plates by overnight incubation at 4°C. Wells were subsequently blocked with PB buffer for 90 min at RT before washing 4X with PT buffer. Wells were then exposed to 100 μl of Fab solution diluted in PT for 30 min, washed 10X with PT buffer, and then incubated with a 1:3,000 dilution of HRP-conjugated anti-HIS antibody (Rockland Biosciences) for 30 min at RT. Plates were washed again X6 with PT buffer and X2 with PBS. Wells were developed with TMB substrate for 5 min and quenched with an equal volume of 1 M H_3PO_4. Plates were read at 450 nm using a SpectraMax 340PC plate reader (Molecular devices). Single-point Fab-ELISA was used to assess Fab specificity at 1 μM Fab concentration, and multi-point Fab-ELISA was used to calculate the EC_{50} for Fab binding to the immobilized target. In multi-point Fab ELISA, ABS_{450} values were obtained for a range of Fab

concentrations (0.01 nM–1 μM) and the EC_{50} was calculated by fitting the data to the one-site specific-binding equation in Prism (Graphpad). Single-point Fab-ELISA values are provided in Supplementary Table 3 and multi-point Fab ELISA fits are provided in Supplementary Figure 1.

Bio-layer interferometry

The ForteBio Octet RED384 system (Pall Corporation) was used to measure the binding kinetics between purified Notch-1 Fabs and target proteins. Fabs were immobilized on amine-reactive generation-2 biosensors or anti-Fab CH1 biosensors according to the manufacturer's instructions. Immobilized Fabs were exposed to increasing concentrations of target proteins, and association and dissociation rates were measured by the shift in wavelength (nm). All reactions were performed at 25°C in PBS. For each sensor-immobilized Fab, at least three different target protein concentrations were used, and K_D (equilibrium dissociation constant) was obtained by globally fitting the data to a 1:1 binding model. Data were collected with Octet Data Acquisition version 7.1.0.87 (ForteBio) and analyzed using Octet Data Analysis version 7.1 (ForteBio). The list of K_{on}, K_{off} and K_{Dapp} values is provided in Supplementary Table 4.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

BM and CG conceived the idea. CG supervised the entire study. BM conducted the experiments and analyzed the results with KB, DH, and AK. AK and KB supervised bioinformatics experiments. BM and CG wrote the manuscript with additions from KB. All authors contributed to the article and approved the submitted version.

Funding

Funding for this work came from Western Economic Diversification Canada 12939, Natural Sciences and Engineering Research Council, grant number 027-25634, and Royal University Hospital Foundation – Nutrien Clinical Research Chair. BM thanks the Department of Biochemistry, University of Saskatchewan and the Government of Saskatchewan for Graduate Scholarships and Awards.

Acknowledgments

The authors thank Sachdev S. Sidhu (University of Toronto) for providing Library-F phage and Library-F mutagenic oligonucleotides.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.931307/full#supplementary-material

References

Alfaleh, M. A., Alsaab, H. O., Mahmoud, A. B., Alkayyal, A. A., Jones, M. L., Mahler, S. M., et al. (2020). Phage display derived monoclonal antibodies: from bench to bedside. *Front. Immunol.* 11:1986. doi: 10.3389/fimmu.2020.01986

Azevedo Reis Teixeira, A., Erasmus, M. F., D'Angelo, S., Naranjo, L., Ferrara, F., Leal-Lopes, C., et al. (2021). Drug-like antibodies with high affinity, diversity and developability directly from next-generation antibody libraries. *MAbs* 13, 1980942. doi: 10.1080/19420862.2021.1980942

Barreto, K., Maruthachalam, B. V., Hill, W., Hogan, D., Sutherland, A. R., Kusalik, A., et al. (2019). Next-generation sequencing-guided identification and reconstruction of antibody CDR combinations from phage selection outputs. *Nucleic Acids Res.* 47:e50. doi: 10.1093/nar/gkz131

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., et al. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chapter 19, Unit 19.10.1-21, doi: 10.1002/0471142727.mb1910s89

Chiorean, E. G., LoRusso, P., Strother, R. M., Diamond, J. R., Younger, A., Messersmith, W. A., et al. (2015). A phase I first-in-human study of enoticumab (REGN421), a fully human delta-like ligand 4 (Dll4) monoclonal antibody in patients with advanced solid tumors. *Clin. Cancer Res.* 21, 2695–2703. doi: 10.1158/1078-0432.CCR-14-2797

Coleman, R. L., Handley, K. F., Burger, R., Molin, G. Z. D., Stagg, R., Sood, A. K., et al. (2020). Demcizumab combined with paclitaxel for platinum-resistant ovarian, primary peritoneal, and fallopian tube cancer: the SIERRA open-label phase Ib trial. *Gynecol. Oncol.* 157, 386–391. doi: 10.1016/j.ygyno.2020.01.042

D'Angelo, S., Kumar, S., Naranjo, L., Ferrara, F., Kiss, C., and Bradbury, A. R. (2014). From deep sequencing to actual clones. *Protein Eng. Des. Sel.* 27, 301–307. doi: 10.1093/protein/gzu032

Espinoza, I., and Miele, L. (2013). Notch inhibitors for cancer treatment. *Pharmacol. Ther.* 139, 95–110. doi: 10.1016/j.pharmthera.2013.02.003

Fellouse, F. A., and Sidhu, S. S. (2006). "Making antibodies in bacteria," in *Making and Using Antibodies: A Practical Handbook*. eds. G. C. Howard and M. R. Kaser (San Francisco, CA: Taylor and Francis), 157–180.

Fellouse, F. A., Esaki, K., Birtalan, S., Raptis, D., Cancasci, V. J., Koide, A., et al. (2007). High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J. Mol. Biol.* 373, 924–940.

Ferrarotto, R., Eckhardt, G., Patnaik, A., LoRusso, P., Faoro, L., Heymach, J. V., et al. (2018). A phase I dose-escalation and dose-expansion study of brontictuzumab in subjects with selected solid tumors. *Ann. Oncol.* 29, 1561–1568. doi: 10.1093/annonc/mdy171

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. 3rd, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318

Glanville, J., D'Angelo, S., Khan, T. A., Reddy, S. T., Naranjo, L., Ferrara, F., et al. (2015). Deep sequencing in library selection projects: what insight does it bring? *Curr. Opin. Struct. Biol.* 33, 146–160. doi: 10.1016/j.sbi.2015.09.001

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11, R86. doi: 10.1186/gb-2010-11-8-r86

Hu, Z. I., Bendell, J. C., Bullock, A., LoConte, N. K., Hatoum, H., Ritch, P., et al. (2019). A randomized phase II trial of nab-paclitaxel and gemcitabine with tarextumab or placebo in patients with untreated metastatic pancreatic cancer. *Cancer Med.* 8, 5148–5157. doi: 10.1002/cam4.2425

Jian, J. W., Chen, H. S., Chiu, Y. K., Peng, H. P., Tung, C. P., Chen, I. C., et al. (2019). Effective binding to protein antigens by antibodies from antibody libraries designed with enhanced protein recognition propensities. *MAbs* 11, 373–387. doi: 10.1080/19420862.2018.1550320

Katoh, M., and Katoh, M. (2020). Precision medicine for human cancers with notch signaling dysregulation (review). *Int. J. Mol. Med.* 45, 279–297. doi: 10.3892/ijmm.2019.4418

Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* 154, 367–382. doi: 10.1016/0076-6879(87)54085-X

Lafkas, D., Shelton, A., Chiu, C., de Leon Boenig, G., Chen, Y., Stawicki, S. S., et al. (2015). Therapeutic antibodies reveal notch control of transdifferentiation in the adult lung. *Nature* 528, 127–131. doi: 10.1038/nature15715

Lee, C. V., Sidhu, S. S., and Fuh, G. (2004). Bivalent antibody phage display mimics natural immunoglobulin. *J. Immunol. Methods* 284, 119–132. doi: 10.1016/j.jim.2003.11.001

Mahon, C. M., Lambert, M. A., Glanville, J., Wade, J. M., Fennell, B. J., Krebs, M. R., et al. (2013). Comprehensive interrogation of a minimalist synthetic CDR-H3 library and its ability to generate antibodies with therapeutic potential. *J. Mol. Biol.* 425, 1712–1730. doi: 10.1016/j.jmb.2013.02.015

Maruthachalam, B. V., El-Sayed, A., Liu, J., Sutherland, A. R., Hill, W., Alam, M. K., et al. (2017). A single-framework synthetic antibody library containing a combination of canonical and variable complementarity-determining regions. *Chembiochem* 18, 2247–2259. doi: 10.1002/cbic.201700279

Marvin, J. S., and Lowman, H. B. (2015). "Antibody humanization and affinity maturation using phage display," in *Phage Display in Biotechnology and Drug Discovery*. eds. S. S. Sidhu and C. R. Geyer (Boca Raton, FL: Taylor and Francis), 347–372.

Nelson, B., and Sidhu, S. S. (2012). Synthetic antibody libraries. *Methods Mol. Biol.* 899, 27–41. doi: 10.1007/978-1-61779-921-1_2

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2016). Biostrings: string objects representing biological sequences, and matching algorithms. R package version 2.42.1.

Persson, H., Ye, W., Wernimont, A., Adams, J. J., Koide, A., Koide, S., et al. (2013). CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J. Mol. Biol.* 425, 803–811. doi: 10.1016/j.jmb.2012.11.037

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Satistical Computing, Vienna.

Rajan, S., and Sidhu, S. S. (2012). Simplified synthetic antibody libraries. *Methods Enzymol.* 502, 3–23. doi: 10.1016/B978-0-12-416039-2.00001-X

Ranganathan, P., Weaver, K. L., and Capobianco, A. J. (2011). Notch signalling in solid tumours: a little bit of everything but not all the time. *Nat. Rev. Cancer* 11, 338–351. doi: 10.1038/nrc3035

Ravn, U., Gueneau, F., Baerlocher, L., Osteras, M., Desmurs, M., Malinge, P., et al. (2010). By-passing in vitro screening-next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Res.* 38:e193. doi: 10.1093/nar/gkq789

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352. doi: 10.1038/nature10242

Takebe, N., Nguyen, D., and Yang, S. X. (2014). Targeting notch signaling pathway in cancer: clinical development advances and challenges. *Pharmacol. Ther.* 141, 140–149. doi: 10.1016/j.pharmthera.2013.09.005

Tonikian, R., Zhang, Y., Boone, C., and Sidhu, S. S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat. Protoc.* 2, 1368–1386. doi: 10.1038/nprot.2007.151

Uprety, D., Remon, J., and Adjei, A. A. (2021). All That glitters is not gold: the story of Rovalpituzumab Tesirine in SCLC. *J. Thorac. Oncol.* 16, 1429–1433. doi: 10.1016/j.jtho.2021.07.012

Valadon, P., Pérez-Tapia, S. M., Nelson, R. S., Guzmán-Bringas, O. U., Arrieta-Oliva, H. I., Gómez-Castellano, K. M., et al. (2019). ALTHEA gold libraries™: antibody libraries for therapeutic antibody discovery. *MAbs* 11, 516–531. doi: 10.1080/19420862.2019.1571879

Wang, H., Yan, K., Wang, R., Yang, Y., Shen, Y., Yu, C., et al. (2021). Antibody heavy chain CDR3 length-dependent usage of human IGHJ4 and IGHJ6 germline genes. *Antib Ther.* 4, 101–108. doi: 10.1093/abt/tbab010

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, NY.

Wu, Y., Cain-Hom, C., Choy, L., Hagenbeek, T. J., de Leon, G. P., Chen, Y., et al. (2010). Therapeutic antibody targeting of individual notch receptors. *Nature* 464, 1052–1057. doi: 10.1038/nature08878

Yang, W., Yoon, A., Lee, S., Kim, S., Han, J., and Chung, J. (2017). Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. *Exp. Mol. Med.* 49:e308. doi: 10.1038/emm.2017.22

Zhou, B., Lin, W., Long, Y., Yang, Y., Zhang, H., Wu, K., et al. (2022). Notch signaling pathway: architecture, disease, and therapeutics. *Signal Transduct. Target. Ther.* 7, 95. doi: 10.1038/s41392-022-00934-y

Zhang, H., Torkamani, A., Jones, T. M., Ruiz, D. I., Pons, J., Lerner, R. A. (2011). Phenotype- information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against complex systems. *Proc. Natl. Acad. Sci. USA* 108, 13456–13461.

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 08 August 2022
DOI 10.3389/fmicb.2022.961093

OPEN ACCESS

EDITED BY
Yoichi Takakusagi,
National Institutes for Quantum and Radiological Science and Technology, Japan

REVIEWED BY
Piergiuseppe De Berardinis,
Consiglio Nazionale delle Ricerche (Bologna), Italy
Noah Reynolds,
University of Illinois at Springfield, United States

*CORRESPONDENCE
Brian K. Kay
bkay@uic.edu

SPECIALTY SECTION
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

RECEIVED 03 June 2022
ACCEPTED 04 July 2022
PUBLISHED 08 August 2022

CITATION
Allen GL, Grahn AK, Kourentzi K, Willson RC, Waldrop S, Guo J and Kay BK (2022) Expanding the chemical diversity of M13 bacteriophage. *Front. Microbiol.* 13:961093. doi: 10.3389/fmicb.2022.961093

Expanding the chemical diversity of M13 bacteriophage

Grace L. Allen[1], Ashley K. Grahn[1], Katerina Kourentzi[2], Richard C. Willson[2], Sean Waldrop[3], Jiantao Guo[3] and Brian K. Kay[1]*

[1]Tango Biosciences, Inc., Chicago, IL, United States, [2]Department of Chemical and Biomolecular Engineering, University of Houston, Houston, TX, United States, [3]Department of Chemistry, University of Nebraska at Lincoln, Lincoln, NE, United States

Bacteriophage M13 virions are very stable nanoparticles that can be modified by chemical and genetic methods. The capsid proteins can be functionalized in a variety of chemical reactions without loss of particle integrity. In addition, Genetic Code Expansion (GCE) permits the introduction of non-canonical amino acids (ncAAs) into displayed peptides and proteins. The incorporation of ncAAs into phage libraries has led to the discovery of high-affinity binders with low nanomolar dissociation constant (K_D) values that can potentially serve as inhibitors. This article reviews how bioconjugation and the incorporation of ncAAs during translation have expanded the chemistry of peptides and proteins displayed by M13 virions for a variety of purposes.

KEYWORDS

bioconjugation, phage-display, stop codon suppression, peptide, cyclization, cross-linking, combinatorial peptide libraries, antibody fragments

Introduction

Since the seminal work of Professor George Smith (Smith, 1985) in displaying a protein fragment on the surface of M13 bacteriophage, a large number of peptides and proteins have been displayed for antibody discovery, protein engineering, and mapping protein-protein interactions. While phage-display is a prize-worthy technique (Smith, 2019), it has generally been limited in chemistry to the canonical set of 20 amino acids of L-chirality. Both chemical modification and genetic methods have been used to expand the types of functional groups displayed on the surface of virions.

Chemically modified virions have been used in a variety of applications (Mohan and Weiss, 2016), such as lateral flow assays (Hagström et al., 2015; Kim et al., 2015, 2017), biosensors (Moon et al., 2019), nanomaterials (Petrenko, 2018), nanomedicine (Ulfo et al., 2022), and batteries (Lee et al., 2009). In M13 bacteriophage (**Figure 1**), there are ~2,700 copies of the major capsid protein, pVIII, which makes it a desirable target for bulk chemical modification of virions (Kehoe and Kay, 2005). The major capsid protein is 50 amino acids long and it contains a number of residues with functional

groups suitable for bioconjugation, such as amines, carboxylic acids, and phenols (Carmody et al., 2021). The reactivity of these functional groups is dependent on steric accessibility, ionization state, and solvent conditions. For example, the ε-amino group of the lysine at position 8 can be derivatized with glutaraldehyde or N-hydroxysuccinimide esters (NHS-esters) to attach fluorescent dyes, biotin, drugs, DNA, enzymes, or gold nanoparticles (Carmody et al., 2021), although the N-terminus of pVIII is preferentially targeted because of its higher solvent accessibility and lower pK_a value (Li et al., 2010). While pVIII lacks cysteines, thiolation of primary amines with 2-Iminothiolane (Traut's Reagent) generates sulfhydryl groups that can be coupled to maleimide-activated antibodies or enzymes. Moreover, ε-amino groups are reactive toward aldehydes that are easily generated by mild periodate-mediated oxidation of the sugars on antibodies or horseradish peroxidase (HRP), enabling favorable conjugation directed away from binding sites and active sites of antibodies or HRP, respectively (Adhikari et al., 2013, 2015). In another scheme, using 4-formyl succinimidyl benzoate, amine groups are converted to aromatic aldehydes that readily react with hydrazide derivatized DNA under mild conditions (Domaille et al., 2013). Carboxylate groups at the C-terminus and within aspartic (D) and glutamic (E) acids of virion coat proteins can be activated with a carbodiimide crosslinker [e.g., EDC (1-ethyl-3-(3-(dimethylamino) propyl)-carbodiimide hydrochloride)]. It is also possible to incorporate a methionine analog, L-azidohomoalanine (Urquhart et al., 2016), in strains of *Escherichia coli* that are auxotrophic for methionine and produce virions displaying hundreds of azide groups for downstream chemical conjugation.

Peptides and proteins that have been displayed on the surface of virions have also been the target of chemical or enzymatic modification. One can engineer a free cysteine (Junutula et al., 2008) for bioconjugation or, alternatively, disulfides in a displayed peptide, which can be reduced and then reacted with a cross-bridging molecule (i.e., linchpin), thereby creating macrocyclic or bicyclic peptides for the purpose of discovering novel peptide ligands of target proteins (Ng and Derda, 2016; Deyle et al., 2017; Chen et al., 2021; Ekanayake et al., 2021). Tyrosine (Y) in pVIII can be selectively activated by laccase to produce a free radical species that can be conjugated to acrylates (Vignali et al., 2018). Virions that have been engineered to display the AviTag (Scholle et al., 2004) or sortase tag (Hess et al., 2012) can be biotinylated by BirA or ligated to a variety of labeled peptides with sortase, respectively. Promising directions in the future will be to exploit the toolbox of enzymes capable of protein ligation (Nuijens et al., 2019; Weeks and Wells, 2020) and the SpyTag/SpyCatcher system (Keeble and Howarth, 2020) to build novel virion structures.

Genetic methods have been separately applied to expanding the chemical space of virions. They are largely based on the pioneering work of Professor Peter Schultz's research group (Liu et al., 1997; Xiao and Schultz, 2016; Young and Schultz, 2018) on recoding the amber codon (TAG) so that it is recognized by a mutant, suppressor tRNA molecule that is charged with an non-canonical amino acid (ncAA). **Figure 2** illustrates the basics of tRNA suppression. One first uses molecular biology techniques to engineer a TAG mutation at a site in the coding region of a protein where the ncAA is desired. When the gene is transcribed, the UAG codon is recognized as a stop codon by the ribosome, thereby terminating translation and yielding a truncated protein; however, if a tRNA has an anti-codon (CUA) that can base pair with the UAG codon in an mRNA, the amino acid attached to the 3' end of the tRNA will form a peptide bond with the nascent peptide in the ribosome, allowing translation to continue and yield a full-length protein. The introduced ncAA can be any that is compatible with charging by an aminoacyl-tRNA synthetase (aaRS). The amino acid pocket of the pyrrolysyl-tRNA synthetase has been shown to be remarkably malleable to engineering recognition and charging of its cognate tRNA with diverse ncAAs (Wan et al., 2014; Tharp et al., 2018). Stop codon suppression has also been achieved with opal (UGA) (Anderson and Schultz, 2003) and ochre (UAA) (Italia et al., 2019) codons. In addition to suppression of stop codons, an analogous method with quadruplet codons (Magliery et al., 2001; Anderson and Schultz, 2003; Anderson et al., 2004; Neumann et al., 2010; Niu et al., 2013; Wang et al., 2014) has been developed (**Figure 2**). Mutations in both the aaRS and tRNA, as well as the use of engineered bacterial hosts (Chatterjee et al., 2014), have enhanced suppression efficiency of a variety of quadruplet codons (Guo and Niu, 2022). To date, >200 different ncAAs have been introduced into different proteins by this technique (Xiao and Schultz, 2016), which has been termed Genetic Code Expansion (GCE). Some of GCE's applications have been to probe the structure and function of proteins, alter their redox potential, introduce fluorophores, infrared, and spin label probes, encode post-translational modifications, and create sites for site-specific bioconjugation. Several recent reviews of how GCE has been used to modify proteins expressed in bacteria, yeast, and mammalian cells can be found elsewhere (Young and Schultz, 2018; Chung et al., 2020; Nikić-Spiegel, 2020; Manandhar et al., 2021; Ros et al., 2021; Shandell et al., 2021; Sanders et al., 2022).

GCE has been applied to phage-display. **Figure 3** shows an *E. coli* cell containing three circular genomes: a plasmid encoding the orthogonal tRNA and cognate aaRS, a phagemid that carries a truncated form of capsid protein III (pIII), which is fused to the coding region of the displayed peptide or protein containing the suppressible amber codon, and an M13 helper virus. The helper virus encodes 10 proteins necessary for viral replication and assembly; it also contains a mutation that leads to preferential packaging of the phagemid genome over the helper virus genome. Secreted virions will display the recombinant peptide or protein and ncAA, only if suppression is successful. The culture medium is supplemented with the ncAA, where it enters the cell and is used by the engineered aaRS to

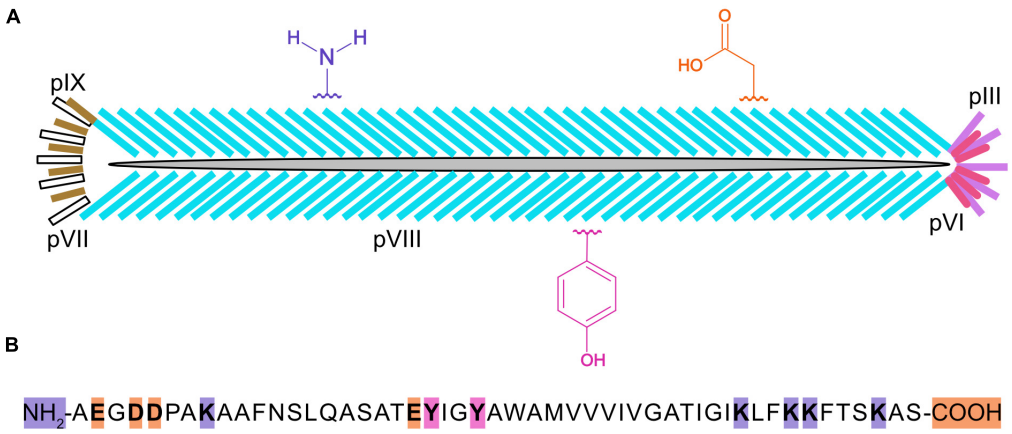


FIGURE 1
Cartoon of an M13 bacteriophage virion, three types of chemically reactive groups, and primary structure of the major capsid protein, pVIII. Each virus particle **(A)** is 900 nanometers long and 7 nanometers wide and contains one single-stranded, circular DNA molecule, and five copies of pIX (yellow), pVII (white), pIII (purple), and pVI (pink), and ~2,700 copies of pVIII (blue). Chemical groups that are available for conjugation include the α-amino groups on the N-terminus and ε-amino groups of lysines (K) of pVIII, the carboxylate groups at the C-terminus and aspartic (D) and glutamic (E) acids, and the phenol groups of tyrosine (Y) are shown in violet, orange, and fushia, respectively. The primary structure for mature pVIII **(B)** is shown with the N-terminus, C-terminus, and residues theoretically capable of being chemically modified highlighted in color. It should be noted that the lysines in the anchoring region (KLFKKFTSKAS) of pVIII may not be sterically accessible for chemical modification.

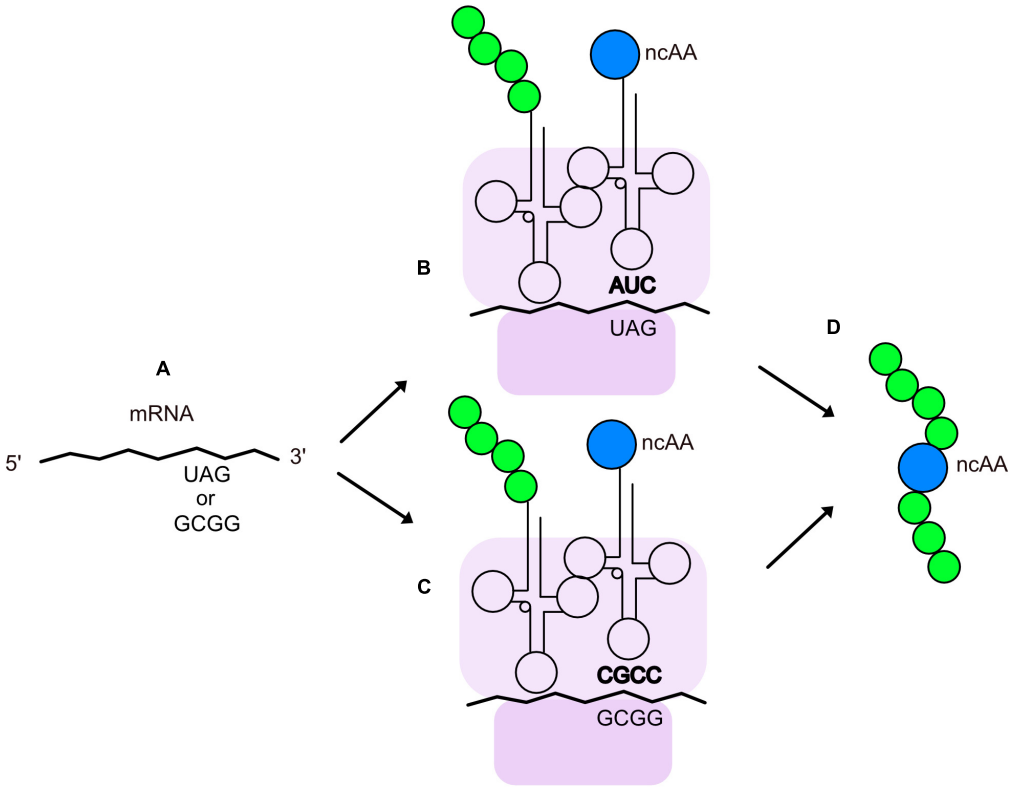


FIGURE 2
Insertion of ncAAs into peptides or proteins by GCE. **(A)** An mRNA containing an engineered quadruplet or stop codon for insertion of an ncAA at a designated location in a protein. During translation, the ribosome (purple) decodes the engineered stop codon **(B)** or quadruplet codon **(C)** with a tRNA charged with the ncAA (blue). The tRNA anticodon is shown in bold and the growing polypeptide chain is shown in green. **(D)** Peptide or protein with an ncAA at the desired location.

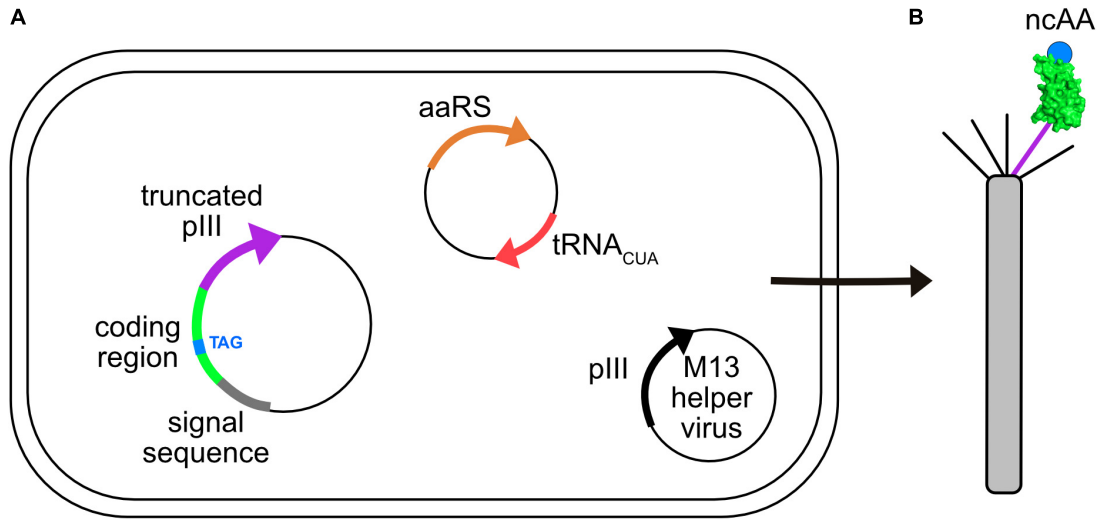


FIGURE 3
Incorporation of ncAAs into phage-displayed constructs. **(A)** *E. coli* cell showing three genomes: phagemid with suppressible mutation (TAG, blue) in the coding region (green) of a protein fused to a truncated form of capsid protein pIII (purple), a plasmid encoding an orthogonal aaRS/tRNA$_{CUA}$ pair (orange/red), and an M13 helper virus (i.e., K07) encoding full length pIII (black). Each of these genomes carry a different antibiotic resistance marker to allow for selective growth of bacterial cells containing all three. **(B)** Virions secreted from such cells contain both wild type (black) and recombinant (purple) pIII capsid proteins, the latter of which display the protein of interest (green) incorporating the desired ncAA (blue).

charge the suppressor tRNA. **Table 1** lists some examples of publications describing the incorporation of ncAAs into virions, which are also summarized in more detail below.

The earliest experiments of engineered replacement of amino acids with M13 bacteriophage took advantage of its life cycle properties. Scientists at New England Biolabs inserted the opal (TGA) stop codon and a downstream selenocysteine insertion sequence (SECIS) upstream of the signal sequence in pIII (Sandman and Noren, 2000). Viral secretion was dependent on the addition of selenium to the culture medium and the incorporation of selenocysteine was confirmed by chemical reactivity. The efficiency of incorporation was influenced by the choice of nucleotide downstream of the opal codon, although the frequency of TGG revertants, encoding tryptophan, was high. Later, the authors (Beech et al., 2015) utilized the selenocysteine-displaying virions to chemically attach five different adenosine receptor ligands and successfully demonstrated that the pentavalently decorated virions could activate the adenosine A1 receptor, a G protein-coupled receptor (GPCR), of cultured cells. An early example of utilizing stop codon suppression to control the replacement of an amino acid was that of the Schultz group (Pastrnak and Schultz, 2001). They introduced an amber codon *in lieu* of the asparagine (N) in a 10-mer peptide sequence (PASTTNKDKL) at the N-terminus of pIII of a phage genome. Virions were propagated in an *E. coli* strain that carried the supE mutation, which encodes a suppressor tRNA that is charged with asparagine. The virions were then used to infect a second bacterial strain that carried a plasmid encoding a yeast tRNA and its cognate aaRS that inserted N into the 10-mer peptide. Only virions secreted by this strain bound well to a monoclonal antibody that had been generated against the 10-mer peptide, demonstrating successful suppression of the stop codon. This proof-of-concept experiment set the stage for engineering aaRSs that could charge their cognate tRNA with ncAAs and suppress amber codons inserted into pIII of virions. Phage-assisted continuous evolution (PACE), phage-assisted non-continuous evolution (PANCE), and phage- and robotic-assisted near-continuous evolution (PRANCE) have enabled rapid laboratory evolution of orthogonal aaRSs over hundreds of generations of mutation, selection, and replication (Esvelt et al., 2011; Bryson et al., 2017; Miller et al., 2020; DeBenedictis et al., 2021, 2022; Fischer et al., 2022). As coat protein pIII is essential for phage, this selection scheme directly links the production and titer of virions with the efficiency of stop codon suppression and incorporation of the desired ncAA in a stop codon inserted into gene III.

The Schultz group was among the first to incorporate ncAAs into displayed peptides and proteins. A variety of ncAAs have been incorporated into peptides permitting bioconjugation (Tian et al., 2004), binding to metal ions (Day et al., 2013), and replacing Zn(II) with Fe(II) for DNA binding domains (Kang et al., 2014). Non-canonical amino acids have also been incorporated into human single-chain variable fragments (scFvs) displayed on virions. In fact, a sulfotyrosine has been incorporated into an scFv that contributed to binding to HIV glycoprotein 120 (Liu et al., 2008). The electrophilic

TABLE 1 Examples of GCE involving M13 virions in the literature.

Peptide or protein displayed on M13 virions	ncAA incorporated through GCE	Outcome	Reference
Combinatorial peptide library	Selenocysteine	Proof-of-principle experiment	Sandman and Noren, 2000
10-mer peptide epitope	Gln with Asp	Directed evolution of aminoacyl-tRNA synthetase	Pastrnak and Schultz, 2001
Short peptide	O-methyl-tyrosine p-azidophenylalanine p-acetylphenylalanine p-benzoylphenylalanine 3-(2-naphthyl)alanine	Proof-of-principle experiments demonstrated fluorescence labeling	Tian et al., 2004
scFv (TAG codon at position 111 of V_H CDR3)	Sulfotyrosine para-acetyl-phenylalanine bipyridyl-alanine 4-borono-phenylalanine	Affinity selection of a sulfotyrosine-containing antibodies that can bind to gp120	Liu et al., 2008
scFv with six random NNK codons in V_H CDR3	p-boronophenylalanine	Affinity selection of an scFv capable of forming a covalent bond to a sugar	Liu et al., 2009
CX_6C, where X = NNK	Bipyridylalanine	Affinity selection of cyclic Ni^{2+} and Zn^{2+} binding peptides	Day et al., 2013
Five residues in the N-terminal finger of zif268 were randomized to include both canonical amino acids and Bpy-Ala	(2,2'-bipyridin-5-yl)alanine (Bpy-Ala)	DNA and Fe(II) binding domain	Kang et al., 2014
Peptide	Selenocysteine	Covalent attachment to five adenosine receptor ligands and activation of cell signaling pathways with decorated virions	Beech et al., 2015
CX_6Z	N^{ϵ}-acryloyl-lysine	Affinity selection of cyclic peptide ligands and inhibitor for TEV protease	Wang et al., 2019
scFv with TAG and quadruplet codons at five different positions	N^{ϵ}-[((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl]-l-lysine p-propargyloxy-l-phenyl-alanine	Dual labeling of scFv on virions	Oller-Salvia and Chin, 2019
CX_6Z	four phenylalanine derivatives N^{ϵ}-butyryl-lysine N^{ϵ} crotonyl-lysine	Affinity selection of cyclic peptide ligands and inhibitors for Sirtuin 2 and TEV protease	Tharp et al., 2020
ZX_6C	O-(2-bromoethyl)-tyrosine	Affinity selection of cyclic peptide ligands to streptavidin, Kelch-like ECH-associated protein 1 (Keap1), and Sonic Hedgehog (Shh)	Owens et al., 2020

C = cysteine.
X = NNK codons, where N is A, C, G, or T and K is G or T.
Z = ncAA.

amino acid *p*-boronophenylalanine has also been successfully incorporated into an scFv, where it can cross-link to a sugar (Liu et al., 2009).

Insertion of ncAAs into phage-displayed peptides offers a number of experimental opportunities. First, it can increase the chemical diversity of the combinatorial peptide libraries for the purpose of identifying novel peptide ligands to a target protein. Second, the ncAA can serve as a site for bioconjugation of ligands that already bind to a target, such as an enzyme, for the purpose of discovering a peptide inhibitor (**Figure 4A**). For example, in a recent report (Tharp et al., 2020), N^{ε} -butyryl-L-lysine (BuK), thiobutyryl (tBuK), and thiomyristoyl (tMyK) analogs were introduced in a combinatorial 11-mer library for the discovery of inhibitors of the NAD-dependent deacetylase, Sirtuin 2 (Sirt2). The ability to genetically encode post-translational modifications by GCE should prove useful for identifying inhibitors of various enzymes and cellular binding proteins in the future.

Recently, GCE has also been used to incorporate ncAAs for the purpose of cyclizing phage-displayed peptides. This effort represents an alternative route for cyclizing phage-displayed peptides to generate macrocycles with novel structures and the potential to inhibit protein-protein interactions in therapeutic applications (Deyle et al., 2017). In the original approach (Heinis et al., 2009), a library of peptides with three fixed cysteines was reduced and then reacted with a trifunctional compound, termed a "linchpin," thereby creating bicyclic peptides. GCE has been used for the same purpose: researchers have built libraries of virions displaying CX_6Z (Wang et al., 2019; Tharp et al., 2020) and ZX_6C (Owens et al., 2020), where X represents an amino acid encoded by NNK codons and Z is an ncAA encoded by the TAG codon that is capable of reacting with the adjacent cysteine residue to form a macrocycle (**Figure 4B**). These libraries were screened by affinity selection and yielded nanomolar inhibitors of Tobacco Etch Virus (TEV) protease (Wang et al., 2019), Sirt2 (Tharp et al., 2020), and Kelch-like ECH-associated protein 1 (Keap1) and Sonic Hedgehog (Shh) (Owens et al., 2020). The

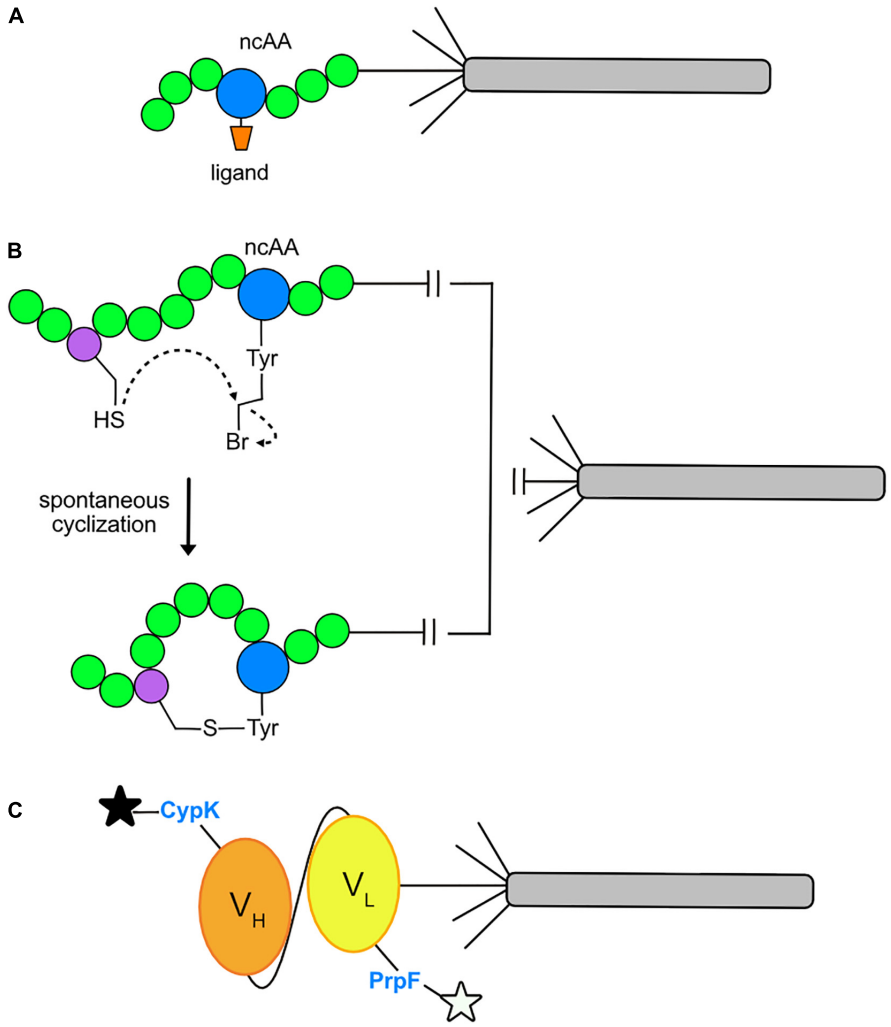


FIGURE 4
Current applications of GCE in phage display. **(A)** Conjugation of a ligand to an ncAA allows direct binding to the active site of a target enzyme. A ligand (orange) may be conjugated to an ncAA (blue) positioned within a peptide (green) and displayed on pIII of an M13 virion. After multiple rounds of phage-display affinity selections to the target protein, a consensus amino acid sequence flanking the ligand conjugated ncAA will be revealed. The final product is an optimized sequence with strong affinity for the target protein. **(B)** Macrocycle formation via spontaneous cyclization. The cysteine-reactive ncAA O2bey (O-(2-bromoethyl)-tyrosine, blue) spontaneously cyclizes with a cysteine residue (purple) to form a macrocycle that is displayed on pIII of an M13 virion. **(C)** Dual fluorophore labeling of a phage-displayed protein. Orthogonal incorporation of the ncAAs PrpF (*p*-propargyloxy-phenylalanine) and CypK (cyclopropene derivative of lysine) into the variable light chain (V_L, yellow) and variable heavy chain (V_H, orange) regions of a single-chain variable fragment (scFv) of a recombinant antibody. This allows for dual labeling with azide- and tetrazine-fluorophores, respectively, as depicted by the black and white stars.

cyclized peptides all showed stronger affinity to their targets than their linear counterparts. It will be exciting to see how this approach unfolds over time. Perhaps covalent inhibitors can be discovered by incorporating ncAAs capable of cross-linking with a target as reported elsewhere (Chen et al., 2021).

GCE has been used to site-specifically dual-fluorophore-label proteins displayed on the surface of M13 virions (**Figure 4C**). Suppression of amber and quadruplet codons facilitated the insertion of $N^{\varepsilon-}$[((2-methylcycloprop-2-en1-yl)methoxy)carbonyl]-l-lysine (CypK) and *p*-propargyloxy-l-phenyl-alanine (PrpF) into separate sites of a phage-displayed anti-Her2 scFv (Oller-Salvia and Chin, 2019). The orthogonal *Methanococus janaaschii* (*Mj*) tyrosyl-trRNA synthetase (*Mj*TyrRs)/$tRNA_{CUA}$ pair was used to incorporate PrpF, which was labeled with an azide-fluorophore. The pyrrolysyl-tRNA synthetase (PylRS)/tRNA orthogonal pair was used to incorporate CypK, which was labeled with a tetrazine-fluorophore. By exploring a variety of variables, the authors were able to achieve near wild-type display levels of the scFv for dual labeled, phage-displayed proteins. This system allows for mutually orthogonal and site-specific, dual-labeling of a protein in a one-pot reaction.

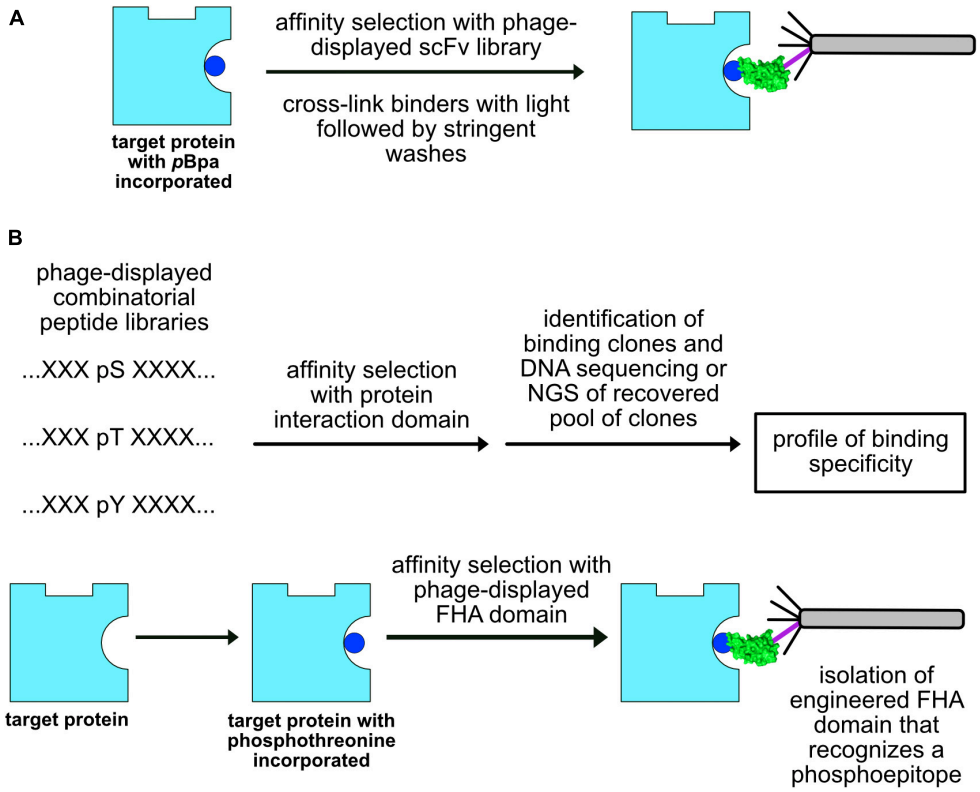


FIGURE 5
Future applications of GCE in phage display. **(A)** Epitope steering. If the target protein (cyan) has incorporated a photoactivatable, crosslinking ncAA (*p*Bpa, blue), recovery of virions can be biased toward those that bind to the labeled epitope. **(B)** Profiling the specificity of domains that recognize peptides that carry post-translational modifications. Phosphoamino acids such as phosphoserine (pS), phosphothreonine (pT), and phosphotyrosine (pY) can be incorporated into phage-displayed combinatorial peptide libraries or recombinant proteins. Affinity selections can be performed with these constructs to elucidate the specificity of protein interaction domains (top) as well as generate affinity reagents that recognize phosphoepitopes (bottom).

Two other recent publications of note describe the incorporation of an ncAA into the target of an affinity selection experiment (**Figure 5A**). Human interleukin-1β (IL-1β) and complement 5a (C5a) proteins were prepared in bacterial cells with *p*-benzoyl-L-phenylalanine utilizing GCE and used to affinity select a phage-display scFv library (Chen et al., 2020). By cross-linking virions to the targets with ultraviolet (UV) irradiation and washing away non-covalently bound virions with pH 2.0 glycine, the bound virions were selectively recovered by trypsin digestion. Over one-third of the recovered scFvs could bind to the targets without cross-linking and their affinities could be improved by mutagenesis. Thus, GCE can be used to steer production of antibodies to epitopes of interest. In a related approach, the 3-nitrotyrosine (nitroTyr) modified form (nY133) of the 14-3-3 signaling protein was used to immunize alpacas and two nanobodies that recognize this post-translational modification at position 133 were recovered by phage-display (Van Fossen et al., 2022).

Perspectives

Several challenges and opportunities remain for optimizing GCE for phage-display. First, in order to incorporate a particular ncAA, the ncAA or its precursor must be supplied to the culture medium for it to cross the bacterial cell membrane where it can cross the cell membrane and be used by an aaRS to charge the suppressor tRNA. This challenge may be solved through the use of mutated strains, the expression of exogenous transporters, and the design of gene clusters capable of cellular synthesis of certain ncAAs. Second the use of bacterial strains that lack release factor 1 (RF-1) (Lajoie et al., 2013; Mukai et al., 2015; Fredens et al., 2019) will suppress TAG codons more efficiently.

One might ask, what are some possible experiments that can be accomplished by combining GCE and phage-display? An example that comes to mind is the production of phage-displayed combinatorial peptide libraries that carry phosphoserine (Park et al., 2011; Zhu et al., 2021),

phosphothreonine (Zhang et al., 2017), or phosphotyrosine (Hoppmann et al., 2017; Luo et al., 2017). Such libraries could then be used to define the specificity of naturally occurring protein interaction modules that recognize phosphoepitopes (Yaffe and Elia, 2001; Liang and Van Doren, 2008; Marasco and Carlomagno, 2020), such as Src Homology 2 (SH2), phosphotyrosine binding domains (PTB), 14-3-3 proteins, BRCA1 C-terminus (BRCT) domains, and Forkhead-associated (FHA) domains (**Figure 5B**). The same libraries could be used to define the specificity of antibodies generated to phosphopeptides (Mandell, 2003). It might even be possible to construct libraries that display more than one post-translational modification (Qin and Liu, 2022) to elucidate the specificity of those domains that recognize two post-translational modifications simultaneously. This general approach should be applicable to defining the specificity of cellular proteins that bind other post-translational modifications. Another application might be to generate recombinant proteins that bear a phosphorylated amino acid at a position found in cellular proteins and then generate recombinant affinity reagents that recognize the folded, phosphorylated target (**Figure 5B**). Thus, it may even be possible to use this general approach to generate recombinant antibodies to many of the site-specific post-translational modifications of interest to cell biologists and biochemists. Finally, it is likely that chemically-and genetically-modified virions will prove beneficial in such emerging applications of phage in drug delivery (Karimi et al., 2016), immune-oncology (Foglizzo and Marchiò, 2021), synthetic biology (Lemire et al., 2018), tissue regeneration (Cao et al., 2019), and vaccines (Henry et al., 2015; González-Mora et al., 2020).

Author contributions

AG and GA prepared the figures. All the authors contributed to the writing or editing the manuscript.

Funding

Funding was provided by grant 1 R43 GM146514-01 from the National Institutes of Health.

Conflict of interest

GA, AG, and BK were employed of a biotech start-up company (Tango Biosciences, Inc.).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adhikari, M., Dhamane, S., Hagström, A. E., Garvey, G., Chen, W. H., Kourentzi, K., et al. (2013). Functionalized viral nanoparticles as ultrasensitive reporters in lateral-flow assays. *Analyst* 138, 5584–5587. doi: 10.1039/c3an00891f

Adhikari, M., Strych, U., Kim, J., Goux, H., Dhamane, S., Poongavanam, M. V., et al. (2015). Aptamer-phage reporters for ultrasensitive lateral flow assays. *Analy. Chem.* 87, 11660–11665. doi: 10.1021/acs.analchem.5b00702

Anderson, J. C., and Schultz, P. G. (2003). Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* 42, 9598–9608. doi: 10.1021/bi034550w

Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S., and Schultz, P. G. (2004). An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U S A* 101, 7566–7571.

Beech, J., Saleh, L., Frentzel, J., Figler, H., Corrêa, I. R., Baker, B., et al. (2015). Multivalent site-specific phage modification enhances the binding affinity of receptor ligands. *Bioconjug. Chem.* 26, 529–536. doi: 10.1021/acs.bioconjchem.5b00011

Bryson, D. I., Fan, C., Guo, L. T., Miller, C., Söll, D., and Liu, D. R. (2017). Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* 13, 1253–1260.

Cao, B., Li, Y., Yang, T., Bao, Q., Yang, M., and Mao, C. (2019). Bacteriophage-based biomaterials for tissue regeneration. *Adv. Drug Del. Rev.* 145, 73–95. doi: 10.1016/j.addr.2018.11.004

Carmody, C. M., Goddard, J. M., and Nugen, S. R. (2021). Bacteriophage Capsid Modification by Genetic and Chemical Methods. *Bioconjug. Chem.* 32, 466–481.

Chatterjee, A., Lajoie, M. J., Xiao, H., Church, G. M., and Schultz, P. G. (2014). A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *Chembiochem* 15, 1782–1786. doi: 10.1002/cbic.201402104

Chen, L., Zhu, C., Guo, H., Li, R., Zhang, L., Xing, Z., et al. (2020). Epitope-directed antibody selection by site-specific photocrosslinking. *Sci. Adv.* 6:eaaz7825. doi: 10.1126/sciadv.aaz7825

Chen, S., Lovell, S., Lee, S., Fellner, M., Mace, P. D., and Bogyo, M. (2021). Identification of highly selective covalent inhibitors by phage display. *Nat. Biotechnol.* 39, 490–498. doi: 10.1038/s41587-020-0733-7

Chung, C. Z., Amikura, K., and Söll, D. (2020). Using genetic code expansion for protein biochemical studies. *Front. Bioeng. Biotechnol.* 8:598577.

Day, J. W., Kim, C. H., Smider, V. V., and Schultz, P. G. (2013). Identification of metal ion binding peptides containing unnatural amino acids by phage display. *Bioorg. Med. Chem. Lett.* 23, 2598–2600. doi: 10.1016/j.bmcl.2013.02.106

DeBenedictis, E. A., Carver, G. D., Chung, C. Z., Söll, D., and Badran, A. H. (2021). Multiplex suppression of four quadruplet codons via tRNA directed evolution. *Nat. Commun.* 12:5706. doi: 10.1038/s41467-021-25948-y

DeBenedictis, E. A., Chory, E. J., Gretton, D. W., Wang, B., Golas, S., and Esvelt, K. M. (2022). Systematic molecular evolution enables robust biomolecule discovery. *Nat. Methods* 19, 55–64. doi: 10.1038/s41592-021-01348-4

Deyle, K., Kong, X. D., and Heinis, C. (2017). Phage selection of cyclic peptides for application in research and drug development. *Acc. Chem. Res.* 50, 1866–1874.

Domaille, D. W., Lee, J. H., and Cha, J. N. (2013). High density DNA loading on the M13 bacteriophage provides access to colorimetric and fluorescent protein microarray biosensors. *Chem. Commun.* 49, 1759–1761. doi: 10.1039/c3cc38871a

Ekanayake, A. I., Sobze, L., Kelich, P., Youk, J., Bennett, N. J., Mukherjee, R., et al. (2021). Genetically encoded fragment-based discovery from phage-displayed macrocyclic libraries with genetically encoded unnatural pharmacophores. *J. Am. Chem. Soc.* 143, 5497–5507. doi: 10.1021/jacs.1c01186

Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature* 472, 499–503.

Fischer, J. T., Söll, D., and Tharp, J. M. (2022). Directed evolution of methanomethylophilus alvus Pyrrolysyl-tRNA synthetase generates a hyperactive and highly selective variant. *Front. Mol. Biosci.* 9:850613. doi: 10.3389/fmolb.2022.850613

Foglizzo, V., and Marchiò, S. (2021). Bacteriophages as Therapeutic and Diagnostic Vehicles in Cancer. *Pharmaceuticals* 14:161.

Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., et al. (2019). Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 569, 514–518.

González-Mora, A., Hernández-Pérez, J., Iqbal, H. M. N., Rito-Palomares, M., and Benavides, J. (2020). Bacteriophage-based vaccines: a potent approach for antigen delivery. *Vaccines* 8:504. doi: 10.3390/vaccines8030504

Guo, J., and Niu, W. (2022). Genetic code expansion through quadruplet codon decoding. *J. Mol. Biol.* 434:167346.

Hagström, A. E., Garvey, G., Paterson, A. S., Dhamane, S., Adhikari, M., Estes, M. K., et al. (2015). Sensitive detection of norovirus using phage nanoparticle reporters in lateral-flow assay. *PLoS One* 10:e0126571. doi: 10.1371/journal.pone.0126571

Heinis, C., Rutherford, T., Freund, S., and Winter, G. (2009). Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* 5, 502–507.

Henry, K. A., Arbabi-Ghahroudi, M., and Scott, J. K. (2015). Beyond phage display: non-traditional applications of the filamentous bacteriophage as a vaccine carrier, therapeutic biologic, and bioconjugation scaffold. *Front. Microbiol.* 6:755. doi: 10.3389/fmicb.2015.00755

Hess, G. T., Cragnolini, J. J., Popp, M. W., Allen, M. A., Dougan, S. K., Spooner, E., et al. (2012). M13 bacteriophage display framework that allows sortase-mediated modification of surface-accessible phage proteins. *Bioconjug. Chem.* 23, 1478–1487. doi: 10.1021/bc300130z

Hoppmann, C., Wong, A., Yang, B., Li, S., Hunter, T., Shokat, K. M., et al. (2017). Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat. Chem. Biol.* 13, 842–844.

Italia, J. S., Addy, P. S., Erickson, S. B., Peeler, J. C., Weerapana, E., and Chatterjee, A. (2019). Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. *J. Am. Chem. Soc.* 141, 6204–6212. doi: 10.1021/jacs.8b12954

Junutula, J. R., Bhakta, S., Raab, H., Ervin, K. E., Eigenbrot, C., Vandlen, R., et al. (2008). Rapid identification of reactive cysteine residues for site-specific labeling of antibody-Fabs. *J. Immunolog. Methods* 332, 41–52. doi: 10.1016/j.jim.2007.12.011

Kang, M., Light, K., Ai, H. W., Shen, W., Kim, C. H., Chen, P. R., et al. (2014). Evolution of iron(II)-finger peptides by using a bipyridyl amino acid. *Chembiochem* 15, 822–825. doi: 10.1002/cbic.201300727

Karimi, M., Mirshekari, H., Moosavi Basri, S. M., Bahrami, S., Moghoofei, M., and Hamblin, M. R. (2016). Bacteriophages and phage-inspired nanocarriers for targeted delivery of therapeutic cargos. *Adv. Drug Del. Rev.* 106, 45–62. doi: 10.1016/j.addr.2016.03.003

Keeble, A. H., and Howarth, M. (2020). Power to the protein: enhancing and combining activities using the Spy toolbox. *Chem. Sci.* 11, 7281–7291. doi: 10.1039/d0sc01878c

Kehoe, J. W., and Kay, B. K. (2005). Filamentous phage display in the new millennium. *Chem. Rev.* 105, 4056–4072. doi: 10.1021/cr000261r

Kim, J., Adhikari, M., Dhamane, S., Hagström, A. E., Kourentzi, K., Strych, U., et al. (2015). Detection of viruses by counting single fluorescent genetically biotinylated reporter immunophage using a lateral flow assay. *ACS Appl. Mater. Interfaces* 7, 2891–2898. doi: 10.1021/am5082556

Kim, J., Vu, B., Kourentzi, K., Willson, R. C., and Conrad, J. C. (2017). Increasing binding efficiency via reporter shape and flux in a viral nanoparticle lateral-flow assay. *ACS Appl. Mater. Interfaces* 9, 6878–6884. doi: 10.1021/acsami.6b15728

Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., et al. (2013). Genomically recoded organisms expand biological functions. *Science* 342, 357–360.

Lee, Y. J., Yi, H., Kim, W. J., Kang, K., Yun, D. S., Strano, M. S., et al. (2009). Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. *Science* 324, 1051–1055. doi: 10.1126/science.1171541

Lemire, S., Yehl, K. M., and Lu, T. K. (2018). Phage-based applications in synthetic biology. *Annu. Rev. Virol.* 5, 453–476.

Li, K., Chen, Y., Li, S., Nguyen, H. G., Niu, Z., You, S., et al. (2010). Chemical modification of M13 bacteriophage and its application in cancer cell imaging. *Bioconjug. Chem.* 21, 1369–1377. doi: 10.1021/bc900405q

Liang, X., and Van Doren, S. R. (2008). Mechanistic insights into phosphoprotein-binding FHA domains. *Acc. Chem. Res.* 41, 991–999. doi: 10.1021/ar700148u

Liu, C. C., Mack, A. V., Brustad, E. M., Mills, J. H., Groff, D., Smider, V. V., et al. (2009). Evolution of proteins with genetically encoded "chemical warheads". *J. Am. Chem. Soc.* 131, 9616–9617. doi: 10.1021/ja902985e

Liu, C. C., Mack, A. V., Tsao, M. L., Mills, J. H., Lee, H. S., Choe, H., et al. (2008). Protein evolution with an expanded genetic code. *Proc. Natl. Acad. Sci. U S A* 105, 17688–17693.

Liu, D. R., Magliery, T. J., Pastrnak, M., and Schultz, P. G. (1997). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U S A* 94, 10092–10097.

Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., et al. (2017). Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* 13, 845–849. doi: 10.1038/nchembio.2405

Magliery, T. J., Anderson, J. C., and Schultz, P. G. (2001). Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*. *J. Mol. Biol.* 307, 755–769. doi: 10.1006/jmbi.2001.4518

Manandhar, M., Chun, E., and Romesberg, F. E. (2021). Genetic code expansion: inception, development, commercialization. *J. Am. Chem. Soc.* 143, 4859–4878. doi: 10.1021/jacs.0c11938

Mandell, J. W. (2003). Phosphorylation state-specific antibodies: applications in investigative and diagnostic pathology. *Am. J. Pathol.* 163, 1687–1698.

Marasco, M., and Carlomagno, T. (2020). Specificity and regulation of phosphotyrosine signaling through SH2 domains. *J. Struct. Biol. X* 4:100026.

Miller, S. M., Wang, T., and Liu, D. R. (2020). Phage-assisted continuous and non-continuous evolution. *Nat. Protoc.* 15, 4101–4127.

Mohan, K., and Weiss, G. A. (2016). Chemically modifying viruses for diverse applications. *ACS Chem. Biol.* 11, 1167–1179.

Moon, J. S., Choi, E. J., Jeong, N. N., Sohn, J. R., Han, D. W., and Oh, J. W. (2019). Research Progress of M13 Bacteriophage-Based Biosensors. *Nanomaterials* 9:E1448. doi: 10.3390/nano9101448

Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., et al. (2015). Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci. Rep.* 5:9699. doi: 10.1038/srep09699

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 464, 441–444. doi: 10.1038/nature08817

Ng, S., and Derda, R. (2016). Phage-displayed macrocyclic glycopeptide libraries. *Org. Biomol. Chem.* 14, 5539–5545. doi: 10.1039/c5ob02646f

Nikić-Spiegel, I. (2020). Expanding the genetic code for neuronal studies. *Chembiochem* 21, 3169–3179.

Niu, W., Schultz, P. G., and Guo, J. (2013). An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* 8, 1640–1645.

Nuijens, T., Toplak, A., Schmidt, M., Ricci, A., and Cabri, W. (2019). Natural occurring and engineered enzymes for peptide ligation and cyclization. *Front. Chem.* 7:829.

Oller-Salvia, B., and Chin, J. W. (2019). Efficient phage display with multiple distinct non-canonical amino acids using orthogonal ribosome-mediated genetic code expansion. *Angew. Chem. Int. Ed. Engl.* 58, 10844–10848. doi: 10.1002/anie.201902658

Owens, A. E., Iannuzzelli, J. A., Gu, Y., and Fasan, R. (2020). MOrPH-PhD: an integrated phage display platform for the discovery of functional genetically encoded peptide macrocycles. *ACS Cent. Sci.* 6, 368–381.

Park, H. S., Hohn, M. J., Umehara, T., Guo, L. T., Osborne, E. M., Benner, J., et al. (2011). Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* 333, 1151–1154.

Pastrnak, M., and Schultz, P. G. (2001). Phage selection for site-specific incorporation of unnatural amino acids into proteins in vivo. *Bioorg. Med. Chem.* 9, 2373–2379. doi: 10.1016/s0968-0896(01)00157-2

Petrenko, V. A. (2018). Landscape phage: evolution from phage display to nanobiotechnology. *Viruses* 10:E311.

Qin, X., and Liu, T. (2022). Recent advances in genetic code expansion techniques for protein phosphorylation studies. *J. Mol. Biol.* 434:167406.

Ros, E., Torres, A. G., Ribas, and de Pouplana, L. (2021). Learning from nature to expand the genetic code. *Trends Biotechnol.* 39, 460–473.

Sanders, J., Hoffmann, S. A., Green, A. P., and Cai, Y. (2022). New opportunities for genetic code expansion in synthetic yeast. *Curr. Opin. Biotechnol.* 75:102691. doi: 10.1016/j.copbio.2022.102691

Sandman, K. E., and Noren, C. J. (2000). The efficiency of *Escherichia coli* selenocysteine insertion is influenced by the immediate downstream nucleotide. *Nucleic Acids Res.* 28, 755–761. doi: 10.1093/nar/28.3.755

Scholle, M. D., Collart, F. R., and Kay, B. K. (2004). In vivo biotinylated proteins as targets for phage-display selection experiments. *Protein Expr. Purif.* 37, 243–252.

Shandell, M. A., Tan, Z., and Cornish, V. W. (2021). Genetic code expansion: a brief history and perspective. *Biochemistry* 60, 3455–3469. doi: 10.1021/acs.biochem.1c00286

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317. doi: 10.1126/science.4001944

Smith, G. P. (2019). Phage display: simple evolution in a petri dish (nobel?lecture). *Angew. Chem. Int. Ed. Engl.* 58, 14428–14437. doi: 10.1002/anie.201908308

Tharp, J. M., Ehnbom, A., and Liu, W. R. (2018). $tRNA^{Pyl}$: Structure, function, and applications. *RNA Biol.* 15, 441–452. doi: 10.1080/15476286.2017.1356561

Tharp, J. M., Hampton, J. T., Reed, C. A., Ehnbom, A., Chen, P. C., Morse, J. S., et al. (2020). An amber obligate active site-directed ligand evolution technique for phage display. *Nat. Commun.* 11:1392. doi: 10.1038/s41467-020-15057-7

Tian, F., Tsao, M. L., and Schultz, P. G. (2004). A phage display system with unnatural amino acids. *J. Am. Chem. Soc.* 126, 15962–15963.

Ulfo, L., Cantelli, A., Petrosino, A., Costantini, P. E., Nigro, M., Starinieri, F., et al. (2022). Orthogonal nanoarchitectonics of M13 phage for receptor targeted anticancer photodynamic therapy. *Nanoscale* 14, 632–641. doi: 10.1039/d1nr06053h

Urquhart, T., Daub, E., and Honek, J. F. (2016). Bioorthogonal modification of the major sheath protein of bacteriophage M13: extending the versatility of bionanomaterial scaffolds. *Bioconjug. Chem.* 27, 2276–2280. doi: 10.1021/acs.bioconjchem.6b00460

Van Fossen, E. M., Grutzius, S., Ruby, C. E., Mourich, D. V., Cebra, C., Bracha, S., et al. (2022). Creating a Selective Nanobody Against 3-Nitrotyrosine Containing Proteins. *Front. Chem.* 10:835229. doi: 10.3389/fchem.2022.835229

Vignali, V., Miranda, S., Lodoso-Torrecilla, I., van Nisselroy, C. A. J., Hoogenberg, B. J., Dantuma, S., et al. (2018). Biocatalytically induced surface modification of the tobacco mosaic virus and the bacteriophage M13. *Chem. Commun.* 55, 51–54. doi: 10.1039/c8cc08042a

Wan, W., Tharp, J. M., and Liu, W. R. (2014). Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* 1844, 1059–1070. doi: 10.1016/j.bbapap.2014.03.002

Wang, K., Sachdeva, A., Cox, D. J., Wilf, N. M., Lang, K., Wallace, S., et al. (2014). Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* 6, 393–403. doi: 10.1038/nchem.1919

Wang, X. S., Chen, P. C., Hampton, J. T., Tharp, J. M., Reed, C. A., Das, S. K., et al. (2019). A genetically encoded, phage-displayed cyclic-peptide library. *Angew. Chem. Int. Ed. Engl.* 58, 15904–15909.

Weeks, A. M., and Wells, J. A. (2020). Subtiligase-catalyzed peptide ligation. *Chem. Rev.* 120, 3127–3160.

Xiao, H., and Schultz, P. G. (2016). At the interface of chemical and biological synthesis: an expanded genetic code. *Cold Spring Harb. Perspect. Biol.* 8:a023945.

Yaffe, M. B., and Elia, A. E. (2001). Phosphoserine/threonine-binding domains. *Curr. Opin. Cell Biol.* 13, 131–138.

Young, D. D., and Schultz, P. G. (2018). Playing with the molecules of Life. *ACS Chem. Biol.* 13, 854–870.

Zhang, M. S., Brunner, S. F., Huguenin-Dezot, N., Liang, A. D., Schmied, W. H., Rogerson, D. T. et al. (2017). Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* 14, 729–736. doi: 10.1038/nmeth.4302

Zhu, P., Franklin, R., Vogel, A., Stanisheuski, S., Reardon, P., Sluchanko, N. N., et al. (2021). PermaPhos ^{Ser} : autonomous synthesis of functional, permanently phosphorylated proteins. *bioRxiv* 2021:465468. doi: 10.1101/2021.10.22.465468

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 22 August 2022
DOI 10.3389/fmicb.2022.968036

OPEN ACCESS

EDITED BY
Jian Huang,
University of Electronic Science and Technology of China, China

REVIEWED BY
Jinhai Huang,
Tianjin University,
China
Huaichao Luo,
Sichuan Cancer Hospital,
China

*CORRESPONDENCE
Chenyi Luo
469453215@qq.com
Jianxun Wang
jianxun.wang@bucm.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

RECEIVED 13 June 2022
ACCEPTED 18 July 2022
PUBLISHED 22 August 2022

CITATION
Chen H, Li S, Wang J, He S, Wang D, Qian Z, Hu D, Qi F, Hu K, Luo C and Wang J (2022) Simultaneous measurement of the antibody responses against SARS-CoV-2 and its multiple variants by a phage display mediated immuno-multiplex quantitative PCR-based assay. *Front. Microbiol.* 13:968036. doi: 10.3389/fmicb.2022.968036

Simultaneous measurement of the antibody responses against SARS-CoV-2 and its multiple variants by a phage display mediated immuno-multiplex quantitative PCR-based assay

Hanyi Chen[1†], Shen Li[1†], Jiali Wang[1], Siqi He[1], Dong Wang[2], Zhaohui Qian[3], Dandan Hu[4], Fangfang Qi[5], Keping Hu[6,7], Chenyi Luo[1,8*] and Jianxun Wang[1,8*]

[1]School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China, [2]School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [3]NHC Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [4]Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China, [5]Department of Anatomy and Neurobiology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China, [6]The Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [7]Andes Antibody Technology Hengshui LL Company, Hengshui City, China, [8]Shenzhen Research Institute, Beijing University of Chinese Medicine, Shenzhen, China

To combat the continued pandemic of COVID-19, multiplex serological assays have been developed to comprehensively monitor the humoral immune response and help to design new vaccination protocols to different SARS-CoV-2 variants. However, multiplex beads and stably transfected cell lines require stringent production and storage conditions, and assays based on flow cytometry is time-consuming and its application is therefore restricted. Here, we describe a phage display system to distinguish the differences of immune response to antigenic domains of multiple SARS-CoV-2 variants simultaneously. Compared with linear peptides, the recombinant antigens displayed on the phage surface have shown some function that requires the correct folding to form a stable structure, and the binding efficiency between the recombinant phage and existing antibodies is reduced by mutations on antigens known to be important for antigen–antibody interaction. By using Phage display mediated immuno-multiplex quantitative PCR (Pi-mqPCR), the binding efficiency between the antibody and antigens of different SARS-CoV-2 variants can be measured in one amplification reaction. Overall, these data show that this assay is a valuable tool to evaluate the humoral response to the same antigen of different SARS-CoV-2 variants or antigens of different pathogens. Combined with high-throughput DNA sequencing technology, this phage display system can be further applied in monitoring humoral immune response in a large population before and after vaccination.

KEYWORDS

phage display, coronavirus, SARS-CoV-2, variants, antibody responses

Introduction

Since coronavirus disease-2019 (COVID-19) emerged at the end of 2019, the continued severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic has caused a worldwide public health crisis. A member of the β-coronavirus genus, SARS-CoV-2, has a 29,903 bp genome that codes for structural proteins, including envelope (E), spike (S), nucleocapsid (N), and membrane (M), and nonstructural proteins required for virus infection (Khailany et al., 2020). In the process of virus infection, the receptor-binding domain (RBD) of spike can bind angiotensin-converting enzyme 2 (ACE2) and mediate SARS-CoV-2 entry into the cell (Lan et al., 2020). Because of its importance in terms of virus tropism and infectivity, the S protein has become the target of most vaccines and antibody drugs (Barnes et al., 2020). However, as a single-stranded positive-strand RNA virus with a high mutation rate, mutations have accumulated during the SARS-CoV-2 pandemic, and the variants with increased fitness and potential to escape the immune response have developed, increasing the chance of spread (Harvey et al., 2020).

Since it was first reported in January 2020, the variant with the D614G mutation has replaced the wild type and become the mainstream variant worldwide. Many studies have shown that D614G can significantly enhance the infectivity of SARS-CoV-2, but this mutation does not affect the neutralizing effect of monoclonal antibodies (Yurkovetskiy et al., 2020). At the end of 2020, the variant B.1.351 (Beta) was first reported in South Africa and repeatedly connected with immune escape. B.1.351 has three mutations in the RBD region, K417N, E484K, and N501Y. Some research has shown that E484K and N501Y change the spatial structure of the RBD and reduce the binding efficiency of existing antibodies (Weisblum et al., 2020). B.1.617.2 (Delta) was first discovered in India and became a mainstream variant in 2021. The mutations L452R and T478K in the RBD region may help the virus resist neutralization (Wall et al., 2021). In addition, the P681R mutation improves the replication efficiency of the virus in the human airway system (Saito et al., 2022). In November 2021, variant B.1.1.529 was named Omicron, which overtook the Delta strain in a short time and currently dominates globally. The Omicron variant has more than 50 mutations, 30 of which are located in spike. Because of its enhanced transmissibility and immune evasion capability, Omicron has caused unprecedented concerns worldwide (Callaway, 2021). In the context of increasing vaccination rates and the emergence of variants, evaluating the humoral immune status in a human population toward the different variants and adjusting the countermeasures will play an essential role in counteracting the spread of the virus.

Serological assays have been performed on linear peptides or full-length antigen proteins, and their binding activity has been observed with antibodies. Since the COVID-19 pandemic started, many serological assays based on the spike, nucleocapsid, and other proteins of SARS-CoV-2 have been developed (Torres et al., 2019; Candel et al., 2020; Zhao et al., 2020). These assays employ different techniques, such as ELISA, lateral flow immunoassay (LFIA), and chemiluminescence enzyme immunoassay (CLIA). However, most of these techniques detect only the antibody level to a certain kind of protein in a test. Based on fluorescence immunoassays, Luminex can assay the presence or absence of antibodies to three different SARS-CoV-2 antigens, such as S1, the RBD, and nucleocapsid (xMAP®, 2022). In 2021, Niklas et al. transfected wild-type (WT) or mutant S proteins into the Ramos human B lymphoma cell line and used a color-based barcoded spike flow cytometric assay (BSFA), which allows comparison of the level of antibodies to the S protein of WT SARS-CoV-2 and variants (Vesper et al., 2021). However, multiplex beads and stably transfected cell lines which used in fluorescence immunoassays require stringent production and storage conditions, and assays based on flow cytometry will still be restricted by throughput. Peptide microarrays can immobilize short peptides on solid planar supports and detect different pathogen-related peptides or epitopes in high throughput (Wang et al., 2020; Vengesai et al., 2022). Based on recombinant antigens, protein microarray can be also used to assay the antibody response to different antigens of SARS-COV-2. But considering the high cost of a commercial protein microarray, this method will still be limited in large population assay (Bostan et al., 2020; Krishnamurthy et al., 2020; Hein et al., 2022). Phage immunoprecipitation sequencing (PhIP-Seq) was first reported in 2011, and phage-displayed antigen libraries are encoded by synthetic oligonucleotides. After immunoprecipitation with serum samples, deep DNA sequencing can permit the quantification of each peptide's antibody-dependent enrichment (Larman et al., 2011). In this way, humoral immune assays can utilize to DNA sequencing, significantly improving the throughput of the assay (Table 1). However, because of the limited length of the synthetic oligonucleotide library, this phage display method can detect only linear epitope-directed antibodies (Mohan et al., 2018). Stoddard et al. (2021) captured immunogenic peptides spanning the entire proteome of SARS-CoV-2 in a phage-displayed antigen library. Through the humoral immune assays of 19 COVID-19 patients, S, N, and ORF1ab were identified as highly immunogenic regions. However, due to the conformational tendency, this study did not identify any antibody-dependent enrichment in the RBD.

Here, we describe a polyvalent phage display system based on the fusion of the entire RBD of SARS-CoV-2 with the M13 bacteriophage Protein III and display it on the phage surface. In

TABLE 1 The throughput of common serological assays to SARS-CoV-2.

Test type	Throughput
ELISA	Single detection
Lateral flow immunoassay (LFIA)	Single detection
Chemiluminescence enzyme immunoassay (CLIA)	Single detection
Fluorescence immunoassays (FIA)	Multiple assays
Peptide microarrays	High-throughput assay
PhIP-Seq	High-throughput assay

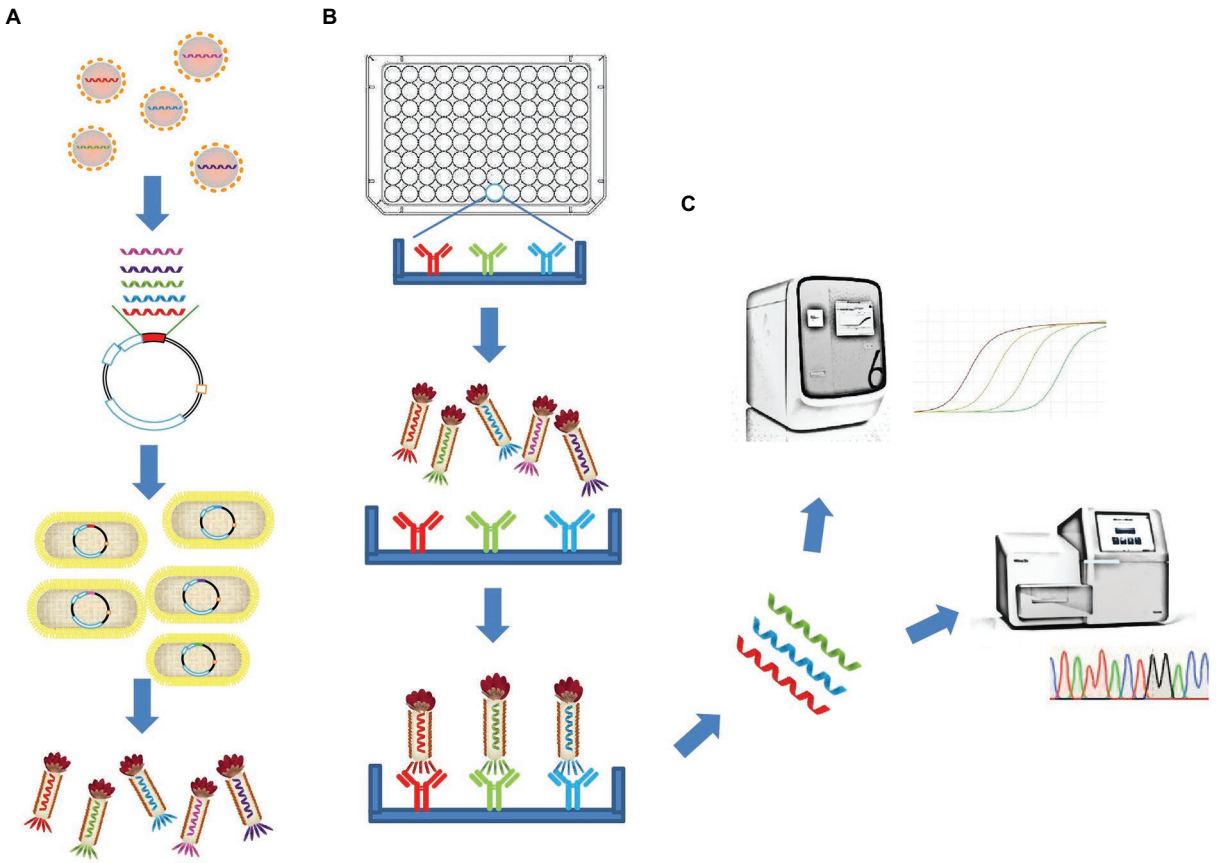


FIGURE 1
Schematic illustration of the polyvalent phage display system and the Pi-mqPCR assay. **(A)** Construction and synthesis of the recombinant phage. **(B)** Phage immunoprecipitation. **(C)** Real-time immuno-PCR and high-throughput DNA sequencing assay.

addition to detecting linear epitope-directed antibodies, the recombinant antigens displayed on the phage surface has shown some function that requires the correct folding to form a stable structure compared with linear peptides, including reduced binding efficiency of existing antibodies by mutations and binding with receptor ACE2. By using Phage display mediated immuno-multiplex quantitative PCR (Pi-mqPCR), the binding efficiency between the antibody and different SARS-CoV-2 variants was compared in the same amplified reaction. In this way, antigen–antibody reaction can be turned into DNA assays and significantly improve the throughput of the assay (Figure 1).

Materials and methods

Plasmids and bacterial strains

The plasmid pUC19 containing the codon-optimized genes for the SARS-CoV-2 spike RBD (WT), spike NTD, nucleocapsid protein, and hemagglutinin HA1 subunit was commercially synthesized (GENERAL BLOT). The gene for the SARS-CoV-2 spike RBD (Omicron) was a gift from Dr. Zhaohui Qian. The protein sequences used for our phage display system were from the SARS-CoV-2 Wuhan strain (MN908947), SARS-CoV-2 Omicron strain (R40B60 BHP 3321001247/2021), influenza virus A/Perth/16/2009(H3N2; KM821346), and A/WSN/1933(H1N1; HE802059). M13KO7 helper phage was provided by New England Biolabs. Competent *Escherichia coli* TG1 and DH5α cells were obtained from Biomed.

Construction of the recombinant phage

The M13KO7 helper phage was used to infect a culture of TG1 cells. After growth overnight at 37°C, the TIANprep Mini Plasmid Kit (TIANGEN) was used to extract the double-stranded DNA (dsDNA) phage chromosome. The insert genes and the M13KO7 phage chromosome were amplified by PCR with primers containing 20 base pair overlap. The PCR products were assembled by an EasyGeno Assembly Cloning Kit (TIANGEN) at 50°C for 30 min. The assembly mix was transformed into competent DH5α cells. The point mutations in the RBD region were generated by QuikChange Lightning (Agilent). The recombinant phage chromosome was sequenced (GENEWIZ), and the sequencing

results were analyzed by SnapGene 4.2 software. To produce the recombinant phage, the DH5α strain transformed with the recombinant phage chromosome was grown overnight at 37°C and centrifuged for 15 min at 8,000 × rpm at 4°C. The supernatant was collected, and the recombinant phage particles were precipitated by adding 20% (w/v) polyethylene glycol (PEG) 8,000 solution to 2.5 M NaCl (LABLEAD) in distilled water. Following incubation on ice for 1 h, the mixture was centrifuged for 30 min at 10,000 × rpm at 4°C, and the pellet was resuspended in 1 ml of PBS and stored at 4°C. After incubation at 95°C for 15 min, real-time fluorescent quantitative PCR was used to detect the titer of the phage.

Western blot assay

Before denaturation at 99°C for 20 min, 10 μl of 5× SDS–PAGE Sample Loading Buffer (LABLEAD) was added to 40 μl of phage (109 copies/μl). Then, the samples were electrophoresed on LabPAGE 4–12% 11-well gels (LABLEAD) and transferred to a nitrocellulose membrane (APPLYGEN). The nitrocellulose membrane was blocked with 5% skim milk for 2 h, after which the anti-myc tag antibody (R&D Systems) was added and incubated overnight at 4°C. The membrane was washed three times with TBST (LABLEAD) and then incubated with a horseradish peroxidase-labeled goat anti-mouse IgG secondary antibody (1:5000, APPLYGEN) at room temperature for 1 h. After washing the membrane three times with TBST, SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher) was added, and the ChemiDoc MP Imaging System (Bio-Rad) was used to record images.

Phage immunoprecipitation

The binding protein was coated in a 96-well microplate with 50 μl of ELISA Coating Buffer (Solarbio) and incubated at 4°C overnight. After washing five times with PBST (LABLEAD), blocking was performed with a 2% BSA (LABLEAD) solution in PBST for 2 h at 37°C. Following further washing with PBST four times, 50 μl/well recombinant phage (10^{11} copies/ml) was added to the microplate for 1 h at 37°C. After washing five times with PBST, 100 μl of DNase/RNase-free water (Solarbio) was added to each well of the microplate and incubated at 95°C for 20 min. The eluted phage particles were used as DNA templates in real-time quantitative PCR.

Real-time immuno-PCR assay

For the real-time fluorescent quantitative PCR, 10 μl of 2× Realab Green PCR Fast Mixture [LABLEAD, forward primer (0.5 μM), reverse primer (0.5 μM); Table 2] and 1 μl of eluted phage templates were added to each PCR tube. DNase/RNase-free water was added to bring the up volume to 20 μl. The thermal cycle conditions included 95°C for 30 s, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. The amplification system of multiplex real-time fluorescent quantitative PCR consisted of 5 μl of 2× FastFire qPCR PreMix (TIANGEN), forward primer (0.3 μM), reverse primer (0.3 μM), four types of probes (0.4 μM each), 0.5 μl of eluted phages, and sterilized water to a final volume of 10 μl. After 1 min of incubation at 95°C, 40 PCR cycles were performed according to the following temperature regime: 95°C for 5 s, 57°C for 10 s, and 72°C for 20 s. The amplification was performed using a Stratagene Mx3005P real-time PCR system (Agilent). The copies of phage chromosomes were determined by plotting Ct versus 10-fold serial dilutions of target gene fragments with known DNA concentrations.

TABLE 2 Primer and probe sequences used in multiplex real-time fluorescent quantitative PCR.

Primer/probe	Sequence (5′–3′)
Forward primer	ACTGCGGCGAGCGGAAAT
Reverse primer	GCCACCACTGATTTGAGCG
Probe-1	FAM-AACAACTGGACCGACCG-BHQ1
Probe-2	CY5-CTGGCTCTGCGTGCTGTGCTC-BHQ2
Probe-3	ROX-CTCAAACCCCCGCGCGTTCCCC-MGB
Probe-4	VIC-TCCAAGCGCTCGCATCGTGG-BHQ1

Statistical analysis

In this study, statistical analysis was performed by GraphPad Prism 8.0 (GraphPad Software) by using a *t*-test, and $p < 0.05$ was considered significant. Experimental data were expressed as the mean ± standard error of the mean.

Results

Construction and synthesis of the recombinant phage

To construct the phage displaying polyvalence, the RBD-coding gene was inserted into the genome of M13 bacteriophage between the signal sequence and Gene III. A myc tag was used to detect the expression of the recombinant protein (Figure 2D). Western blot results indicated that recombinant protein III migrated at the expected molecular weight of approximately 65 kDa (Ledsgaard et al., 2018; Tai et al., 2020), while Protein III of wild-type M13KO7 could not bind with the anti-myc-tag antibody (Figure 2A).

The binding activity of recombinant phage was evaluated by real-time fluorescent quantitative PCR. After incubation at 95°C for 20 min, both the microplates coated with anti-RBD antibody (Sino Biological, Cat. # 40150-R007, RRID Number: AB_2827979) and anti-myc-tag antibody (APPLYGEN, Cat. # C1302-100) showed significant enrichment of phage DNA

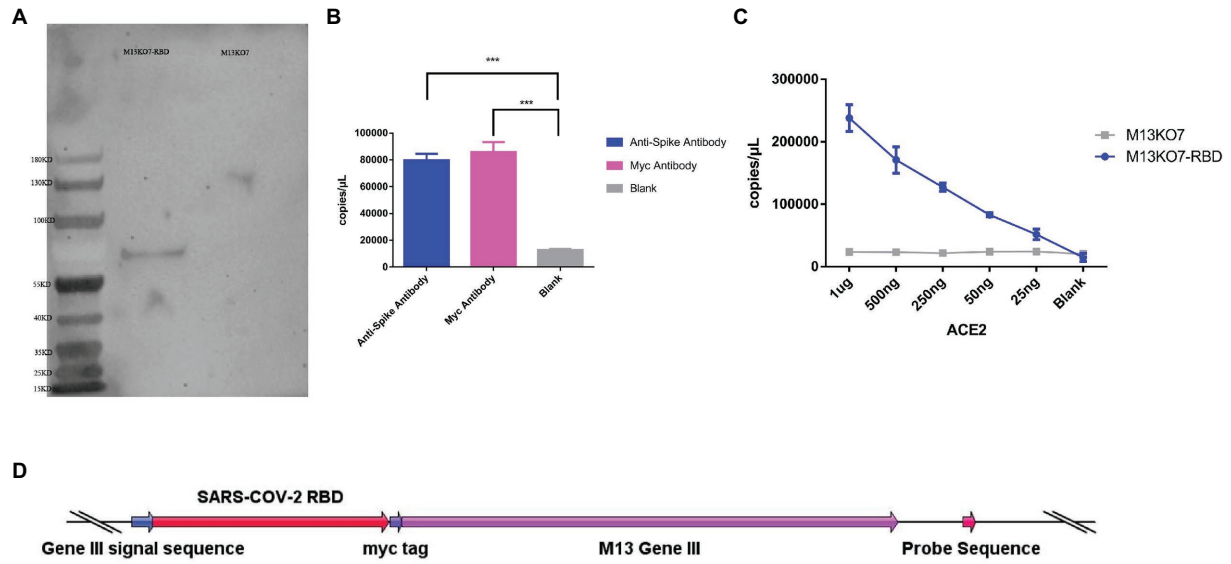


FIGURE 2
Construction and synthesis of the recombinant phage. **(A)** Detection of the expression of the myc tag on wild-type M13KO7 and the recombinant phage by Western blotting. **(B)** The enrichment of the recombinant phage among the antibodies targeted to the spike protein of SARS-CoV-2 and myc tag; ***$p<0.001$. **(C)** The enrichment of the RBD-displaying phage and the wild-type M13KO7 by ACE2. **(D)** The phage vectors used to express the polyvalent display phage. The values shown are the average of three independent experiments and their standard errors.

(Figure 2B). The result shows that the recombinant protein has displayed on the surface of the phage and could be identified by the antibodies.

The function of the recombinant RBD constructs

To test the function of the recombinant RBD constructs, we coated the recombinant human ACE2 protein in the microplate and assayed the enrichment of recombinant phage and WT M13KO7 after immunoprecipitation. In contrast with M13KO7, quantitative PCR showed that the recombinant phage can bind with human ACE2 protein in a concentration-dependent manner (Figure 2C). Next, we wanted to learn more about how the mutations in the phage-displayed RBD construct influenced the recognition of anti-RBD antibodies. We introduced the RBD-displayed phage with L452R, T478K mutations (B.1.617.2, Delta variant), L452Q, F490S mutations (C37, Lambda variant), L452R, E484Q mutations (B.1.617.1, Kappa variant), and N501Y, E484K mutations, which are characteristic of the Alpha and Beta variants (Figure 3B). After immunoprecipitation with two commercially available anti-RBD antibodies (R007 and R118), quantitative PCR showed that the enrichment of recombinant phage could be reduced by mutations in the RBD construct, especially the L452R, T478K, N501Y, and E484K mutations (Figures 3A,B). These data demonstrate that the recombinant RBD displayed on the surface of the phage has more function which requires the correct folding to form a stable structure than linear peptides (Chen et al., 2020).

Phage display mediated immuno-multiplex quantitative PCR assay

To directly assay the enrichment of different recombinant phage, a sequence that could be identified by probes was inserted into the phage chromosome (Figure 2D). Based on this system, Pi-mqPCR can compare the binding activity of recombinant phages displaying different antigens through the same amplification reaction. To identify the binding specificity between the recombinant phage and different antibodies, we first constructed a recombinant phage with the N-terminal domain of the S protein and a C-terminal truncated version of the nucleocapsid protein from SARS-CoV-2. In addition, the hemagglutinin HA1 subunit from influenza virus A/Perth/16/2009(H3N2) and A/WSN/1933(H1N1) was also chosen to construct the recombinant phage. After coating the microplate with the corresponding antibodies, different recombinant phages were then equally mixed for immunoprecipitation (Figure 4A). The result of Pi-mqPCR shows that all the types of recombinant phage showed enrichment in only the microplate coated with the respective labeled antibodies. Based on this result, we found that this system can be used to identify different antibodies not only targeting different viruses (Figure 4B) but also targeting different regions of SARS-CoV-2 (Figure 4C).

Having demonstrated that Pi-mqPCR system works well in the evaluation of binding specificity, we next used it to identify the phage-displayed RBD constructs from different variants. We observed binding activity between the phage displaying the RBD of the Delta variant (B.1.617.2) and Omicron variant

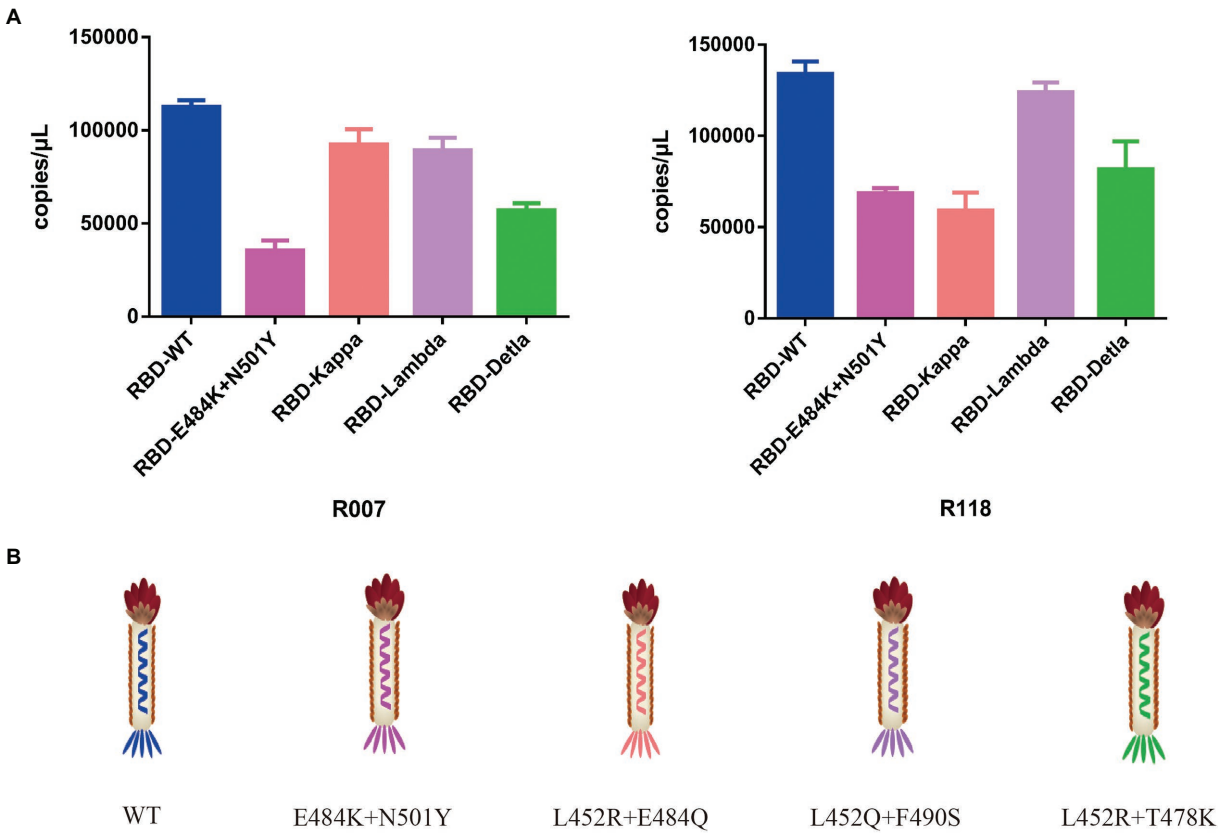


FIGURE 3
(A) The binding activity between the recombinant phage with different mutations in the RBD construct and two commercially available anti-RBD antibodies. **(B)** Schematic drawing of the phage of the phage-displayed RBD construct with different mutations. The values shown are the average of three independent experiments and their standard errors.

(B.1.1.529) and that with N501Y and E484K mutations and four commercially available anti-RBD antibodies (Sino Biological), including a polyclonal antibody (T62: Cat. # 40591-T62) and three monoclonal antibodies (R007: Cat. # 40150-R007, RRID Number: AB_2827979, R118: Cat. # 40592-R118, and MM48: Cat. # 40591-MM48). In addition, six anti-SARS-CoV-2 RBD nanobodies selected in previous research (N1-N6) were also used for immunoprecipitation with the recombinant phage. We observed that all four types of recombinant phage could bind the polyclonal antibody in a concentration-dependent manner. However, compared with the wild type, the recombinant phage that displayed RBD mutants showed more reduced enrichment (Figure 5A). Interestingly, the titration of the three RBD-specific monoclonal antibodies showed a significant difference in the enrichment between the RBD constructs from wild-type and different SARS-CoV-2 variants. The RBD constructs from the Delta variant could still be recognized by the antibodies R007 and R118 in a concentration-dependent manner. However, the binding activity of the phage-displayed RBD with N501Y and E484K point mutations could be observed only at high concentrations of R007 and MM48, while the RBD region from Omicron showed little enrichment with all three monoclonal antibodies (Figures 5B–D).

Similar to the monoclonal antibodies, the six types of anti-SARS-CoV-2 RBD nanobodies showed a greater decline in binding activity with RBD constructs from different variants, especially the Omicron variant (Figure 6). In addition, we compared the binding activity between nanobody N4, N6, and SARS-CoV-2 (2019-nCoV) Spike RBD Recombinant Protein from wild type (Sino Biological, Cat: 40592-V05H) and Omicron variant (Sino Biological, Cat: 40592-V05H3). In line with Pi-mqPCR, the result of the standard ELISA-based method showed that the binding activity between nanobody N4, N6, and RBD Recombinant Protein from Omicron variant reduced significantly compared with the wild type (Supplementary Figure S1). These results are in line with research showing that mutations in the RBD region reduce the binding efficiency of existing antibodies by changing the spatial structure of the RBD (Garcia-Beltran et al., 2021; Li et al., 2021; Liu C, et al., 2021; Liu Z, et al., 2021; Mlcochova et al., 2021), and the Omicron variant shows the most resistance to neutralization by monoclonal and convalescent plasma antibodies (Zhang et al., 2022). Based on these data, the phage-displayed antigen system can be used to evaluate the fine specificity of the antibody response to different SARS-CoV-2 variants.

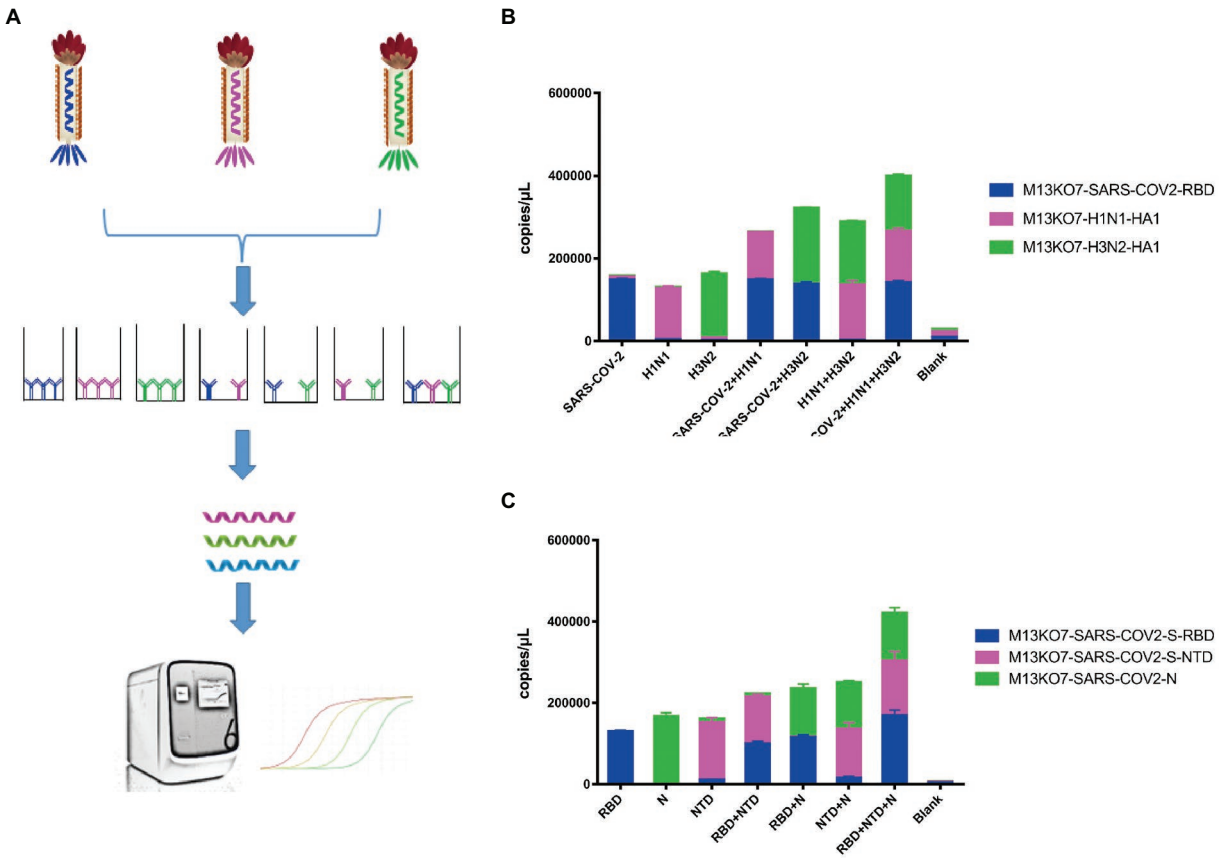


FIGURE 4
Pi-mqPCR was used to identify the binding specificity of recombinant phage. **(A)** Schematic drawing of the Pi-mqPCR. **(B,C)** Distribution of the phage chromosome that displayed antigens from different viruses **(B)** and different regions of SARS-CoV-2 **(C)** following immunoprecipitation with the respective labeled antibodies. The values shown are the average of three independent experiments and their standard errors.

Discussion

In this study, we constructed a polyvalent antigen display system based on the M13 bacteriophage. We show here that the recombinant antigen could be displayed on the surface of the phage and identified by antibodies. Interestingly, we observed binding activity between ACE2 and the phage that displayed the RBD region of SARS-CoV-2. These data demonstrated that the phage-displayed antigen has properties similar to the protein found on the viral membrane. Based on this result, the phage-displayed antigen can be used not only to detect linear-directed antibodies but also in some research that requires the correct structure. For example, Uppalapati et al. (2022) reconstituted the receptor-binding motif (RBM) of Middle East respiratory syndrome coronavirus (MERS-CoV) by phage-display conformer libraries and selected two reconstituted RBM conformers that cross-reacted with a panel of 7 neutralizing monoclonal antibodies. Combined with phage display selection, phage display antigen libraries can also be used to predict RBD constructs that have a higher affinity for ACE2.

To isolate biomolecules with high affinity, the dual plasmid helper phage display system has been used for phage display selection (Kumar et al., 2022); compared with this, the polyvalent display system can reduce nonspecific binding and has more advantages in diagnostic assays. In this research, our system could identify antibodies that target more than three antigens in the same multiplex real-time fluorescent quantitative PCR. Several studies have shown that serological assays based on multiple antigens can indeed increase the specificity of testing and provide a more comprehensive picture of the humoral immune response (Gillot et al., 2020). For example, the research of Grossberg et al. (Grossberg et al., 2021) showed that the antibody levels of S1-RBD IgA, NP IgG, and S2 IgA can be used to identify severe, mild, and asymptomatic groups of COVID-19 patients. Considering increasing vaccination rates, combined detection of anti-NP and anti-Spike antibodies can also be used to differentiate the immune response from viral infection and

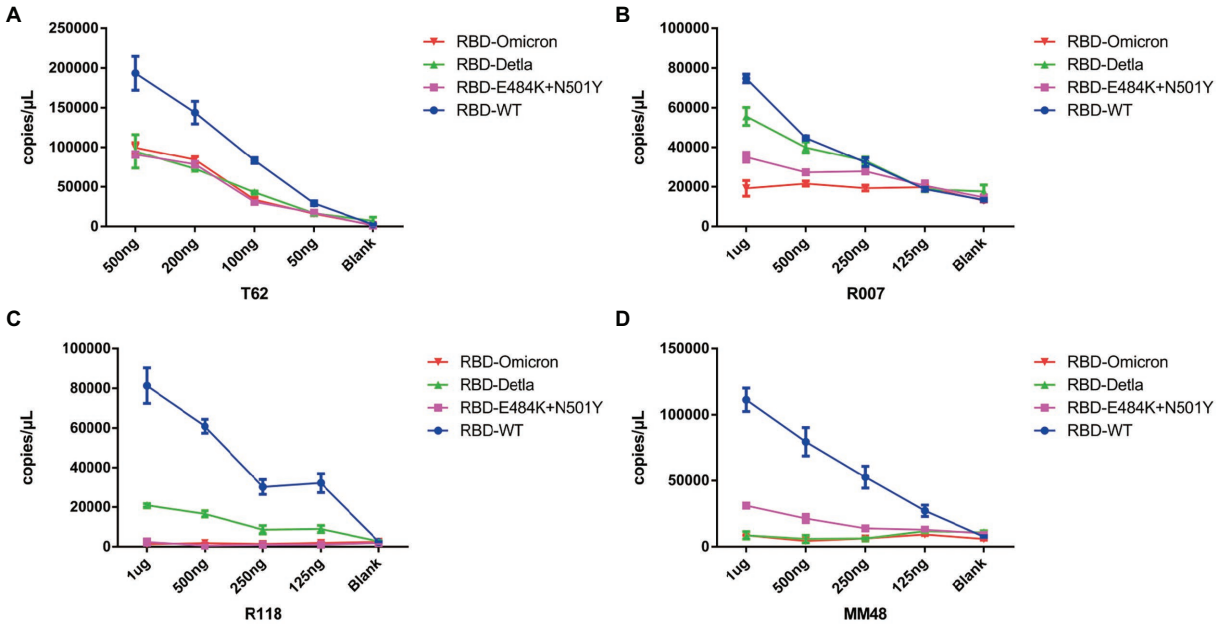


FIGURE 5
Pi-mqPCR assay to explore the enrichment of the recombinant phage-displayed RBD constructs from wild-type SARS-CoV-2 and different variants by a polyclonal antibody **(A)** and three monoclonal antibodies **(B–D)**. The values shown are the average of three independent experiments and their standard errors.

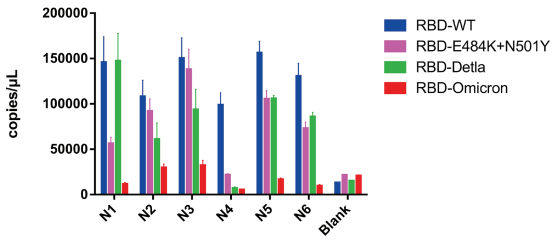


FIGURE 6
Pi-mqPCR assay to explore the enrichment of the recombinant phage-displayed RBD constructs from wild-type SARS-CoV-2 and different variants by six anti-SARS-CoV-2 RBD nanobodies. The values shown are the average of three independent experiments and their standard errors.

accurately assess immunity (Brochot et al., 2022). Compared with flow cytometry, the assay based on Pi-mqPCR can be combined with nucleic acid amplification tests and is more suitable for clinical serological diagnosis.

Since 2020, many studies have shown that mutations in the RBD region of SARS-CoV-2 can reduce the binding efficiency of monoclonal antibodies and convalescent plasma. However, limited to the linear epitope, there are very few phage-displayed antigen libraries available for differentiating the immune response to SARS-CoV-2 variants (Krishnamurthy et al., 2020; Liu, VanBlargan et al., 2021). In this research, we observed that the RBD region of SARS-CoV-2 displayed on M13 bacteriophage can be identified by antibodies and that mutations in the RBD decrease the binding efficiency. In line with a previous study, the Omicron RBD was mostly unable to bind all the types of monoclonal antibodies and nanobodies, while the RBD constructs from the Delta variant and those with the N501Y and E484K point mutations could still bind with specific antibodies. Based on this result, our system can evaluate the humoral response to different SARS-CoV-2 variants at least to a certain extent. Due to anti-infection measures, the convalescent plasma of COVID-19 patients is unavailable for our research. However, based on the results of this research, short DNA sequences, such as probes used in Pi-mqPCR or synthetic barcodes, can be inserted into the genome of M13 bacteriophage and used to measure the phage quantity. For high-throughput assay, the phage which has special barcodes can be used for immunoprecipitation with sera of the people before and after vaccination. After this, high-throughput DNA sequencing can analyze the enrichment of phage DNA to measure the humoral immune response from individual people samples. So combined with high-throughput DNA sequencing technology, this phage display system can be further applied in monitoring humoral immune response in a large population before and after vaccination. Compared with that of antigen expressed in eukaryotic systems, the dramatically lower manufacturing costs will also expand the application ranges of phage-based immunoassays.

Overall, these data show that this Pi-mqPCR assay is a valuable tool to evaluate the humoral response to the same antigen of different SARS-CoV-2 variants or antigens of different pathogens.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JnW conceived and designed the experiment, with help from DW, ZQ, DH, FQ, KH and CL. HC and SL performed experiments and analyzed the data. JlW and SH generated anti-SARS-CoV-2 RBD nanobodies. HC wrote the manuscript, with help from all other authors. All authors contributed to the article and approved the submitted version.

Funding

JnW was supported by startup funds from Beijing University of Chinese Medicine (Grant No. 1000041510051). DH was supported by a grant from Science and Technology Key Project for People's Livelihood of Guangzhou, China (Grant No. 202206010060).

Acknowledgments

The authors thank Dr. Sen Wu from the College of Biological Sciences, China Agricultural University, for discussion and suggestion. The authors thank John Myers, PhD, from American Journal Experts (http://www.aje.cn/pricing/) for editing the English text of a draft of this manuscript.

Conflict of interest

KH was employed by Andes Antibody Technology Hengshui LL Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.968036/full#supplementary-material

References

Barnes, C. O., Jette, C. A., Abernathy, M. E., Dam, K.-M. A., Esswein, S. R., Gristick, H. B., et al. (2020). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* 588, 682–687. doi: 10.1038/s41586-020-2852-1

Bostan, M., Ataman, M., Bostan, I. S., and Bleotu, C. (2020). Targets and assay types for COVID-19 diagnosis. *J. Immunoassay Immunochem.* 41, 946–959. doi: 10.1080/15321819.2020.1862866

Brochot, E., Souplet, V., Follet, P., Ponthieu, P., Olivier, C., Even, G., et al. (2022). A multiplex serological assay for the characterization of IgG immune response to SARS-CoV-2. *PLoS One* 17:e0262311. doi: 10.1371/journal.pone.0262311

Callaway, E. (2021). Heavily mutated omicron variant puts scientists on alert. *Nature* 600:21. doi: 10.1038/d41586-021-03552-w

Candel, F. J., Viñuela-Prieto, J. M., González Del Castillo, J., Barreiro García, P., Fragiel Saavedra, M., Hernández Píriz, A., et al. (2020). Utility of lateral flow tests in SARS-CoV-2 infection monitorization. *Rev. Esp. Quimioter.* 33, 258–266. doi: 10.37201/req/052.2020

Chen, Y., Guo, Y., Pan, Y., and Zhao, Z. J. (2020). Structure analysis of the receptor binding of 2019-nCoV. *Biochem. Biophys. Res. Commun.* 525, 135–140. doi: 10.1016/j.bbrc.2020.02.071

Garcia-Beltran, W. F., Lam, E. C., St Denis, K., Nitido, A. D., Garcia, Z. H., Hauser, B. M., et al. (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* 184, 2372.e9–2383.e9. doi: 10.1016/j.cell.2021.03.013

Gillot, C., Douxfils, J., Cadrobbi, J., Laffineur, K., Dogné, J.-M., Elsen, M., et al. (2020). An original ELISA-based multiplex method for the simultaneous detection of 5 SARS-CoV-2 IgG antibodies directed against different antigens. *J. Clin. Med.* 9:3752. doi: 10.3390/jcm9113752

Grossberg, A. N., Koza, L. A., Ledreux, A., Prusmack, C., Krishnamurthy, H. K., Jayaraman, V., et al. (2021). A multiplex chemiluminescent immunoassay for serological profiling of COVID-19-positive symptomatic and asymptomatic patients. *Nat. Commun.* 12:740. doi: 10.1038/s41467-021-21040-7

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., et al. (2020). SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19, 409–424. doi: 10.1038/s41579-021-00573-0

Hein, S., Herrlein, M. L., Mhedhbi, I., Bender, D., Haberger, V., Benz, N., et al. (2022). Analysis of BNT162b2- and CVnCoV-elicited sera and of convalescent sera toward SARS-CoV-2 viruses. *Allergy* 77, 2080–2089. doi: 10.1111/all.15189

Khailany, R. A., Safdar, M., and Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* 19:100682. doi: 10.1016/j.genrep.2020.100682

Krishnamurthy, H. K., Jayaraman, V., Krishna, K., Rajasekaran, K. E., Wang, T., Bei, K., et al. (2020). Antibody profiling and prevalence in US patients during the SARS-CoV2 pandemic. *PLoS One* 15:e0242655. doi: 10.1371/journal.pone.0242655

Kumar, G., Sterrett, S., Hall, L., Tabengwa, E., Honjo, K., Larimer, M., et al. (2022). Comprehensive mapping of SARS-CoV-2 peptide epitopes for development of a highly sensitive serological test for total and neutralizing antibodies. *Protein Eng. Des. Sel.* 35:gzab033. doi: 10.1093/protein/gzab033

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5

Larman, H. B., Zhao, Z., Laserson, U., Li, M. Z., Ciccia, A., Gakidis, M. A., et al. (2011). Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* 29, 535–541. doi: 10.1038/nbt.1856

Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J., and Laustsen, A. H. (2018). Basics of antibody phage display technology. *Toxins* 10:236. doi: 10.3390/toxins10060236

Li, Q., Nie, J., Wu, J., Zhang, L., Ding, R., Wang, H., et al. (2021). SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape. *Cell* 184, 2362.e9–2371.e9. doi: 10.1016/j.cell.2021.02.042

Liu, C., Ginn, H. M., Dejnirattisai, W., Supasa, P., Wang, B., Tuekprakhon, A., et al. (2021). Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* 184, 4220–4236. doi: 10.1016/j.cell.2021.06.020

Liu, Z., VanBlargan, L. A., Bloyet, L.-M., Rothlauf, P. W., Chen, R. E., Stumpf, S., et al. (2021). Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* 29, 477–488. doi: 10.1016/j.chom.2021.01.014

Mlcochova, P., Kemp, S. A., Dhar, M. S., Papa, G., Meng, B., Ferreira, I., et al. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* 599, 114–119. doi: 10.1038/s41586-021-03944-y

Mohan, D., Wansley, D. L., Sie, B. M., Noon, M. S., Baer, A. N., Laserson, U., et al. (2018). PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nat. Protoc.* 13, 1958–1978. doi: 10.1038/s41596-018-0025-6

Saito, A., Irie, T., Suzuki, R., Maemura, T., Nasser, H., Uriu, K., et al. (2022). Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* 602, 300–306. doi: 10.1038/s41586-021-04266-9

Stoddard, C. I., Galloway, J., Chu, H. Y., Shipley, M. M., Sung, K., Itell, H. L., et al. (2021). Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. *Cell Rep.* 35:109164. doi: 10.1016/j.celrep.2021.109164

Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., et al. (2020). Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* 17, 613–620. doi: 10.1038/s41423-020-0400-4

Torres, J. P., Piñera, C., De La Maza, V., Lagomarcino, A. J., Simian, D., Torres, B., et al. (2019). Severe acute respiratory syndrome coronavirus 2 antibody prevalence in blood in a large school community subject to a coronavirus disease 2019 outbreak: a cross-sectional study. *Clin. Infect. Dis.* 73, e458–e465. doi: 10.1093/cid/ciaa955

Uppalapati, L., Roitburd-Berman, A., Weiss-Ottolenghi, Y., Graham, B. S., Dimitrov, D. S., Ying, T., et al. (2022). Functional reconstitution of the MERS CoV receptor binding motif. *Mol. Immunol.* 145, 3–16. doi: 10.1016/j.molimm.2022.03.006

Vengesai, A., Kasambala, M., Mutandadzi, H., Mduluza-Jokonya, T. L., Mduluza, T., and Naicker, T. (2022). Scoping review of the applications of peptide microarrays on the fight against human infections. *PLoS One* 17:e0248666. doi: 10.1371/journal.pone.0248666

Vesper, N., Ortiz, Y., Bartels-Burgahn, F., Yang, J., de la Rosa, K., Tenbusch, M., et al. (2021). A barcoded flow cytometric assay to explore the antibody responses against SARS-CoV-2 spike and its variants. *Front. Immunol.* 12:730766. doi: 10.3389/fimmu.2021.730766

Wall, E. C., Wu, M., Harvey, R., Kelly, G., Warchal, S., Sawyer, C., et al. (2021). Neutralising antibody activity against SARS-CoV-2 VOCs B.1.617.2 and B.1.351 by BNT162b2 vaccination. *Lancet (London, England)* 397, 2331–2333. doi: 10.1016/S0140-6736(21)01290-3

Wang, H., Wu, X., Zhang, X., Hou, X., Liang, T., Wang, D., et al. (2020). SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. *ACS Cent. Sci.* 6, 2238–2249. doi: 10.1021/acscentsci.0c00742

Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J. C., et al. (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* 9:e61312. doi: 10.7554/eLife.61312

xMAP® (2022). SARS-CoV-2 antibody testing. Available at: https://www.luminexcorp.com/xmap-sars-cov-2-antibody-testing (Accessed April 29, 2022).

Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., et al. (2020). Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* 183, 739–751. doi: 10.1016/j.cell.2020.09.032

Zhang, L., Li, Q., Liang, Z., Li, T., Liu, S., Cui, Q., et al. (2022). The significant immune escape of pseudotyped SARS-CoV-2 variant omicron. *Emerg. Microbes Infect.* 11, 1–5. doi: 10.1080/22221751.2021.2017757

Zhao, J., Yuan, Q., Wang, H., Liu, W., Liao, X., Su, Y., et al. (2020). Antibody responses to SARS-CoV-2 in patients with novel coronavirus disease 2019. *Clin. Infect. Dis.* 71, 2027–2034. doi: 10.1093/cid/ciaa344

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 26 September 2022
DOI 10.3389/fmicb.2022.962124

OPEN ACCESS

EDITED BY
Sangryeol Ryu,
Seoul National University,
South Korea

REVIEWED BY
Minsuk Kong,
Seoul National University of Science and Technology, South Korea
Devin Tesar,
Genentech, Inc., United States

*CORRESPONDENCE
Frederico Aires-da-Silva
fasilva@fmv.ulisboa.pt

SPECIALTY SECTION
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

RECEIVED 05 June 2022
ACCEPTED 01 September 2022
PUBLISHED 26 September 2022

CITATION
André AS, Moutinho I, Dias JNR and Aires-da-Silva F (2022) *In vivo* Phage Display: A promising selection strategy for the improvement of antibody targeting and drug delivery properties. *Front. Microbiol.* 13:962124. doi: 10.3389/fmicb.2022.962124

In vivo Phage Display: A promising selection strategy for the improvement of antibody targeting and drug delivery properties

Ana S. André[1,2], Isa Moutinho[1,2], Joana N. R. Dias[1,2] and Frederico Aires-da-Silva[1,2]*

[1]Centro de Investigação Interdisciplinar em Sanidade Animal (CIISA), Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, Lisbon, Portugal, [2]Laboratório Associado para Ciência Animal e Veterinária (AL4AnimalS), Lisbon, Portugal

The discovery of hybridoma technology, described by Kohler and Milstein in 1975, and the resulting ability to generate monoclonal antibodies (mAbs) initiated a new era in antibody research and clinical development. However, limitations of the hybridoma technology as a routine antibody generation method in conjunction with high immunogenicity responses have led to the development of alternative approaches for the streamlined identification of most effective antibodies. Within this context, display selection technologies such as phage display, ribosome display, yeast display, bacterial display, and mammalian cell surface display have been widely promoted over the past three decades as ideal alternatives to traditional hybridoma methods. The display of antibodies on phages is probably the most widespread and powerful of these methods and, since its invention in late 1980s, significant technological advancements in the design, construction, and selection of antibody libraries have been made, and several fully human antibodies generated by phage display are currently approved or in various clinical development stages. With evolving novel disease targets and the emerging of a new generation of therapeutic antibodies, such as bispecific antibodies, antibody drug conjugates (ADCs), and chimeric antigen receptor T (CAR-T) cell therapies, it is clear that phage display is expected to continue to play a central role in antibody development. Nevertheless, for non-standard and more demanding cases aiming to generate best-in-class therapeutic antibodies against challenging targets and unmet medical needs, *in vivo* phage display selections by which phage libraries are directly injected into animals or humans for isolating and identifying the phages bound to specific tissues offer an advantage over conventional *in vitro* phage display screening procedures. Thus, in the present review, we will first summarize a general overview of the antibody therapeutic market, the different types of antibody fragments, and novel engineered variants that have already been explored. Then, we will discuss the state-of-the-art of *in vivo* phage display methodologies as a promising emerging selection strategy for improvement antibody targeting and drug delivery properties.

KEYWORDS

phage display, *in vivo*, antibody discovery, antibody engineering, antibody selection, therapeutic antibodies

Introduction

The discovery of hybridoma technology, described by Köhler and Milstein (1975), and the subsequent ability to develop monoclonal antibodies (mAbs) initiated a paradigm shift in antibody research and their clinical development. Yet, despite representing a major breakthrough in antibody-based therapy, early clinical studies demonstrated that unmodified murine mAbs presented properties that limited their use in the clinical setting. One of the most important shortcomings was the high immunogenic character of these mAbs that resulted in the generation of human anti-mouse antibody response (HAMA) that limited their therapeutic utility. Furthermore, murine mAbs demonstrated decreased serum half-life and inability to elicit human effector responses (Khazaeli et al., 1994; Hwang and Foote, 2005; Presta, 2006). To overcome these limitations, antibody engineering techniques have been explored and used to manipulate murine mAbs into chimeric or humanized antibodies by modifying their constant regions in human variants, which led to a reduction in HAMA response while promoting an efficacy optimization. More recently, fully human mAbs have being developed using hybridoma technology in transgenic mice models that have integrated into their germline human immunoglobulin (Ig) loci, such as HuMabMouse and XenoMouse platforms (Aires da Silva et al., 2008). Despite all these advances in antibody engineering and transgenic mice methods, the phage display technology has been considered, since its invention in the late 1980s, the most powerful technique for antibody discovery and development. By representing a robust and reliable method to identify specific high-affinity antigen binders from large combinatorial libraries of potentially clinically relevant antibodies, phage display technology has played a key role in the remarkable progress of discovering and optimizing antibodies for diverse applications, particularly antibody-based drugs. Nowadays, there are several antibodies and peptides generated by *in vitro* phage display currently approved or in advanced clinical stages. However, for more demanding cases aiming to generate best-in-class molecules against difficult targets, conventional *in vitro* approaches are not sufficient to fulfill unmet needs. Thus, *in vivo* phage display offers a technology capable of surpassing the drawbacks of the *in vitro* methods, making it a valuable tool to screen increasingly specific molecules. The current review provides a general overview of the antibody therapeutic market, the different types of antibody fragments, and engineered variants that have already been explored. Furthermore, we summarize the history and development of different types of antibody libraries and methods of selection with special focus on the phage display, particularly *in vivo*. Lastly, we review what has been achieved using this methodology for antibody fragments as well peptide libraries and then discuss how this strategy can improve the development of best-in-class monoclonal antibodies for cancer and other diseases.

Monoclonal antibodies

The unique specificity and high efficacy of mAbs have made them effective molecules for therapeutic and diagnostic applications. Since the approval of Orthoclone OKT3®, the first monoclonal antibody approved by Emmons and Hunsicker (1987), the use of mAbs has become a new way to target antigens in a wide variety of diseases and conditions. Besides cancer and autoimmune disorders, mAbs are being used to treat over 50 other major diseases. Applications include treatment for heart disease, allergic conditions such as asthma, and prevention of organ rejection after transplantation. Mabs are also under investigation for the treatment of central nervous disorders such as Alzheimer's disease, metabolic diseases like diabetes, and the prevention of migraines. Importantly, as medicine evolves into a new era of personalized therapy, the use of mAbs is at the core of this forefront. In fact, mAbs continue to be developed for newly identified biological targets and can be implemented into different formats to enhance their functionality and use, paving the way toward tailored medicines to each patient needs. Based on their commercial success and clinical potential, mAbs are among the top-selling drugs globally and their market size continues to grow annually. Presently, there are over 100 mAb products on the market; however, considering the thousands of candidates that are under preclinical and clinical trials and the novel advances in this field, this number is expected to increase substantially in the near future. The global monoclonal antibody market is estimated to generate the revenue of $300 billion by the end of 2025. In addition to their therapeutic applicability, the diagnosis is another area of potential expansion for these molecules. The global market for antibodies applied in diagnostics was valued at US$ 20,000 in 2017 and is projected to reach US$ 35,000 by 2026 at a CAGR of 5% from 2018 to 2026 (Parray et al., 2020).

Antibody structure and function

Antibodies, also named as immunoglobulins (Igs), are heterodimeric glycoproteins produced by B cells during the adaptive immune response. The diversity of antibody responses necessary to recognize and neutralize a wide range of antigens is achieved by recombination and somatic hypermutation of a set of genes (Aires da Silva et al., 2008; Chi et al., 2020). In mammals, the antibody basic structure consists of two identical heavy chains (H) and two identical light chains (L) in a Y-shaped format. The L chains belong to the kappa (κ) or lambda (λ) subtypes and the H chains to the α, δ, ε, γ or μ isotypes. Antibodies are divided into five different classes or isotypes: IgA, IgD, IgE, IgG, and IgM, based on sequence and length of heavy-chain constant domains, each presenting a specific structure and role in immunological processes. Due to its high prevalence in human serum, importance for immune response and excellent specificity, IgG represents the leading isotype used in immunotherapy. In the IgG class, each

heavy chain is constituted by three constant domains (CH1, CH2, and CH3) and one variable domain (VH), while the light chain consists of a single constant (CL) and a variable domain (VL; Figure 1A). The antigen-binding fragment (Fab) is composed by the variable domains positioned into the antibody N-terminus and the CL and CH1 regions. A flexible sequence (hinge region) links this fragment to the CH2 and CH3 domains, components of the fragment crystallizable (Fc; Figure 1A). Several inter-domain disulfide bonds along with the highly conserved intra-domain bonds maintain antibody integrity. On one hand, the variable domains promote antibody specificity and affinity toward antigen mainly through three hypervariable loops, collectively known as complementary determining regions (CDRs). The conformation of the VH and VL chain CDRs result in six hypervariable loop structures (H1, H2, H3, L1, L2, and L3) that form the antigen-binding site. These domains also present four relatively conserved β-sheet strands, the framework sequences, which act as scaffolds that support the CDR loops. On the other hand, the Fc region is composed only by constant domains and plays a major role in mediating antibody effector functions through antibody-dependent cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC). Furthermore, this IgG region is responsible for the prolonged antibody half-life through a recycling mechanism dependent on neonatal Fc receptor binding (FcRn).

In antibody-based therapies, the goal is to eliminate or neutralize the pathogenic organism or the disease target. Within this context, mAbs can achieve their therapeutic effect through various mechanisms. In some applications, antibody binding can directly and effectively block the activity of many pathogens or targets such as virus or receptors expressed on tumor cells. In other settings, effective treatment requires a more general immune response, and antibodies must boost effector functions, such as ADCC and/or CDC responses. In ADCC responses, upon antibody recognition of an antigen, the Fc domain engages Fc receptors (FcγRs) on the surface of effector cells such as macrophages and natural killer cells that mediate phagocytosis or lysis of the target antigen. In CDC responses, antibodies promote directly target cell death through the development of a complement chain membrane attack complex (Aires da Silva et al., 2008). These effector functions contribute to the therapeutic efficacy of several antibodies in clinical settings in which the destruction of target cells is desired, such as the removal of tumor cells or virally infected cells (Reff et al., 1994).

Antibody fragments and single-domain antibodies as promising therapeutics

The vast majority of mAbs on the market are in the form of a full IgG. These molecules have a long half-life and the ability to induce effector functions if necessary. However, the use of a whole IgG has significant limitations, such as, high molecular weight of these molecules, limiting their penetration into solid tumors, and

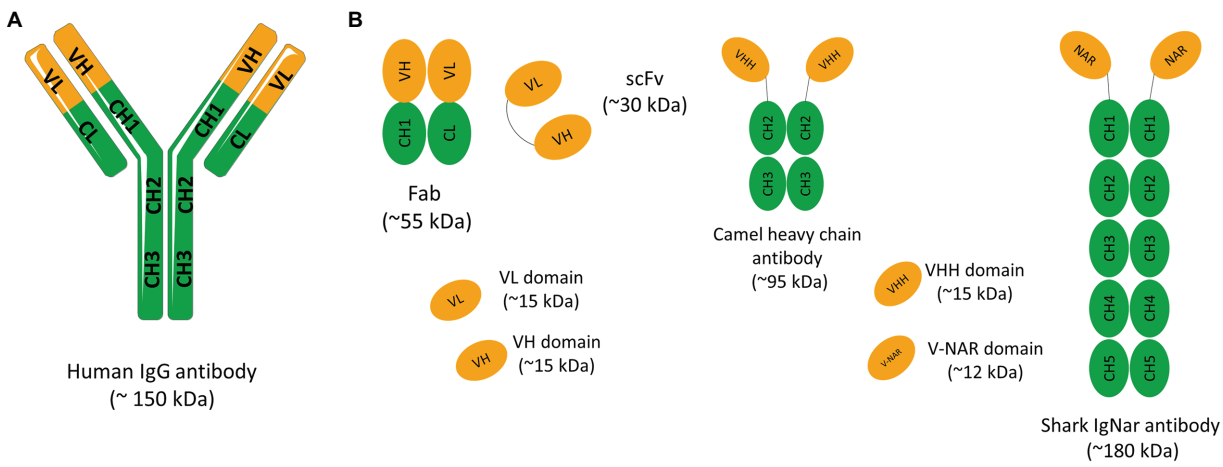


FIGURE 1
(A) Schematic representation of a conventional IgG antibody. Immunoglobulin G (IgG) have approximately 150kDa and consist of two identical light chains (L) and two identical heavy chains (H) connected by disulfide bonds. Light chains are made up of a variable domain (VL) and a constant domain (CL). Heavy chains are made up of a variable domain (VH) and three constant domains (CH_1, CH_2 e CH_3). The IgG molecule can also be divided into two main fragments: the antigen-binding-domain (Fab) and the fragment crystallizable (Fc) domain. The Fab fragment consists of two constant domains (CH_1 and CL) and two variable domains (VH and VL). **(B)** Schematic representation of different antibody fragments. Fragment of antigen binding (Fab) composed of VL and a constant domain of the light chain (CL) linked to VH and a constant domain of the heavy chain (CH_1) by a disulfide bond between the CL and CH_1 domains. Single chain fragment variable (scFv) composed only of variable regions, one from the heavy chain (VH) and the other from the light chain (VL). The two variable regions are linked by a flexible glycine-serine linker $(Gly_4Ser)_3$. Camelid and shark immunoglobulin composed of only heavy chains. They present no light chain, and the displayed V domains bind their targets separately. Camelid heavy-chain antibodies composed a homodimer of one variable domain (VHH) and two C-like constant domains (CH). Shark new antigen receptor antibodies (IgNARs) composed of one variable domain (V-NAR) and five C-like constant domains (CH).

high production costs. By selecting specific antibody domains, some properties such as half-life, tissue penetration, and affinity can be managed according to the intended application. So, different antibody fragments have been explored in the past years. The use of antibody fragments instead of antibodies in their full structure is associated with numerous advantages. Despite the removal of some domains, these fragments retain the specificity of the original immunoglobulin and for this reason, the antigen recognition remains unchanged. The smaller size of these molecules allows the access to hidden epitopes that would not be available to larger molecules such as full-size mAbs. In addition, due to the lack of glycosylation, antibody fragments can be produced using prokaryotic expression systems at a lower production cost. The lack of the Fc domain is also associated with lower immunogenicity, no antibody-dependent enhancement (ADE) effect stimulation, and lower nonspecific uptake in tissues that highly express Fc receptors (Holliger and Hudson, 2005). Therefore, to avoid Fc-associated effects in some clinical settings and address the size limitations of IgGs, smaller antibody molecules such as the antigen-binding fragment (Fab) or the single-chain variable fragment (scFv) may be produced and become more attractive as therapeutics and diagnostic agents (Figure 1B). Before the development of recombinant DNA technologies and protein engineering methods, Fabs could only be generated through proteolytic cleavage of antibodies. Currently, it is possible to clone the variable (V) and the constant (C) domain genes of the light (L) and heavy (H) chains of an IgG molecule to obtain a recombinant Fab (Figure 1B). Similarly, cloning the VH and VL variable genes of an IgG alone permits the production of an antibody in the scFv format. In this case, the scFv is constituted by a VL-VH pair connected by a flexible peptide linker, in order to improve antibody folding and stability (Figure 1B; Holliger and Hudson, 2005). Currently, several antibodies in the Fab and scFv format have been approved. For instance, Abciximab, Ranibizumab, Certolizumab pegol, and Idarucizumab have been approved in the Fab format, and Blinatumomab and Brolucizumab in the scFv format. With these antibody fragments already on the market and with several under development it is clear that antibody fragments are becoming an increasingly important class of therapeutic agents. Single-domain antibodies (sdAbs) are another class of antibody scaffolds that have gained attention as promising therapeutics.

sdAbs are the smallest functional antigen-binding fragments of an antibody that can be isolated from conventional IgGs or obtained from naturally occurring antibodies devoid of light chains that were discovered in two types of organisms, the camelids (VHH, from camels and llamas) and cartilaginous fish (V_{NAR}, from wobbegong and nurse shark; Hamers-Casterman et al., 1993; Greenberg et al., 1995; Muyldermans et al., 2001; Holt et al., 2003; Aires da Silva et al., 2004; Holliger and Hudson, 2005; Figure 1B). Due to their small size, sdAbs in the VH, VL, VHH, or V_{NAR} formats show improved tissue penetration and are able to reach targets that are not easily accessible by conventional antibodies or antibody fragments, such as enzyme active sites, hidden epitopes, or canyons in receptor molecules. Moreover, similar to Fab and scFv, sdAbs lack the Fc domain of a full IgG antibody, presenting a low nonspecific uptake in tissues that highly express Fc receptors. Additional important advantages include their high stability, low immunogenicity, and lower manufacturing costs (Holliger and Hudson, 2005). These unique characteristics and features make sdAbs promising candidates for the development of a new generation of antibody-based therapeutics. Indeed, several sdAbs have been developed and evaluated in different clinical trials and Caplacizumab is a first-in-class that was recently approved for treatment of acquired thrombotic thrombocytopenic purpura (aTTP), a rare disease characterized by excessive blood clotting in small blood vessels (Morrison, 2019). Caplacizumab is a sdAb in the VHH format that works by binding to the A1 domain of von Willebrand factor (VWF) and blocking platelets from binding to VWF and aggregating.

The small size of antibody fragments and single domains improves their ability to penetrate tumors and leads to rapid clearance from circulation through the kidney. In some therapeutic applications, rapid clearance is beneficial. However, in other cases, it is desirable to increase the half-life of the antibody. This can be achieved by linking the antibody fragment or a single domain to polyethylene glycol (PEG; Kontermann, 2011). This process, named PEGylation, increases the serum half-life and simultaneously reduces the immunogenicity of proteins by chemical coupling of PEG to amino groups on the antibody. Although this approach is an industry-established method, several studies also indicate that PEGylation can lead to reduced antigen binding and bioactivity of the PEGylated protein. As an alternative, other half-life extension methods have been developed by exploring the neonatal Fc receptor (FcRn) mediated recycling, responsible for the long half-life of the human immunoglobulin G (IgG, ~21 days) and human serum albumin protein (HSA, ~19 days). Essentially, these strategies are based on the fusion to the IgG Fc region, fusion to serum albumin and fusion to anti-albumin antibodies or peptides. Another promising strategy that has been explored is to fuse the antibody with naturally occurring albumin-binding domains (ABD) found in protein G of certain streptococcal strains (Dennis et al., 2002; Stork et al., 2007; Kontermann, 2011; Sleep et al., 2013; Cantante et al., 2017).

Engineering multivalent, bispecific, and new generation of antibodies

Converting whole IgG antibodies into Fab, scFv, and single-domains might lead to a decrease in the antigen-binding activity due to the loss of avidity. Nonetheless, this loss in binding activity can be compensated by engineering multivalent antibody fragments. Over the past years, several studies have been applied to construct multivalent antibodies through the use of either chemical or genetic cross-links. These engineered antibodies span a molecular-weight range of 60–150 kDa and valences from two to four binding sites. For instance, several Fab fragments have been chemically cross-linked into di-and trivalent multimers, leading to an increased functional affinity

(Casey et al., 1996, 2002; Weir et al., 2002). On the other hand, several strategies have been devised to genetically develop multimeric scFvs, by simple engineering the size of the amino acid linker between the VH and VL domains (Huston et al., 1993; Holliger and Hudson, 2005). These antibody engineering methods have led to the formation of dimeric molecules (diabodies, ~60 kDa), trimers (triabodies, ~90 kDa) or tetramers (tetrabodies, ~120 kDa). All these multivalent antibody fragments are usually generated as monospecific molecules. Antibody engineering strategies have also been applied to generate bispecific antibodies (BsAb). These types of antibodies bind to two different antigens, or two different epitopes on the same antigen (Hudson and Souriau, 2003). The high interest in these molecules occurs because many diseases involve parallel signal pathways or multiple inhibition of different receptors or ligands. In fact, the simultaneous blocking of different targets can limit disease progression as well as the development of possible mechanisms of resistance derived from single or combinatory therapy. BsAb that mediate cytotoxicity by recruiting and activating CD3+ T immune cells are also a promising immunotherapeutic strategy for treatment of hematological malignancies and solid cancers. BsAbs can be developed and engineered in the IgG format as well in different antibody fragment constructs. Each of these molecules, depending on their structure, demonstrate different properties of valency for antigens, half-life and in some cases effector functions (Spiess et al., 2015). Currently, several BsAbs are under development for treatment of cancer and other diseases, and five have already received FDA approval: blinatumomab, emicizumab, amivantamab, tebentafusp and faricimab.

Antibody-drug conjugates (ADCs) and chimeric antigen receptor (CAR)-T-cell-based therapeutics, are two other promising classes of engineered antibodies that have emerged as promising therapeutic strategies in oncology. ADCs comprise an antibody conjugated to a highly cytotoxic compound *via* a chemical linker that directed toward a target antigen expressed on the cancer cell surface, reducing systemic exposure and therefore toxicity (Sievers and Senter, 2013). CAR-T-cell-based therapeutics consist of an antibody-derived targeting domain fused to T-cell signaling domains that, when expressed by a T-cell, endows the T-cell with antigen specificity determined by the targeting domain of the CAR (Sterner and Sterner, 2021). Currently, 12 ADCs and 6 CAR T-cell therapies have been approved by the FDA and EMA. With these advances in modern biotechnology and miniaturization and multifunctionalization of antibodies the market for therapeutic antibody drugs has experienced explosive growth. As a result, therapeutic antibodies have become the predominant class of new drugs developed in recent years for the treatment of various human diseases, including many cancers, autoimmune, metabolic, and infectious diseases.

Selection and screening of monoclonal antibodies

The mouse hybridoma technology was an important step in the development of mAbs. A mouse hybridoma is a hybrid cell produced by injecting a specific antigen into a mouse, collecting the antibody producing cell from the mouse's spleen, and fusing it with a long-lived cancerous cell (myeloma). The resulting hybrid cell can be isolated and expanded producing many identical offsprings. Each of these daughter clones will secrete over a long period of time, the immune cell product, the antibody. A B-cell hybridoma secretes a single specific antibody known as a monoclonal antibody (Figure 2). Following the discovery of the Nobel-prize winning hybridoma technology, the door was opened for the use of mouse antibodies as human therapeutics. However, as mentioned above, murine mAbs present several properties that limit their clinical utility. Therefore, over the past few years several antibody screening methods have been explored to develop chimeric, humanized, and human high-affinity antibodies. Some of these platforms consist in display technologies, such as phage display, ribosome display, yeast display, bacterial display, mammalian cell surface display, and transgenic mice platforms that express human immunoglobulin genes. The phage display is the most commonly used *in vitro* technology to generate recombinant therapeutic antibodies from different sources and antibody formats (Figure 2). Indeed, over the past years, phage display methodology has been refined and advanced to enable the discovery of antibodies against well-validated targets as well challenging targets and unmet medical needs. Up to date, 14 FDA/EMA-approved therapeutic antibodies have been developed using phage display (Supplementary Table 1), including the world best-selling antibody adalimumab (Alfaleh et al., 2020). In the next section, we will first summarize a general overview of the classic *in vitro* phage display technology and then discuss the state-of-the-art of *in vivo* phage display methodologies as a promising innovative strategy for the improvement of antibody targeting and drug delivery properties.

Phage display technology

Phage display was firstly described in 1985 when George P. Smith demonstrated that a filamentous bacteriophage f1 was capable of displaying a fusion protein on the virion surface after the insertion of a foreign DNA fragment into the phage coat protein gene (Smith, 1985). In the early 90s, this technology was further developed and improved for the display of antibodies, mainly by the groups of Winter and McCafferty at the Laboratory of Molecular Biology (Cambridge, United Kingdom), and by the groups of Lerner and Barbas at The Scripps Research Institute (La Jolla, United States; McCafferty et al., 1990; Barbas et al., 1991). In these systems, the antibody genes are linked to the amino-terminus region of the phage minor coat protein pIII. Thus, when expressed, during normal phage biogenesis, mature phage particles incorporate the encoded fusion product, linking the antibody genotype and phenotype (Figure 3). By producing phages that express on their surface antibodies while possessing in their genome the antibody encoding gene, phage display allows the enrichment of antigen-specific phage antibodies, using immobilized or labeled antigens (Winter et al., 1994; Hoogenboom, 2002). Four main steps are part of the process of a

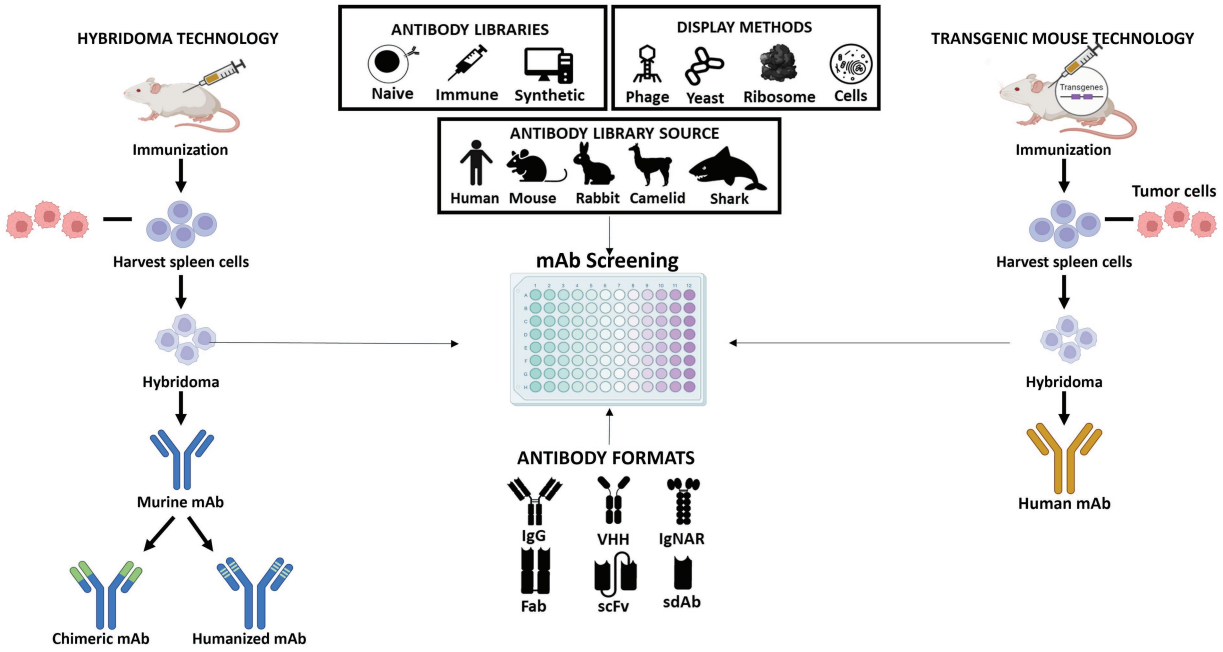


FIGURE 2
Schematic representation of the different approaches for producing monoclonal antibodies. Hybridoma technology allows the production of murine antibodies which can be modified to generate chimeric or humanized antibodies (shown at left); The procedure of obtaining human antibodies using genetically engineered mice with human immunoglobulin genes (shown at right); The central part of the figure shows the different antibody libraries (naïve, immune, and synthetic) which can be constructed from different sources (human, mouse, rabbit, camelid, or shark). The antibodies that are part of the libraries can be presented in different formats (IgG, VHH, IgNAR, Fab, scFv, or sdAb) and their screening is performed using different display technologies (phage, yeast, ribosome, or cell display).

classic phage display selection: coating of antigen; incubation of phage repertoire with antigen; washing to remove non-specific phages; and elution and reamplification of antigen-specific phages (Hoogenboom, 1997, 2005; Griffiths and Duncan, 1998; Figure 3). After the incubation of the phage library with the antigen, the unbound phages are washed away. This step is crucial to avoid the selection of non-specific binders. To obtain high-affinity phages, the stringency can be incremented in each round by increasing the number of washing steps. Moreover, the increase in stringency can also be attained through the modification of the washing buffer, for example by adding detergents (Smith and Petrenko, 1997). Regarding the elution step, different conditions can be used, such as changing the pH level, proteolytic cleavage or competition with free antigens. To facilitate cleavage, some libraries possess a cleavage site, such as trypsin, between the antibody and the pIII protein (Ward et al., 1996; Kristensen and Winter, 1998). In general, 3 to 6 rounds of binding, elution, and amplification are sufficient to recover antibodies with high specificity and affinity (Hoogenboom, 1997, 2005; Griffiths and Duncan, 1998).

The same panning principle and protocol can be applied to run a phage display to select antibodies against antigens expressed on whole cells, liposomes, virus-like particles or other systems (Kirsch et al., 2008; Lipes et al., 2008; Dominik and Kossiakoff, 2015; Jones et al., 2016; Stark et al., 2017).

Various types of phages can be used as phage display systems and depending on the type of phage, the systems can be classified as filamentous M13, T7, bacteriophage lambda (λ phage) or T4 phage display systems. Filamentous phages, namely M13, are the most common used in phage display, as they do not lyse cells during their lifecycle. As shown in Figure 3, M13 consists of a circular ssDNA and five coat proteins (pVIII, pIX, pVII, pVI, pIII). There are two major types of M13 libraries based on the peptide or antibody fusing positions of phage, pIII and pVIII library. Both contain an N-terminal signal sequence that directs them to the inner bacterial membrane before phage assembly. The size of the pIII coat protein is about 406 amino acids long and is involved in phage-host interaction during infection. The great advantage of the pIII phage library is that it can display large foreign proteins on its surface with one to five copies, that makes the pIII system the most commonly used in phage display libraries. The main advantage of this phagemid system is their smaller size and ease of cloning (Bábíčková et al., 2013; Tan et al., 2016). With the growing use of antibody fragments and single-domain antibodies, grows the necessity to improve the phage display systems. For instance, in order to improve the number of scFvs presented on filamentous phage particles, a new kind of phages were developed, known as hyperphages. Hyperphages have a pIII phenotype and are able to infect *E. coli*, however, lack the functional pIII gene, which means that the phagemid-encoded pIII-antibody fusion is the source of pIII in phage assembly. These unique features increase the phage particles that are carrying the scFv on their surface, increasing the antigen-binding activity more

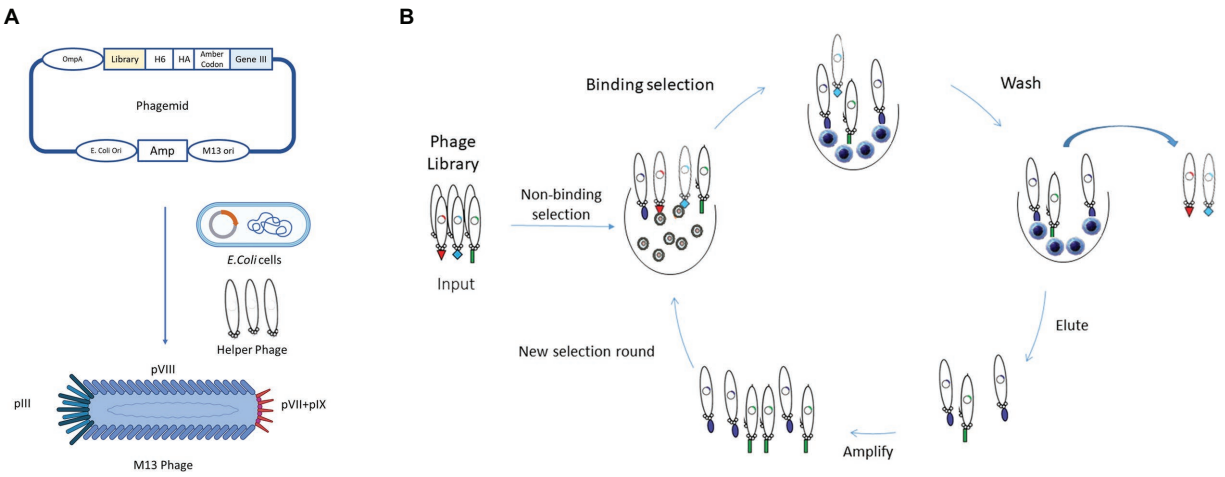


FIGURE 3
Schematic representation of *in vitro* Phage Display Technology. **(A)** In phage display systems, antibody genes are linked to the amino terminus region of the phage minor coat protein pIII, as shown in the phagemid. When expressed, mature phages will incorporate the encoded fusion product, creating a link between antibody genotype and phenotype. **(B)** *In vitro* phage display, selection is composed of several steps: coating of the antigen or preparation of the cells; incubation of phage repertoire with antigen; washing to remove non-specific phages; and elution and reamplification of antigen-specific phages. The stringency of selection could be increased by the increase in the number of washing steps.

than 50% compared to 3% when conventional phages were used (Rondot et al., 2001; Pande et al., 2010).

A critical feature to maximize the chance of discovering an antibody with the desired characteristics is the type of antibody source used. There are three main types of antibody libraries: immune, naïve, and synthetic (Winter et al., 1994; Hoogenboom, 2002; Kügler et al., 2018). Immunized libraries use as antibody source the lymphoid tissues of a donor that was previously immunized with a specific antigen or, in some instances, individuals with a particular disease, such as cancer or a particular infection. Normally, this type of library results in higher-affinity antibodies than those generated from hybridomas and naïve libraries, as gene fragments go through an *in vivo* maturation process (Winter et al., 1994; Griffiths and Duncan, 1998; Hoogenboom, 2002; Bradbury and Marks, 2004). However, as immunizations are required, immune libraries include certain limitations. These disadvantages are related with the time required to immunize animals, unpredictability of the immune response to the target antigen, lack of immune response to some antigens and a new library must be constructed for each antigen. Moreover, due to ethical concerns, human libraries can only be generated from patients B cells. On the other hand, issues related with immunizations are not present in the naïve libraries. These types of libraries are recovered from large naïve repertoires of antibody fragments from non-immunized donors and represent the germline diversity of the antibody repertoire. Due to their characteristics, naïve libraries have some advantages over immunized libraries. Foremost, antigen-specific antibodies can be produced without the need of previous immunizations, being able to get antibodies against self, non-immunogenic or toxic antigens. Furthermore, unlike immune libraries, naïve libraries can be used for all antigens, if large enough. At last, antibody generation is a much quicker process and in less than 2 weeks human antibodies can be isolated (Winter et al., 1994; Hoogenboom, 2002). The main disadvantage is the large size of the naïve library. The large size makes the exact nature of V-gene repertoire largely unknown and uncontrollable (Maynard and Georgiou, 2000). Additionally, one of the main problems reported are the poor expression and toxicity to the host bacteria with antibodies isolated from naïve libraries. This is due to the fact that antibody genes are representative of the human immunological repertoire and because of that there is no guarantee that the clone obtained can be expressed in bacteria (Maynard and Georgiou, 2000). The problems reported may be bypassed by using synthetic antibody libraries. Synthetic libraries are *in vitro* created using oligonucleotides that introduce areas of complete or tailored degeneracy into the CDRs of one or more V genes. The degeneracy introduced into specific codon positions of synthetic oligonucleotides allows the control of the degree of randomization. The first synthetic library was constructed by Marks et al. (1991). For this purpose, a repertoire of human VH genes from 49 human germline VH-gene segments *in vitro* rearranged was developed to create a synthetic CDR3 of five to eight residues. For phage display, rearranged VH genes were cloned with a human Vλ3 light chain as single-chain variable fragments (scFvs). The selection resulted in the isolation of many antibodies against haptens and one against a protein antigen. These great results were part of the first proof of concept for synthetic antibodies. Thereafter, to cover the natural length diversity of the loop of the CDR3, the diversity was expanded from 4 to 12 residues (Nissim et al., 1994). The success of the expanded library showed the importance of diversity in a longer CDR-H3 loop. To improve the synthetically rearranged VH gene repertoire of Nissim et al. (1994) and Griffiths et al. (1994) developed a novel system to add light-chain diversity to the

previous VH gene repertoire. In this study, 26 human germline Vk and 21 germline Vλ segments were assembled into complete V genes using PCR with CDR3 loops partially randomized to mimic the diversity generated by V-J gene recombination *in vivo*. In order to create a large synthetic repertoire of Fab fragments displayed on a filamentous phage, the heavy and light-chain V-gene repertoires were combined on a phage vector in bacteria using the lox-Cre-site-specific-recombination system. This resulted in the development of antibodies with affinities similar to those obtained from a secondary immune response in mice (Griffiths et al., 1994). A summary of some of the phage display antibody libraries used in the discovery of therapeutic antibodies is shown in Table 1.

In vivo phage display

With the emergence of new classes of therapeutic antibodies and novel disease targets, it has become increasingly necessary to improve the conventional phage display methodology to generate better molecules against challenging targets. Thus, to fulfill these unmet needs and to select best-in-class antibodies against antigens in their native conformation and *in vivo* settings, innovative ways to perform the phage display selection have been developed, namely the *in vivo* phage display. *In vivo* phage display consists in the selection of phage libraries using biopannings in living animals. This approach is identical to the *in vitro* phage display, however, the main difference is that the phage library is directly intravenously injected into animals and the phages are allowed to circulate in order to allow the antibodies expressed at the phage surface to bind directly to the specific target, organ or tissues. In the end, animals are perfused to wash the unbound/unspecific phages, euthanized and the desired organs are collected to recover the phages (Figure 4). With this methodology antibodies are selected in the complicated milieu of the animal based on desired pharmacokinetic and targeting specificity properties. In order to have a well-designed *in vivo* phage display experiment, it is necessary to be aware of some parameters that can modulate the expected results. Some of these are related to phage survival and pharmacokinetics in the living animal and the route of administration that can limit the success *of in vivo* selection. These key aspects of an *in vivo* phage display panning will be overviewed and discussed below.

Experiment design of *in vivo* phage display

Phage library and phage survival

The survival of phage particles in the animal is a critical issue when an *in vivo* phage display is performed. It is well known that the immune system and, in particular, the reticuloendothelial system (RES) have a preponderant role in the phage uptake. In fact, this is clearly demonstrated by the high uptake observed in the liver and spleen. To better understand these limitations, various studies compared the survival time of phages in different mice models. Zoe et al. described that 5 min after injection, the percentages of phages in circulation decreased in all mice strains and the highest level of phage particles was detected at 5 min post-injection in CF-1 and SCID mice, and 15 min post-injection in nude mice (Zou et al., 2004). Furthermore, a minor phage accumulation in the spleen was found in the immune-compromised nude and SCID strains, suggesting a facilitated leakage and reducing of non-specific trapping by RES (Zou et al., 2004). Interestingly, Srivastava and colleagues screened a phage library in 6 different mice models and suggested that the main cause for phage degradation was related to B-cell interaction (Srivastava et al., 2004). However, Zou *et al.* described contradictory results, reporting higher phage survival in the

TABLE 1 Phage display antibody libraries used in the discovery of therapeutic antibodies.

Library name	Company/ laboratory	Repertoire	Display format	Size	References
-	Dyax	Naïve	Fab	3.7×10^{10}	de Haard et al., 1999
BMV	CAT	Naïve	scFv	1.4×10^{10}	Vaughan et al., 1996
CS	CAT	Naïve	scFv	1.29×10^{11}	Lloyd et al., 2009
BMV	CAT	Naïve	scFv	1.2×10^{11}	Groves et al., 2006
HuCAL	Morphosys's	Synthetic	Fab	2.1×10^{9}	Knappik et al., 2000
HuCAL GOLD	Morphosys's	Synthetic	Fab	1.6×10^{10}	Rothe et al., 2008
HuCAL PLATINUM	Morphosys's	Synthetic	Fab	4.5×10^{10}	Prassler et al., 2011
pIX V3.0	Janssen Bio	Synthetic	Fab	3.0×10^{10}	Shi et al., 2010
XFab1	Xoma	Naïve	Fab	3.1×10^{11}	Schwimmer et al., 2013
XscFv2	Xoma	Naïve	scFv	3.6×10^{11}	Schwimmer et al., 2013
PHILODiamond	ETH Zurich	Synthetic	scFv	4.1×10^{10}	Weber et al., 2014
ALTHEA Gold Libraries	GlobalBio/ADL	Semi-synthetic	scFv	2.1×10^{10}	Valadon et al., 2019
HAL9/10	Technische Universität Braunschweig	Naïve	scFv	1.5×10^{10}	Kügler et al., 2015

Library size corresponds to the number of colony forming units (cfu) after the transformation of E. coli with the corresponding DNA library.

athymic mice strain that possess B-cells (Zou et al., 2004). Additionally, Srivastava et al. evaluated phage recovery after 5 min post-injection and verified that no more than 50% was recovered, while confirming that this value decreased with incubation time (Srivastava et al., 2004). Overall, these *in vivo* phage display findings supported incubation times ranging between 5 to 15 min after intravenous (i.v.) administration. Altogether, these studies concluded that the inclusion of a phage library circulation time optimization study is recommended for each type of phage and animal model before initiating an *in vivo* phage display experiment (Molenaar et al., 2002).

Route of administration

The route of administration is another important issue to be considered in the design of an *in vivo* phage display experiment. Different routes have been used for phage library administration, and each route may influence the uptake to the target tissue or organ. Due to the rapid systemic exposure of the phage particles, intravenous injection is the route most widely used. However, before choosing the preferential route of administration, it is advisable to consider the targeted organ or tissue. This is particularly true in organs such as the brain that have a less effective systemic delivery due to the blood–brain barrier. The complexity of the target is another problem to have in consideration. For example, in the lungs the vasculature is heterogeneous which can hinder the identification of lung specific receptors by systemic administration of the library. To overcome this problem, phage administration *via* intratracheal instillation may be executed (Wu et al., 2003). Another route with great potential is intestinal and oral administration, considering the important role that the gastrointestinal tract may play in drug diffusion and metabolism. Nevertheless, few studies have reported the use of this route, probably due to the expected degradation caused by digestive juices. Even so, Duerr et al. performed an experiment to recover peptides capable of crossing the intestinal mucosal barrier. Based on the assumption that spleen accumulation is due to the translocation of the trans-mucosal barrier, phages were retrieved from the spleen. However, no consensus sequence in the phages recovered was discovered (Duerr et al., 2004). A few years later, another study from Hamzeh-Mivehroud et al. (2008) obtained similar results, proving that trans-mucosal transport is non-specific (Hamzeh-Mivehroud et al., 2008). Another study of Akita et al. (2006) aimed to identify new peptides for peritoneal tumors of gastric cancer. For this purpose, an *in vivo* intraperitoneal biopanning was performed. The phages were injected directly into the abdominal cavity of mice presenting peritoneal tumors of human gastric cancer. This work allowed to identify peptides for the treatment of peritoneal metastasis of gastric cancer (Akita et al., 2006). Overall, these findings show that the selection of an administration route near the target tissue can have some advantages, such as the reduced uptake by other tissues (Bábíčková et al., 2013).

In vivo phage display discovery applications

In vivo phage display of peptides

In vivo phage display was first described by Pasqualini and Ruoslahti (1996). In this study, a peptide that targets the brain and kidneys was discovered using phage peptide libraries. This first study obtained promising results, culminating in the identification of the tripeptide SRL that specifically targets the brain and the CLPAVASC peptide amino acid sequence that targets the kidney (Pasqualini and Ruoslahti, 1996). Years later, another group used the same methodology to isolate other tripeptide motifs to the brain and kidney tissues (Kolonin et al., 2006). Since then, *in vivo* use of phage display has allowed the identification of several receptor-ligand pairs in malignant and benign disease contexts and different animal models. Interestingly, in 2002 the first *in vivo* phage display in humans was performed by Arap et al. (2002b). In this work an *in vivo* screening of a phage displayed peptide library was performed to investigate the molecular diversity of receptors specific of human vascular beds. For that, a peptide phage library was intravenously infused, and 15 min later tissue biopsies were recovered to retrieve the phages from various organs. A total of 47,160 peptide motifs were then shown to be localized to different organs. This large-scale screening indicated that the tissue distribution of circulating peptides was nonrandom. To validate the specificity of the isolated peptides, the interaction of an IL-11 peptide mimic isolated from the phage display screening to interleukin-11 receptor (IL-11R) of normal prostate endothelium and epithelium tissue was confirmed (Arap et al., 2002b). Later, IL-11R was validated by Zurita et al. (2004) as a molecular target for prostate cancer therapeutics. Afterward, the library obtained by Arap et al. (2002b) was further screened in two subsequent cancer patients to uncover ligand-receptors common or specific to certain vascular beds. These results allowed the identification of four native ligand-receptors, three of which have not been previously reported. Two are expressed in several tissues (integrin α4/annexin A4 and cathepsinB/Apolipoprotein E3) and the other two have a specific distribution in normal tissues (prohibitin/annexin A2 in white adipose tissue) or cancer (RAGE/leukocyte proteinase-3 in bone metastases; Staquicini et al., 2011). Altogether, these studies validated a functional screening strategy based on an *in vivo* phage display for vascular ligand-receptor mapping, aiming to identify potential markers for biotechnology and medical applications.

Another study conducted by Arap et al. (2002a) identified a new peptide through an *in vivo* screening of phage-displayed random peptide libraries targeting the prostate vasculature. One of the selected peptides homed to prostate tissue 10 to 15 times more than other organs, and was therefore coupled to a proapoptotic peptide for the development of a drug delivery system. This proapoptotic peptide is responsible for the disruption of mitochondrial membranes and has shown antitumoral efficacy in prostate cancer prone transgenic mice. Similarly, treatment of prostate proned transgenic mice with this drug delivery system

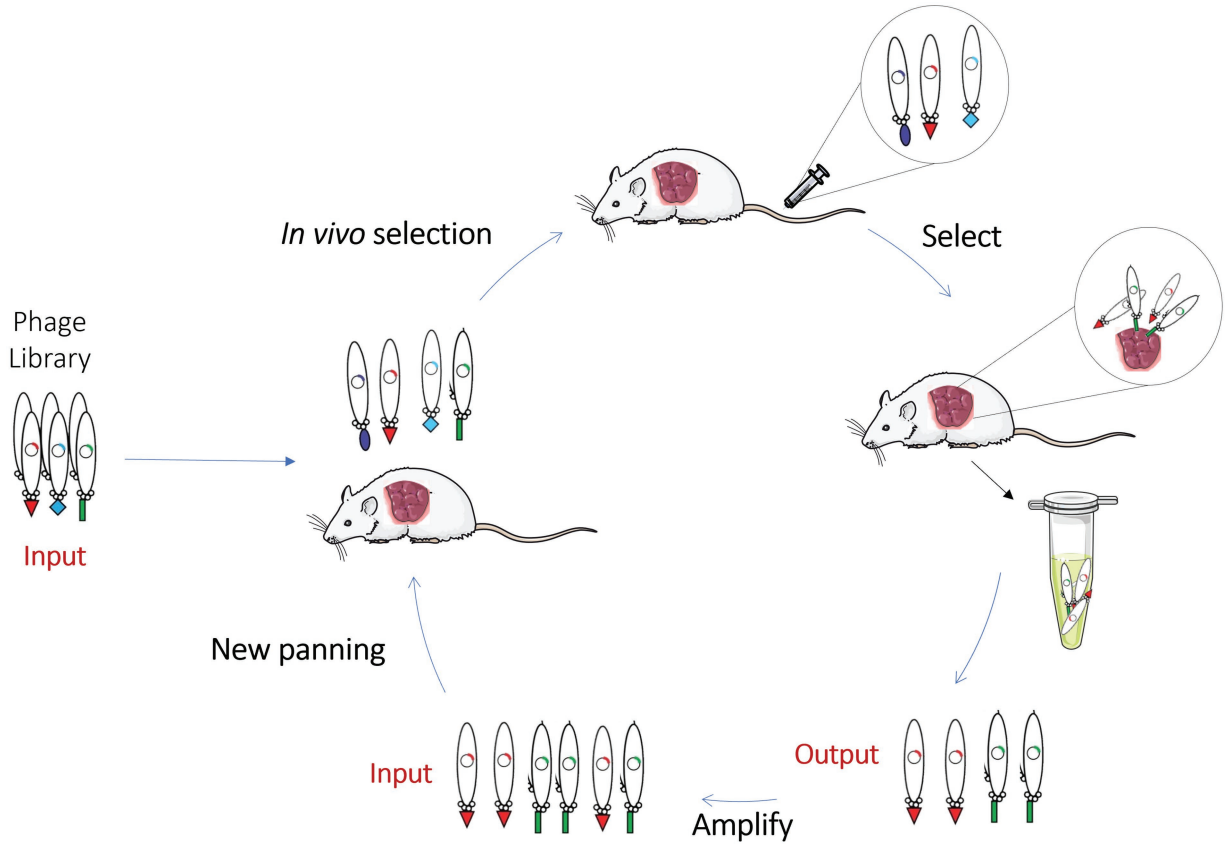


FIGURE 4
Schematic representation of *in vivo* Phage Display. *In vivo* phage display selects phage libraries using a living animal. In this methodology, the library is directly injected into animals and antibodies are allowed to bind directly to the specific organ or tissues. Non-binding phages are washed and, in the end, animals are euthanized, and the desired organs collected to recover the phages. This scheme contains a representation of an *in vivo* experiment performed on a xenograft mouse model where the phages are recovered from the tumor.

demonstrated that the chimeric peptide inhibited cancer development in this animal model (Arap et al., 2002a). Similar methodologies have been applied in the identification of peptides targeting other solid cancers, such as breast cancer. For example, an *in vivo* biopanning with a phage-displayed peptide library allowed the recognition of a cyclic nonapeptide that selectively homes normal breast tissue and identified aminopeptidase P as the respective receptor. Relevantly, this study also confirmed the binding of this novel peptide to the vasculature of hyperplastic and malignant lesions in transgenic breast cancer mice, suggesting that this peptide could be used to selectively target breast tissue and precancerous breast lesions (Essler and Ruoslahti, 2002). In turn, Laakkonen et al. (2004) aimed to investigate tumor angiogenesis related markers. For that purpose, this work designed an *ex vivo* and *in vivo* phage display to select peptides that favor tumor-homing targets accessible to circulating phages with the exception of blood vessels. Screening was performed on MDA-MB-435 breast carcinoma xenografts and enabled the recovery of multiple copies of a cyclic-9-amino-acid peptide, LyP-1. Gathered data revealed that LyP-1 also recognized tumor cells derived from an osteosarcoma xenograft mouse model, and spontaneous prostate and breast cancer in transgenic mice. LyP-1 peptide labeled with fluorescein was detected in tumor structures positive for three lymphatic endothelial markers and negative for three blood vessel markers, suggesting that this peptide may target tumor lymphatic vessels. Moreover, systemic treatment of breast tumor xenograft mice with this LyP-1 peptide was performed. Curiously, tumor treated samples showed foci of apoptotic cells in the absence of lymphatic vessels, revealing a surprising antitumoral activity of this LyP-1 peptide. Another study implemented an *in vivo* screening of a phage-displayed peptide library to discover peptide ligands that target human ovarian cancer, making it a promising candidates for drug delivery (Ma et al., 2013). This work allowed to isolate a novel peptide WSGPGVWGASVK targeting ovarian cancer derived tumor cells and angiogenic endothelial cells. Altogether these studies clearly demonstrate that *in vivo* phage display selection is a powerful platform for the discovery of cancer targeting peptides, with suitable stability and biodistribution. However, most of these peptides target the tumor vasculature rather tumor cells or tissues. Thus, there is a pressing need to improve phage display selection to identify peptides with the ability to extravasate vessels,

penetrate tumors and bind target cells while presenting adequate pharmacokinetic properties. To tackle these issues, Soendergaard et al. (2014) devised a two-tier method that includes *in vivo* selection rounds in xenografted SKOV-3 tumor-bearing nude mice followed by *in vitro* screening on cultured tumor SKOV-3 cells. These studies, resulted in the selection of a specific peptide—RSLWSDFYASASRGP (J18)—with good tumor uptake.

The use of a parallel *in vivo* and *in vitro* phage display selections was also employed in the identification of novel tumor-homing activatable cell-penetrating peptides. For the *in vivo* panning, phages displaying a library of unique activatable cell-penetrating peptides were injected into tumor-bearing mice and cleaved phage were isolated from homogenized tumors. Simultaneously, an *in vitro* selection consisting in the sequential exposure of phage to normal and tumor tissue sample extracts and its subsequent isolation from uncleaved and cleaved phages was carried out. Selected sequences were synthesized as fluorescently labeled peptides, and tumor-specific cleavage was confirmed by digestion with tissue extracts. The most promising cleaved peptide presenting the sequence RLQLKL showed to accumulate within primary and metastatic tumors in a protease-dependent manner. This candidate was therefore selected to detect protease activity in tumors, directing therapeutic or imaging agents (Whitney et al., 2010). Finally, Veleva et al. (2011) group designed a strategy to identify ligands from a phage display random peptide library, selective for cells homing angiogenic tumors derived from circulating bone marrow, that included an *in vitro* phage display selection on blood outgrowth endothelial cells followed by an *in vivo* selection based on homing of bone marrow-bound phage to angiogenic tumors. This dual-functional strategy enabled the isolation of the peptide containing the QFPPKLTNNSML sequence. Further experiments confirmed the ability of this peptide to deliver *in vivo* payloads to angiogenesis sites, reinforcing its clinical potential. Overall these studies described herein paved the way for *in vivo* phage display selections and are summarized in Table 2.

In vivo phage display of antibodies

Although *in vivo* phage display was initially explored for identifying new peptide sequences that revealed interactions with natural ligands and for the recognition of known or new targets, mAbs have also gained considerable attention due to their many advantages over other binders, namely associated with their high specificity and affinity for antigens and therapeutic properties. Within the antibody fragment repertoire, scFvs and sdAbs have been the antibody formats mainly selected and explored for *in vivo* phage display screenings. Within this context, Johns et al. (2000) described for the first time the development of an *in vivo* phage-scFv display selection procedure in the mouse. This study demonstrated the potential for the selection of scFvs against the murine thymus from a phage-scFv library, opening the possibility of implementing the same method to other models. Briefly, a scFv phage library was injected into a thymic murine model and then phages were recovered from the thymus after four rounds of selection. The specificity of some clones was tested and, in the end, CDR3 sequences of two clones were identified. The authors at the time were unable to identify the targets; however, it was reported that one scFv antigen was solely expressed on the thymic endothelium, while the second was present on both thymic endothelium and the perivascular epithelium. Later, Robert et al. (2006) combined an *in vivo* biopanning of a scFv antibody library with a high-throughput screening to identify human antibodies that target lesions developed early in atherosclerosis. First, the *in vivo* selection was performed in Bourgogne brown rabbits submitted to an atherogenic diet with 0.3% of cholesterol. Then, specificity of the selected scFv was confirmed in homozygous ApoE gene-inactivated mice. To prevent the elimination of scFvs against rare epitopes, only one round of selection was performed. This work allowed to select two scFv clones with high affinity to the early atherosclerotic lesions. Years later, another study of the same group performed an *in vivo* phage display using a human phage scFv antibody library injected into a high-fat-diet-fed ApoE$^{-/-}$ mice. Three different fractions were recovered from the aortic tissue: endothelial cell surface binding phages, and extravasated and internalized phages. Three *in vivo* pannings were conducted and six clones were selected. The results obtained allowed to conclude that *in vivo* phage display in hypercholesterolemic animals enables the identification of scFvs specific to atherosclerotic endothelial and subendothelial tissues, and biomarkers associated with lesions (Deramchia et al., 2012b). Following this study, another study was conducted to evaluate molecular components involved in atherosclerosis. For that, scFv antibody fragments from a semi-synthetic human antibody library were injected in a rabbit model of atherosclerosis induced by lesions. Following, phages were recovered from the injured endothelium, the underlying lesional tissue and the cells within the intima. Clones were selected based on the presence of key amino acids. With this method, seven scFv antibodies that demonstrated specific binding to endothelial cells and inflamed intima-related regions of rabbit tissue sections were selected. This study proved that sequence-based selection can be a suitable method to select scFv libraries by *in vivo* phage display (Deramchia et al., 2012a). More recently, Hemadou et al. (2018) used an *in vivo* phage display coupled with an high-throughput screening using flow cytometry to screen a human combinatorial library of scFv phages on an atherosclerosis rabbit model. Three *in vivo* pannings were performed and then high-throughput flow cytometry analysis allowed to recover 209 clones specific for atherosclerotic proteins. Immunohistochemistry experiments confirmed the *ex vivo* reactivity of 60% of these scFv-phages to rabbit atheromas. This multitude of atherosclerosis specific scFvs identified opened new perspectives in the discovery of new biomarkers. Recently, the same group reported the selection within the previously retrieved clones, of a promising scFv against a targetable biomarker of atheroma plaque progression in the pathological microenvironment. With that in mind, clones were reformulated as scFv fused to the scFv-Fc and their reactivity against in aorta sections from both animal models of

TABLE 2 Summary of peptides discovered by *in vivo* phage display.

Sequence	Animal model	Target	Tissue/organ	References
SRL	Balb/c mice	Unknown	Brain	Pasqualini and Ruoslahti (1996)
CLPAVASC	Balb/c mice	Unknown	Kidney	Pasqualini and Ruoslahti (1996)
LGG	C57BL/6 mice	Unknown	Brain	Kolonin et al. (2006)
SMSIARL	CD1 mice	Unknown	Prostate vasculature	Arap et al. (2002b)
CPGPEGAGC	CD1 mice	Aminopeptidase P	Breast cancer	Essler and Ruoslahti (2002)
LyP-1	Balb/c mice	Unknown	Breast cancer	Laakkonen et al. (2004)
WSGPGVWGASVK	Balb/c mice	EGFR	Ovarian Cancer	Ma et al. (2013)
RSLWSDFYASASRGP	Nude mice	Unknown	Ovarian cancer	Soendergaard et al. (2014)
RLQLKL	PytM mice	Unknown	Cancer	Whitney et al. (2010)
QFPPKLTNNSML	C57BL/6 mice	Unknown	Lewis lung carcinoma	Veleva et al. (2011)

atherosclerosis and human specimens of atherosclerotic arteries was tested using flow cytometry and immunoassays. The identification of the promising P3 scFv antibody antigen allowed to propose galectin-3 as an ideal molecular biomarker that can be targeted for diagnostic purposes. Curiously, galectin-3 has been recently proposed as a high-value biomarker for the assessment of coronary and carotid atherosclerosis (Hemadou et al., 2021). Notably, *in vivo* phage display has also been used in diabetes research field. In this context, two different methods were reported for the selection of antibodies targeting pancreatic islets. A scFv phage library was screened using two distinct approaches – an *in vivo* and *in vitro* screening. In the case of the *in vivo* selection, the scFv phage library was intravenously injected into rats and phages were isolated from the pancreatic islets. In the *in vitro* selection, the scFv library was panned on pancreatic islets directly isolated from rats. This work reported the successful isolation of four scFv antibodies that are selectively internalized into rodent β-cells *in vivo* and bind to human β-cells *in situ*. Furthermore, two additional scFv antibodies targeting islet α-cells were identified. These scFv antibodies were highly specific for their target cells over other tissues. Moreover, these radiolabeled scFv antibodies showed to reliably predict β-cell mass in rodent models of diabetes after *in vivo* administration, thereby representing promising candidates for *in vivo* imaging and determination of β-cell mass in humans (Ueberberg et al., 2009; Ueberberg and Schneider, 2010). As expected, also in cancer, *in vivo* phage display has been explored as a promising discovery platform for antibody libraries. Krag et al. (2006) reported a phase I clinical study where infusion of phage-display libraries in cancer patients were performed to identify tumor-targeting ligands. In this study, eight patients with advanced cancer, including breast, melanoma, and pancreas, were intravenously infused with a phage-displayed scFv library or a peptide library. Following 30 min post-injection, tumors were excised and tumor-targeting phages were recovered. In three patients, serial pannings were performed by infusing phage recovered and amplified from the same patient's tumor. Importantly, there was no significant toxicity observed in this preliminary study. Furthermore, the patients did not exhibit detectable antiphage serum IgG before infusion and high antibody response was not observed over the 10-day study infusion limit. Overall, the data obtained allowed the identification of several amino acid motifs among tumor-homing phage clones, suggesting selective tumor accumulation. Six different clones were isolated from the scFv library while only one was isolated from the peptide library. Thus, this study confirmed the feasibility of this technology in personalized medicine, making it possible to select patient tumor-binding ligands in a customized manner. Another study identified a new tumor-specific vascular targeting antibody *via in vivo* phage display of a llama VHH sdAb phage library in an orthotopic mouse model of diffuse glioma. C-C7, the VHH identified, recognized a subpopulation of tumor blood vessels in glioma xenografts and clinical glioma samples. The use of C-C7 VHH as bait in yeast-2-hybrid screens made it possible to recognize dynactin-1-p150Glued as its receptor, a novel targetable protein in activated endothelial cells and macrophages. Thus, validating the combinatorial approach of *in vivo* phage display and yeast-2-hybrid screenings as a promising strategy to identify tumor-targeting antibodies and their binding counterparts (van Lith et al., 2016). *In vivo* phage display screenings have also been explored to select antibodies that can cross the blood–brain barrier (BBB) and specifically target the brain. For instance, a naïve human library was used to identify brain targeting scFv antibodies using an *in vivo* phage display selection in a rat model. Two scFvs, namely scFV40 and scFV4, were selected, and their binding and BBB crossing properties were confirmed (Stutz et al., 2018). Our group has also been working toward the discovery of specific antibodies that target the BBB. We have recently immunized rabbits with brain mice endothelial cells and constructed an sdAb immune library that was then used for an *in vivo* phage display selection in a mouse model. After three rounds of *in vivo* phage display, five potential sdAbs were identified with potent brain targeting and BBB crossing properties. Moreover, a superior brain accumulation was observed for the RG3 clone, (0.82 ± 0.05% I.A./g) and, as far as we are aware, positioning it as one of the most competent sdAb in BBB translocation described so far. To demonstrate the brain targeting properties and efficacy of our platform, the RG3 clone was conjugated to liposomes containing a well-known cytotoxic drug. Notably, the RG3-sdAb

conjugate liposomes demonstrated an efficient BBB translocation and a potent antitumoral activity against glioblastoma (Aguiar et al., 2021). Our group has also been exploring the potential of *in vivo* phage display to select highly specific and potent ADC molecules. In this case, an immune sdAb library against non-hodgkin lymphoma (NHL) receptors was first submitted to three rounds of an *in vitro* whole-cell phage display. Then, the selected pool was subjected to a final *in vivo* panning on a xenograft mouse model of NHL. The recovered phage pool was submitted to next-generation sequencing and a panel of highly specific candidates were identified. One of the most promising sdAbs was selected and used to be conjugated with a cytotoxic compound to develop a new ADC (Silva et al., 2021). Taking into account the results obtained so far using *in vivo* antibody phage display, it is clear that *in vivo* screenings can be a promising selection strategy for the improvement of antibody targeting and drug delivery for several clinical applications. Table 3 summarizes the *in vivo* phage display described herein.

Discussion

Since the discovery of hybridoma (Köhler and Milstein, 1975), new perspectives have been launched in the field of antibody research and clinical development. Although promising results have been achieved, as advances in molecular tools and the disadvantages of hybridoma were more evident, new alternatives for antibody development were pursued. Due to its characteristics, display selection technologies have emerged as a promising alternative to traditional methods. Nowadays, phage display is the most widespread and powerful display technology to select antibodies and peptides. Barderas and Benito-Peña (2019) awarded with the Nobel Prize for Chemistry for the development of phage display for peptides and antibodies. The phage display created a fast route for the approval and development of molecules selected by this technique. The biggest advantage of this display method is the linkage between phenotype and encapsulated genotype (Tan et al., 2016). Furthermore, the use of phage display facilitates the cloning process due to the phage's genome small size and the high efficiency of phage infection. There are still other advantages that can be associated with this methodology, such as the small size of the phages, easy manipulation, safety, low cost of phage preparation and propagation, and the possibility of high-throughput screening of peptides and antibodies (Bábíčková et al., 2013). Additionally, upon the construction of the phage library, phages can been screened against any target antigen within several weeks (Greenwood et al., 1991). However, as with all techniques there are some pitfalls. The capacity of the phage display library and its molecular diversity is limited since it depends on bacterial transformation, phage packing and transmembrane secretion processes (Shim, 2017). Furthermore, the expression system is also limited because not all the sequences can be efficiently expressed in phages, considering that certain proteins require the correct folding to acquire its functions (Alfaleh et al., 2020). Nevertheless, phage display continues to be considered the best tool to detect the protein spatial structure, exploring binding sites between receptors and ligands, as well as the search for ligands with high affinity and biological activity. Despite the great results, out of all the therapeutic molecules approved by the FDA since 2000, only 14 monoclonal antibodies and around 30 peptides were selected using phage display, while hybridoma remains the most used methodology (Alfaleh et al., 2020; Wang et al., 2022). Within this context, it is evident that there is still a need for improvement the selected molecules by phage display so that they can be used in clinical practice. Along with this, and to surpass some of the pitfalls associated with phage display in general, the methodology has been going through some improvements in order to better meet the needs. One of the unmet needs is the increasing necessity for more specific molecules. This specificity can be achieved through *in vivo* phage display. Despite its early development, *in vivo* phage display has gained a prominent place over *in vitro* techniques. *In vivo* phage display allows the screening of molecules on a living animal or human patients, enabling the molecules to

TABLE 3 Summary of antibodies fragments discovered by *in vivo* phage display.

Name	Format	Animal model	Target	Tissue/organ	References
E3A1	scFv	CBA mice	Unknown	Thymus	Johns et al. (2000)
K3.1	scFv	Bourgogne brown rabbits	Carbonic anhydrase II	Atherosclerosis	Deramchia et al. (2012a); Deramchia et al. (2012b)
P3	scFv	Rabbit	Galectin-3	Atherosclerosis	Hemadou et al. (2018); Hemadou et al. (2021)
SCA A1 and SCA B1	scFv	Rats	Unknown	Pancreatic islets	Ueberberg et al. (2009); Ueberberg and Schneider (2010)
269	scFv	Cancer patients	Unknown	Tumor-targeting ligands	Krag et al. (2006)
C-C7	Llama nanobody	Mice	Dynactin-1-p150Glued	Glioma	van Lith et al. (2016)
scFv 4 and scFv 40	scFV	Rat	Unknown	BBB	Stutz et al. (2018)
RG3	sdAb	CD1 mice	Unknown	BBB	Aguiar et al. (2021)
C5	sdAb	Mice	Unknown	cNHL	Silva et al. (2021)

be selected in their natural environment which leads to a more specific molecule selection. Indeed, as selection is performed on an *in vivo* milieu, negative selection is naturally performed, excluding the non-specific phages that bind to non-target organs or tissues. Nevertheless, some drawbacks can be attributed to *in vivo* phage display, such as biodegradability of the phages by the immune system, interspecies differences and low reproducibility (Bábíčková et al., 2013). However, as shown in the reviewed studies, these issues can be easily surpassed *via* the optimization of some parameters, such as the time of incubation and the type of models used for panning.

It is important to take in consideration that according to the European Union (EU) Directive 2010/63/EU and the EU Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) recommendations, the EU and its members are encouraged to promote the generation of new molecules using non-animal methodologies (Gray et al., 2020a, b; Bradbury et al., 2021). These concerns are entirely valid and that is why the *in vivo* phage display can be very useful to reduce animal usage upfront in the subsequent efficacy and safety pre-clinical studies. The *in vivo* phage display has the advantage of selecting the best antibody or protein candidates given its characteristics of specificity, pharmacokinetics and stability. Indeed, with an *in vivo* phage display screening we can identify the best lead candidates *in vivo* and reduce the number of candidates to be tested in pre-clinical studies. Furthermore, a single round of *in vivo* phage display can be performed in combination with *in vitro* phage display experiments, as previously described by us and other groups.

Over the years, *in vivo* phage display has been proving itself to be an effective and powerful technique not only for the selection of peptides but also for antibody fragments. This is evident in the reviewed studies, where new molecules have been identified for many purposes, such as biomarkers and drug delivery, among many others. Furthermore, with evolving novel and more challenging disease targets and with a new generation of therapeutic antibodies emerging it is clear that *in vivo* phage display will play a key role in antibody development. Thus, with the current advances in phage display techniques, we envision that in the upcoming years many antibodies and peptides will be approved as therapeutic molecules in different areas using the classic *in vitro* phage display technology as well novel *in vivo* phage display methodologies.

Author contributions

AA: conceptualization, investigation, writing the original draft, review, and editing. IM, JD, and FA-d-S: conceptualization, investigation, writing, review, and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Portuguese Funding Agency, Fundação para a Ciência e Tecnologia, FCT IP (SAICT/2017/32085 and Ph.D. fellowship SFRH/BD/131468/2017 to AA and 2020.08209.BD to IM). CIISA has provided support through Project UIDB/CVT/00276/2020, funded by FCT and LA/P/0059/2020-AL4AnimalS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.962124/full#supplementary-material

SUPPLEMENTARY TABLE 1
Approved therapeutic antibody derived from phage display technology.

References

Aguiar, S. I., Dias, J. N. R., André, A. S., Silva, M. L., Martins, D., Carrapiço, B., et al. (2021). Highly specific blood-brain barrier transmigrating single-domain antibodies selected by an in vivo phage display screening. *Pharmaceutics* 13:1598. doi: 10.3390/pharmaceutics13101598

Aires da Silva, F., Corte-Real, S., and Goncalves, J. (2008). Recombinant antibodies as therapeutic agents: pathways for modeling new biodrugs. *BioDrugs* 22, 301–314. doi: 10.2165/00063030-200822050-00003

Aires da Silva, F., Santa-Marta, M., Freitas-Vieira, A., Mascarenhas, P., Barahona, I., Moniz-Pereira, J., et al. (2004). Camelized rabbit-derived VH single-domain intrabodies against Vif strongly neutralize HIV-1 infectivity. *J. Mol. Biol.* 340, 525–542. doi: 10.1016/j.jmb.2004.04.062

Akita, N., Maruta, F., Seymour, L. W., Kerr, D. J., Parker, A. L., Asai, T., et al. (2006). Identification of oligopeptides binding to peritoneal tumors of gastric cancer. *Cancer Sci.* 97, 1075–1081. doi: 10.1111/j.1349-7006.2006.00291.x

Alfaleh, M. A., Alsaab, H. O., Mahmoud, A. B., Alkayyal, A. A., Jones, M. L., Mahler, S. M., et al. (2020). Phage display derived monoclonal antibodies: from bench to bedside. *Front. Immunol.* 11:1986. doi: 10.3389/fimmu.2020.01986

Arap, W., Haedicke, W., Bernasconi, M., Kain, R., Rajotte, D., Krajewski, S., et al. (2002a). Targeting the prostate for destruction through a vascular address. *Proc. Natl. Acad. Sci. U. S. A.* 99, 1527–1531. doi: 10.1073/pnas.241655998

Arap, W., Kolonin, M. G., Trepel, M., Lahdenranta, J., Cardó-Vila, M., Giordano, R. J., et al. (2002b). Steps toward mapping the human vasculature by phage display. *Nat. Med.* 8, 121–127. doi: 10.1038/nm0202-121

Bábíčková, J., Tóthová, Ľ., Boor, P., and Celec, P. (2013). *In vivo* phage display — a discovery tool in molecular biomedicine. *Biotechnol. Adv.* 31, 1247–1259. doi: 10.1016/j.biotechadv.2013.04.004

Barbas, C. F., Kang, A. S., Lerner, R. A., and Benkovic, S. J. (1991). Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7978–7982. doi: 10.1073/pnas.88.18.7978

Bradbury, A. R. M., Dübel, S., Knappik, A., and Plückthun, A. (2021). Animal-versus in vitro-derived antibodies: avoiding the extremes. *MAbs* 13:1950265. doi: 10.1080/19420862.2021.1950265

Bradbury, A. R. M., and Marks, J. D. (2004). Antibodies from phage antibody libraries. *J. Immunol. Methods* 290, 29–49. doi: 10.1016/j.jim.2004.04.007

Barderas, R., and Benito-Peña, E. (2019). The 2018 Nobel Prize in Chemistry: phage display of peptides and antibodies. *Anal. Bioanal. Chem.* 411, 2475–2479. doi: 10.1007/s00216-019-01714-4

Cantante, C., Lourenço, S., Morais, M., Leandro, J., Gano, L., Silva, N., et al. (2017). Albumin-binding domain from streptococcus zooepidemicus protein Zag as a novel strategy to improve the half-life of therapeutic proteins. *J. Biotechnol.* 253, 23–33. doi: 10.1016/j.jbiotec.2017.05.017

Casey, J., King, D., Chaplin, L., Haines, A., Pedley, R., Mountain, A., et al. (1996). Preparation, characterisation and tumour targeting of cross-linked divalent and trivalent anti-tumour fab' fragments. *Br. J. Cancer* 74, 1397–1405. doi: 10.1038/bjc.1996.555

Casey, J. L., Napier, M. P., King, D. J., Pedley, R. B., Chaplin, L. C., Weir, N., et al. (2002). Tumour targeting of humanised cross-linked divalent-fab′ antibody fragments: a clinical phase I/II study. *Br. J. Cancer* 86, 1401–1410. doi: 10.1038/sj.bjc.6600198

Chi, X., Li, Y., and Qiu, X. (2020). V (D) J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology* 160, 233–247. doi: 10.1111/imm.13176

de Haard, H. J., van Neer, N., Reurs, A., Hufton, S. E., Roovers, R. C., Henderikx, P., et al. (1999). A large non-immunized human fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies*. *J. Biol. Chem.* 274, 18218–18230. doi: 10.1074/jbc.274.26.18218

Dennis, M. S., Zhang, M., Meng, Y. G., Kadkhodayan, M., Kirchhofer, D., Combs, D., et al. (2002). Albumin binding as a general strategy for improving the pharmacokinetics of proteins *. *J. Biol. Chem.* 277, 35035–35043. doi: 10.1074/jbc.M205854200

Deramchia, K., Jacobin-Valat, M.-J., Laroche-Traineau, J., Bonetto, S., Sanchez, S., Dos Santos, P., et al. (2012a). By-passing large screening experiments using sequencing as a tool to identify sc Fv fragments targeting atherosclerotic lesions in a novel in vivo phage display selection. *IJMS* 13, 6902–6923. doi: 10.3390/ijms13066902

Deramchia, K., Jacobin-Valat, M.-J., Vallet, A., Bazin, H., Santarelli, X., Sanchez, S., et al. (2012b). *In vivo* phage display to identify new human antibody fragments homing to atherosclerotic endothelial and subendothelial tissues [corrected]. *Am. J. Pathol.* 180, 2576–2589. doi: 10.1016/j.ajpath.2012.02.013

Dominik, P. K., and Kossiakoff, A. A. (2015). "Chapter eleven-phage display selections for affinity reagents to membrane proteins in Nanodiscs," in *Methods in Enzymology Membrane Proteins—Engineering, Purification and Crystallization*. ed. A. K. Shukla (Cambridge, Massachusetts: Academic Press), 219–245.

Duerr, D. M., White, S. J., and Schluesener, H. J. (2004). Identification of peptide sequences that induce the transport of phage across the gastrointestinal mucosal barrier. *J. Virol. Methods* 116, 177–180. doi: 10.1016/j.jviromet.2003.11.012

Emmons, C., and Hunsicker, L. G. (1987). Muromonab-CD3 (Orthoclone OKT3): the first monoclonal antibody approved for therapeutic use. *Iowa Med* 77, 78–82.

Essler, M., and Ruoslahti, E. (2002). Molecular specialization of breast vasculature: a breast-homing phage-displayed peptide binds to aminopeptidase P in breast vasculature. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2252–2257. doi: 10.1073/pnas.251687998

Gray, A., Bradbury, A. R. M., Knappik, A., Plückthun, A., Borrebaeck, C. A. K., and Dübel, S. (2020a). Animal-free alternatives and the antibody iceberg. *Nat. Biotechnol.* 38, 1234–1239. doi: 10.1038/s41587-020-0687-9

Gray, A. C., Bradbury, A. R. M., Knappik, A., Plückthun, A., Borrebaeck, C. A. K., and Dübel, S. (2020b). Animal-derived-antibody generation faces strict reform in accordance with European Union policy on animal use. *Nat. Methods* 17, 755–756. doi: 10.1038/s41592-020-0906-9

Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C., and Flajnik, M. F. (1995). A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* 374, 168–173. doi: 10.1038/374168a0

Greenwood, J., Willis, A. E., and Perham, R. N. (1991). Multiple display of foreign peptides on a filamentous bacteriophage. *J. Mol. Biol.* 220, 821–827. doi: 10.1016/0022-2836(91)90354-9

Griffiths, A. D., and Duncan, A. R. (1998). Strategies for selection of antibodies by phage display. *Curr. Opin. Biotechnol.* 9, 102–108. doi: 10.1016/S0958-1669(98)80092-X

Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Crosbyl, W. L., Jones, P. T., et al. (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* 16, 3245–3260. doi: 10.1002/j.1460-2075.1994.tb06626.x

Groves, M., Lane, S., Douthwaite, J., Lowne, D., Gareth Rees, D., Edwards, B., et al. (2006). Affinity maturation of phage display antibody populations using ribosome display. *J. Immunol. Methods* 313, 129–139. doi: 10.1016/J.JIM.2006.04.002

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hammers, C., Songa, E. B., et al. (1993). Naturally occurring antibodies devoid of light chains. *Nature* 363, 446–448. doi: 10.1038/363446a0

Hamzeh-Mivehroud, M., Mahmoudpour, A., Rezazadeh, H., and Dastmalchi, S. (2008). Non-specific translocation of peptide-displaying bacteriophage particles across the gastrointestinal barrier. *Eur. J. Pharm. Biopharm.* 70, 577–581. doi: 10.1016/j.ejpb.2008.06.005

Hemadou, A., Fontayne, A., Laroche-Traineau, J., Ottones, F., Mondon, P., Claverol, S., et al. (2021). In vivo human single-chain fragment variable phage display-assisted identification of Galectin-3 as a new biomarker of atherosclerosis. *JAHA* 10:e016287. doi: 10.1161/JAHA.120.016287

Hemadou, A., Laroche-Traineau, J., Antoine, S., Mondon, P., Fontayne, A., Le Priol, Y., et al. (2018). An innovative flow cytometry method to screen human sc Fv-phages selected by *in vivo* phage-display in an animal model of atherosclerosis. *Sci. Rep.* 8:15016. doi: 10.1038/s41598-018-33382-2

Holliger, P., and Hudson, P. J. (2005). Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* 23, 1126–1136. doi: 10.1038/nbt1142

Holt, L. J., Herring, C., Jespers, L. S., Woolven, B. P., and Tomlinson, I. M. (2003). Domain antibodies: proteins for therapy. *Trends Biotechnol.* 21, 484–490. doi: 10.1016/j.tibtech.2003.08.007

Hoogenboom, H. (1997). Designing and optimizing library selection strategies for generating high-affinity antibodies. *Trends Biotechnol.* 15, 62–70. doi: 10.1016/S0167-7799(97)84205-9

Hoogenboom, H. R. (2002). "Overview of antibody phage-display technology and its applications," in *Antibody phage display: Methods and protocols*. eds. P. M. O. Brien and R. Aitken (Totowa, NJ: Humana Press), 1–37.

Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* 23, 1105–1116. doi: 10.1038/nbt1126

Hudson, P. J., and Souriau, C. (2003). Engineered antibodies. *Nat. Med.* 9, 129–134. doi: 10.1038/nm0103-129

Huston, J. S., McCartney, J., Tai, M. S., Mottola-Hartshorn, C., Jin, D., Warren, F., et al. (1993). Medical applications of single-chain antibodies. *Int. Rev. Immunol.* 10, 195–217. doi: 10.3109/08830189309061696

Hwang, W. Y. K., and Foote, J. (2005). Immunogenicity of engineered antibodies. *Methods* 36, 3–10. doi: 10.1016/j.ymeth.2005.01.001

Johns, M., George, A. J. T., and Ritter, M. A. (2000). *In vivo* selection of sFv from phage display libraries. *J. Immunol. Methods* 239, 137–151. doi: 10.1016/S0022-1759(00)00152-6

Jones, M. L., Alfaleh, M. A., Kumble, S., Zhang, S., Osborne, G. W., Yeh, M., et al. (2016). Targeting membrane proteins for antibody discovery using phage display. *Sci. Rep.* 6:26240. doi: 10.1038/srep26240

Khazaeli, M. B., Conry, R. M., and LoBuglio, A. F. (1994). Human immune response to monoclonal antibodies. *J. Immunother. Emphasis Tumor Immunol.* 15, 42–52. doi: 10.1097/00002371-199401000-00006

Kirsch, M. I., Hülseweh, B., Nacke, C., Rülker, T., Schirrmann, T., Marschall, H.-J., et al. (2008). Development of human antibody fragments using antibody phage display for the detection and diagnosis of Venezuelan equine encephalitis virus (VEEV). *BMC Biotechnol.* 8:66. doi: 10.1186/1472-6750-8-66

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., et al. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* 296, 57–86. doi: 10.1006/JMBI.1999.3444

Köhler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256, 495–497. doi: 10.1038/256495a0

Kolonin, M. G., Sun, J., Do, K., Vidal, C. I., Ji, Y., Baggerly, K. A., et al. (2006). Synchronous selection of homing peptides for multiple tissues by *in vivo* phage display. *FASEB J.* 20, 979–981. doi: 10.1096/fj.05-5186fje

Kontermann, R. E. (2011). Strategies for extended serum half-life of protein therapeutics. *Curr. Opin. Biotechnol.* 22, 868–876. doi: 10.1016/j.copbio.2011.06.012

Krag, D. N., Shukla, G. S., Shen, G.-P., Pero, S., Ashikaga, T., Fuller, S., et al. (2006). Selection of tumor-binding ligands in cancer patients with phage display libraries. *Cancer Res.* 66, 7724–7733. doi: 10.1158/0008-5472.CAN-05-4441

Kristensen, P., and Winter, G. (1998). Proteolytic selection for protein folding using filamentous bacteriophages. *Fold. Des.* 3, 321–328. doi: 10.1016/S1359-0278(98)00044-3

Kügler, J., Tomszak, F., Frenzel, A., and Hust, M. (2018). "Construction of human immune and naive sc Fv libraries," in *Phage display: Methods and protocols*. eds. M. Hust and T. S. Lim (New York, NY: Springer New York), 3–24.

Kügler, J., Wilke, S., Meier, D., Tomszak, F., Frenzel, A., Schirrmann, T., et al. (2015). Generation and analysis of the improved human HAL9/10 antibody phage display libraries. *BMC Biotechnol.* 15, 1–15. doi: 10.1186/S12896-015-0125-0

Laakkonen, P., Åkerman, M. E., Biliran, H., Yang, M., Ferrer, F., Karpanen, T., et al. (2004). Antitumor activity of a homing peptide that targets tumor lymphatics and tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9381–9386. doi: 10.1073/pnas.0403317101

Lipes, B. D., Chen, Y.-H., Ma, H., Staats, H. F., Kenan, D. J., and Gunn, M. D. (2008). An entirely cell-based system to generate single-chain antibodies against cell surface receptors. *J. Mol. Biol.* 379, 261–272. doi: 10.1016/j.jmb.2008.03.072

Lloyd, C., Lowe, D., Edwards, B., Welsh, F., Dilks, T., Hardman, C., et al. (2009). Modelling the human immune response: performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens. *Protein Engineer. Design Select.* 22, 159–168. doi: 10.1093/PROTEIN/GZN058

Ma, C., Yin, G., Yan, D., He, X., Zhang, L., Wei, Y., et al. (2013). A novel peptide specifically targeting ovarian cancer identified by *in vivo* phage display: an ovarian cancer-targeting peptide identified by *in vivo* phage display. *J. Pept. Sci.* 19, 730–736. doi: 10.1002/psc.2555

Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D., and Winter, G. (1991). By-passing immunization. *J. Mol. Biol.* 222, 581–597. doi: 10.1016/0022-2836(91)90498-U

Maynard, J., and Georgiou, G. (2000). *Antibody Engineering. Annual Review of Biomedical Engineering 2. Vol. 2.* (Palo Alto, California, US: Annual Review), 339–376.

McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348, 552–554. doi: 10.1038/348552a0

Molenaar, T. J. M., Michon, I., de Haas, S. A. M., van Berkel, T. J. C., Kuiper, J., and Biessen, E. A. L. (2002). Uptake and processing of modified bacteriophage M13 in mice: implications for phage display. *Virology* 293, 182–191. doi: 10.1006/viro.2001.1254

Morrison, C. (2019). Nanobody approval gives domain antibodies a boost. *Nat. Rev. Drug Discov.* 18, 485–487. doi: 10.1038/D41573-019-00104-W

Muyldermans, S., Cambillau, C., and Wyns, L. (2001). Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. *Trends Biochem. Sci.* 26, 230–235. doi: 10.1016/s0968-0004(01)01790-x

Nissim, A., Hoogenboom, H. R., Tomlinson, I. M., Flynn, G., Midgley, C., Lane, D., et al. (1994). Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO J.* 13, 692–698. doi: 10.1002/j.1460-2075.1994.tb06308.x

Pande, J., Szewczyk, M. M., and Grover, A. K. (2010). Phage display: concept, innovations, applications and future. *Biotechnol. Adv.* 28, 849–858. doi: 10.1016/j.biotechadv.2010.07.004

Parray, H. A., Shukla, S., Samal, S., Shrivastava, T., Ahmed, S., Sharma, C., et al. (2020). Hybridoma technology a versatile method for isolation of monoclonal antibodies, its applicability across species, limitations, advancement and future perspectives. *Int. Immunopharmacol.* 85:106639. doi: 10.1016/j.intimp.2020.106639

Pasqualini, R., and Ruoslahti, E. (1996). Organ targeting *in vivo* using phage display peptide libraries. *Nature* 380, 364–366.

Prassler, J., Thiel, S., Pracht, C., Polzer, A., Peters, S., Bauer, M., et al. (2011). HuCAL PLATINUM, a synthetic fab library optimized for sequence diversity and superior performance in mammalian expression systems. *J. Mol. Biol.* 413, 261–278. doi: 10.1016/J.JMB.2011.08.012

Presta, L. G. (2006). Engineering of therapeutic antibodies to minimize immunogenicity and optimize function. *Adv. Drug Deliv. Rev.* 58, 640–656. doi: 10.1016/j.addr.2006.01.026

Reff, M. E., Carner, K., Chambers, K. S., Chinn, P. C., Leonard, J. E., Raab, R., et al. (1994). Depletion of B cells *in vivo* by a chimeric mouse human monoclonal antibody to CD20. *Blood* 83, 435–445. doi: 10.1182/blood.V83.2.435.435

Robert, R., Jacobin-Valat, M.-J., Daret, D., Miraux, S., Nurden, A. T., Franconi, J.-M., et al. (2006). Identification of human sc Fvs targeting atherosclerotic lesions. *J. Biol. Chem.* 281, 40135–40143. doi: 10.1074/jbc.M609344200

Rondot, S., Koch, J., Breitling, F., and Dübel, S. (2001). A helper phage to improve single-chain antibody presentation in phage display. *Nat. Biotechnol.* 19, 75–78. doi: 10.1038/83567

Rothe, C., Urlinger, S., Löhning, C., Prassler, J., Stark, Y., Jäger, U., et al. (2008). The human combinatorial antibody library HuCAL GOLD combines diversification of all six CDRs according to the natural immune system with a novel display method for efficient selection of high-affinity antibodies. *J. Mol. Biol.* 376, 1182–1200. doi: 10.1016/J.JMB.2007.12.018

Schwimmer, L. J., Huang, B., Giang, H., Cotter, R. L., Chemla-Vogel, D. S., Dy, F. V., et al. (2013). Discovery of diverse and functional antibodies from large human repertoire antibody libraries. *J. Immunol. Methods* 391, 60–71. doi: 10.1016/J.JIM.2013.02.010

Shi, L., Wheeler, J. C., Sweet, R. W., Lu, J., Luo, J., Tornetta, M., et al. (2010). De novo selection of high-affinity antibodies from synthetic fab libraries displayed on phage as pIX fusion proteins. *J. Mol. Biol.* 397, 385–396. doi: 10.1016/J.JMB.2010.01.034

Shim, H. (2017). "Antibody phage display," in *Recombinant antibodies for infectious diseases advances in experimental medicine and biology*. ed. T. S. Lim (Cham: Springer International Publishing), 21–34.

Sievers, E. L., and Senter, P. D. (2013). Antibody-drug conjugates in cancer therapy. *Annu. Rev. Med.* 64, 15–29. doi: 10.1146/annurev-med-050311-201823

Silva, F., Andre, A., Dias, J., Aguiar, S., Oliveira, S., and Tavares, L. (2021). Highly specific rabbit single- domain antibodies for drug delivery in immunotherapy applications. PCT/IB2022/056303.

Sleep, D., Cameron, J., and Evans, L. R. (2013). Albumin as a versatile platform for drug half-life extension. *Biochim. Biophys. Acta Gen. Subj.* 1830, 5526–5534. doi: 10.1016/j.bbagen.2013.04.023

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the Virion surface. *Science* 228, 1315–1317. doi: 10.1126/science.4001944

Smith, G. P., and Petrenko, V. A. (1997). Phage display. *Chem. Rev.* 97, 391–410. doi: 10.1021/cr960065d

Soendergaard, M., Newton-Northup, J. R., and Deutscher, S. L. (2014). *In vivo* phage display selection of an ovarian cancer targeting peptide for SPECT/CT imaging. *Am. J. Nucl. Med. Mol. Imaging* 4, 561–570.

Spiess, C., Zhai, Q., and Carter, P. J. (2015). Alternative molecular formats and therapeutic applications for bispecific antibodies. *Mol. Immunol.* 67, 95–106. doi: 10.1016/J.MOLIMM.2015.01.003

Srivastava, A. S., Kaido, T., and Carrier, E. (2004). Immunological factors that affect the *in vivo* fate of T7 phage in the mouse. *J. Virol. Methods* 115, 99–104. doi: 10.1016/j.jviromet.2003.09.009

Staquicini, F. I., Cardó-Vila, M., Kolonin, M. G., Trepel, M., Edwards, J. K., Nunes, D. N., et al. (2011). Vascular ligand-receptor mapping by direct combinatorial selection in cancer patients. *Proc. Natl. Acad. Sci. U. S. A.* 108, 18637–18642. doi: 10.1073/pnas.1114503108

Stark, Y., Venet, S., and Schmid, A. (2017). "Whole cell panning with phage display," in *Synthetic antibodies: Methods and protocols methods in molecular biology*. ed. T. Tiller (New York, NY: Springer), 67–91.

Sterner, R. C., and Sterner, R. M. (2021). CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer J.* 11, 69–11. doi: 10.1038/s41408-021-00459-7

Stork, R., Muller, D., and Kontermann, R. E. (2007). A novel tri-functional antibody fusion protein with improved pharmacokinetic properties generated by fusing a bispecific single-chain diabody with an albumin-binding domain from streptococcal protein G. *Protein Eng. Des. Sel.* 20, 569–576. doi: 10.1093/protein/gzm061

Stutz, C. C., Georgieva, J. V., and Shusta, E. V. (2018). Coupling brain perfusion screens and next generation sequencing to identify blood–brain barrier binding antibodies. *AICHE J.* 64, 4229–4236. doi: 10.1002/aic.16360

Tan, Y., Tian, T., Liu, W., Zhu, Z., and Yang, J. (2016). Advance in phage display technology for bioanalysis. *Biotechnol. J.* 11, 732–745. doi: 10.1002/biot.201500458

Ueberberg, S., Meier, J. J., Waengler, C., Schechinger, W., Dietrich, J. W., Tannapfel, A., et al. (2009). Generation of novel single-chain antibodies by phage-display technology to direct imaging agents highly selective to pancreatic β-or α-cells in vivo. *Diabetes* 58, 2324–2334. doi: 10.2337/db09-0658

Ueberberg, S., and Schneider, S. (2010). Phage library-screening: a powerful approach for generation of targeting-agents specific for normal pancreatic islet-cells and islet-cell carcinoma *in vivo*. *Regul. Pept.* 160, 1–8. doi: 10.1016/j.regpep.2009.11.017

Valadon, P., Pérez-Tapia, S. M., Nelson, R. S., Guzmán-Bringas, O. U., Arrieta-Oliva, H. I., Gómez-Castellano, K. M., et al. (2019). ALTHEA Gold libraries™: antibody libraries for therapeutic antibody discovery. *MAbs* 11, 516–531. doi: 10.1080/19420862.2019.1571879

van Lith, S. A. M., Roodink, I., Verhoeff, J. J. C., Mäkinen, P. I., Lappalainen, J. P., Ylä-Herttuala, S., et al. (2016). *In vivo* phage display screening for tumor vascular targets in glioblastoma identifies a llama nanobody against dynactin-1-p150Glued. *Oncotarget* 7, 71594–71607. doi: 10.18632/oncotarget.12261

Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., et al. (1996). Human antibodies with sub-nanomolar affinities

isolated from a large non-immunized phage display library. *Nat. Biotechnol.* 14, 309–314. doi: 10.1038/NBT0396-309

Veleva, A. N., Nepal, D. B., Frederick, C. B., Schwab, J., Lockyer, P., Yuan, H., et al. (2011). Efficient in vivo selection of a novel tumor-associated peptide from a phage display library. *Molecules* 16, 900–914. doi: 10.3390/molecules16010900

Wang, L., Wang, N., Zhang, W., Cheng, X., Yan, Z., Shao, G., et al. (2022). Therapeutic peptides: current applications and future directions. *Sig. Transduct. Target Ther.* 7:48. doi: 10.1038/s41392-022-00904-4

Ward, R. L., Clark, M. A., Lees, J., and Hawkins, N. J. (1996). Retrieval of human antibodies from phage-display libraries using enzymatic cleavage. *J. Immunol. Methods* 189, 73–82. doi: 10.1016/0022-1759(95)00231-6

Weber, M., Bujak, E., Putelli, A., Villa, A., Matasci, M., Gualandi, L., et al. (2014). A highly functional synthetic phage display library containing over 40 billion human antibody clones. *PLoS One* 9:e100000. doi: 10.1371/JOURNAL.PONE.0100000

Weir, A. N. C., Nesbitt, A., Chapman, A. P., Popplewell, A. G., Antoniw, P., and Lawson, A. D. G. (2002). Formatting antibody fragments to mediate specific therapeutic functions. *Biochem. Soc. Trans.* 30, 512–516. doi: 10.1042/bst0300512

Whitney, M., Crisp, J. L., Olson, E. S., Aguilera, T. A., Gross, L. A., Ellies, L. G., et al. (2010). Parallel in vivo and in vitro selection using phage display identifies protease-dependent tumor-targeting peptides. *J. Biol. Chem.* 285, 22532–22541. doi: 10.1074/jbc.M110.138297

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily--domains for cell surface recognition. *Annu. Rev. Immunol.* 6, 381–405. doi: 10.1146/annurev.iy.06.040188.002121

Winter, G., Griffiths, A. D., Hawkins, R. E., and Hoogenboom, H. R. (1994). Making antibodies by phage display technology. *Annu. Rev. Immunol.* 23, 433–455.

Wu, M., Pasula, R., Smith, P. A., and Martin, W. J. (2003). Mapping alveolar binding sites *in vivo* using phage peptide libraries. *Gene Ther.* 10, 1429–1436. doi: 10.1038/sj.gt.3302009

Zou, J., Dickerson, M. T., Owen, N. K., Landon, L. A., and Deutscher, S. L. (2004). Biodistribution of filamentous phage peptide libraries in mice. *Mol. Biol. Rep.* 31, 121–129. doi: 10.1023/b:mole.0000031459.14448.af

Zurita, A. J., Troncoso, P., Cardó-Vila, M., Logothetis, C. J., Pasqualini, R., and Arap, W. (2004). Combinatorial screenings in patients: the Interleukin-11 receptor α as a candidate target in the progression of human prostate cancer. *Cancer Res.* 64, 435–439. doi: 10.1158/0008-5472.CAN-03-2675

frontiers | Frontiers in Microbiology

TYPE Methods
PUBLISHED 02 December 2022
DOI 10.3389/fmicb.2022.875556

OPEN ACCESS

EDITED BY
Yoichi Takakusagi,
National Institutes for Quantum and Radiological Science and Technology, Japan

REVIEWED BY
Amelie Fradet-Turcotte,
Laval University, Canada
Hoang Duc Nguyen,
Ho Chi Minh City University of Science, Vietnam

*CORRESPONDENCE
Małgorzata Lisowska
malgorzata.lisowska@ug.edu.pl

†These authors have contributed equally to this work

‡PRESENT ADDRESSES
Anne-Sophie Huart,
ZoBio, Leiden, Netherlands
Adam Krejci,
Aelian Biotechnology GmbH,
Vienna, Austria

SPECIALTY SECTION
This article was submitted to Phage Biology, a section of the journal Frontiers in Microbiology

RECEIVED 14 February 2022
ACCEPTED 13 October 2022
PUBLISHED 02 December 2022

CITATION
Lisowska M, Lickiss F, Gil-Mir M, Huart A-S, Trybala Z, Way L, Hernychova L, Krejci A, Muller P, Krejcir R, Zhukow I, Jurczak P, Rodziewicz-Motowidło S, Ball K, Vojtesek B, Hupp T and Kalathiya U (2022) Next-generation sequencing of a combinatorial peptide phage library screened against ubiquitin identifies peptide aptamers that can inhibit the *in vitro* ubiquitin transfer cascade. *Front. Microbiol.* 13:875556. doi: 10.3389/fmicb.2022.875556

Next-generation sequencing of a combinatorial peptide phage library screened against ubiquitin identifies peptide aptamers that can inhibit the *in vitro* ubiquitin transfer cascade

Małgorzata Lisowska[1*†], Fiona Lickiss[1†], Maria Gil-Mir[2†], Anne-Sophie Huart[2†‡], Zuzanna Trybala[1], Luke Way[2], Lenka Hernychova[3], Adam Krejci[3‡], Petr Muller[3], Radovan Krejcir[3], Igor Zhukow[4], Przemyslaw Jurczak[5], Sylwia Rodziewicz-Motowidło[5], Kathryn Ball[2], Borivoj Vojtesek[3], Ted Hupp[2] and Umesh Kalathiya[1]

[1]International Centre for Cancer Vaccine Science, University of Gdańsk, Gdańsk, Poland, [2]University of Edinburgh, Institute of Genetics and Molecular Medicine, Edinburgh, United Kingdom, [3]Research Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, Brno, Czechia, [4]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland, [5]Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland

Defining dynamic protein–protein interactions in the ubiquitin conjugation reaction is a challenging research area. Generating peptide aptamers that target components such as ubiquitin itself, E1, E2, or E3 could provide tools to dissect novel features of the enzymatic cascade. Next-generation deep sequencing platforms were used to identify peptide sequences isolated from phage-peptide libraries screened against Ubiquitin and its ortholog NEDD8. In over three rounds of selection under differing wash criteria, over 13,000 peptides were acquired targeting ubiquitin, while over 10,000 peptides were selected against NEDD8. The overlap in peptides against these two proteins was less than 5% suggesting a high degree in specificity of Ubiquitin or NEDD8 toward linear peptide motifs. Two of these ubiquitin-binding peptides were identified that inhibit both E3 ubiquitin ligases MDM2 and CHIP. NMR analysis highlighted distinct modes of binding of the two different peptide aptamers. These data highlight the utility of using next-generation sequencing of combinatorial phage-peptide libraries to isolate peptide aptamers toward a protein target that can be used as a chemical tool in a complex multi-enzyme reaction.

KEYWORDS

ubiquitin, phage-peptide, next-generation sequencing, aptamers, molecular dynamics, protein–peptide binding

Introduction

The ubiquitin conjugation system has emerged as a compelling landscape in the drug discovery field (Gong et al., 2010) with PROTAC technology as an example of an innovative synthetic tool (Hu and Crews, 2021). Ubiquitin and its orthologs NEDD8, SUMO, and ISG15 are major post-translational adaptors that target proteins for a variety of molecular fates including protein degradation or stabilization, intracellular localization *via* trafficking, and/or altered biochemical function by direct or allosteric mechanisms (Ciechanover, 2015). The altered wiring of the ubiquitin and ubiquitin-like (UBL) conjugation network in age-related diseases like Alzheimer's and cancer highlights the importance in developing tools, technologies, and drug screens to intervene therapeutically in diseases involving proteostasis and proteotoxicity (Labbadia and Morimoto, 2015).

The ubiquitin and UBL conjugation reaction is catalyzed by a multi-enzyme cascade that transfers the ubiquitin or UBL molecule from an initial priming enzyme named E1, through a cascade of transfer molecules including those named E2, E3, and sometimes an E4 that ensures specific covalent linkage to the substrate molecule (Scheffner and Kumar, 2014). The E1 represents the least diverse set of priming orthologs in this conjugation system, in which there is estimated to be approximately four that charge UBL conjugation in different signaling events. These different ubiquitin like molecules include viral infection that utilizes ISG15 (Bialas et al., 2015; Ketscher and Knobeloch, 2015); cytokine signaling that utilizes FAT10, and additional UBLs named SUMO or NEDD8 (Heride et al., 2014). The E2 family of UBL conjugation system is more diverse than the E1 family and is thought to provide a degree of diversity and specificity to more localized proteomes (Middleton et al., 2014). The E3 class is the most diverse group of adaptor molecules that interacts with the E2 to transfer ubiquitin molecules directly to substrate (Das et al., 2013; Metzger et al., 2014).

Small molecule discovery in the ubiquitin system has centered on targeting E1, E2, and E3 ubiquitin (like) conjugation enzymes. The most well-developed examples include ligands that target the E1 enzyme that catalyzes NEDD8 conjugation (Bailly et al., 2016), the E2 ubiquitin conjugation enzyme CDC34 (Ceccarelli et al., 2011), and the peptide-binding pocket of the E3 ubiquitin-NEDD8-SUMO ligase MDM2 (Vassilev et al., 2004). The ubiquitin conjugation machine functions as a dynamic multi-enzyme system and provides multiple contacts for both allosteric control and alterations in protein–protein contacts (Das et al., 2013). The Cullin and MDM2 E3 ubiquitin ligases have formed a model system to identify dynamic allosteric stages in E3-mediated substrate docking and ubiquitination (Wallace et al., 2006; Deshaies and Joazeiro, 2009; Wawrzynow et al., 2009; Robson et al., 2012; Fraser et al., 2015). As these molecular machineries provide a model to study the dynamics of protein–protein interactions in an enzyme conjugation system, there is a need for novel strategies to study dynamics of protein–protein interactions, and reaction mechanisms and identify potential approaches for drug discovery.

Discovery of protein–protein interactions is a fundamental goal in the life sciences (Morelli and Hupp, 2012). The existence of millions of potential interfaces driven by linear motifs provides a large interaction landscape requiring novel tools to dissect signaling mechanisms (Tompa et al., 2014). An example is the use of ribosome display that can select for peptides binding to a target from a combinatorial pool (Fuchs et al., 2013). We have set up a robust next-generation peptide-phage library screening assay to identify peptide aptamers that could be used as tools to dissect ubiquitin reaction mechanisms. The peptide libraries are derived from bacteriophage-peptide combinatorial libraries (Smith, 1985) that are screened against a target antigen. Iterative cycles of screening and amplification can select for peptides that bind with a high specificity for a target protein and has been used previously on the E3 ubiquitin ligase MDM2 (Bottger et al., 1996; Burch et al., 2004). This method is labor intensive and always limited to an analysis of a maximum of dozens of phage clones, analyzed usually one by one. Thus, a limitation of the approach methodologically has been the low sequencing throughout of individual selected bacteriophage. Here, we exploit next-generation sequencing methodologies to deep sequence bacteriophage-peptide pools selected against Ubiquitin. With the advance of next-generation sequencing, it is now possible to sequence thousands of inserts in parallel (Dias-Neto et al., 2009; Derda et al., 2010, 2011), which we now use to identify specific peptides that bind to a target protein (ubiquitin). This methodological approach can be used as a template for dissecting many complex protein–protein interactions with an impact on the proteomic research field.

Materials and equipment

The major tools used in the data acquisition are listed below:

Ph.D. Phage Display Library−12 (New England Biolabs).
Nickel-coated ELISA 96-well plate (Thermo Scientific).
Roche454 Junior sequencing platform.
Varian Inova 500 NMR spectrometer.

Materials and methods

Peptide phage display

Peptide phage display was carried out with Ph.D. Phage Display Library−12 (New England Biolabs). NEB provides an extensive and well-explained instruction manual therefore only changes in panning procedure are described here. Random 12-mer peptides are fused to a minor coat protein (pIII) of M13 phage with pentavalent display on each phage. Displayed peptides are expressed at the N-terminus of the coding region of wild-type pIII and separated by a short spacer (Gly–Gly–Gly–Ser). The complexity

of the libraries was defined as 10^9 independent clones and the titer was about 10^{13} pfu/ml. To avoid wild-type phage contamination, all the solutions were either sterilized by filtration or autoclave.

Panning procedure

A nickel-coated ELISA 96-well plate (Thermo Scientific) was washed with TBST (TBS 1× (50 mM Tris pH 7.5, 150 mM NaCl)—0.1% (*v*/*v*) Tween-20) then His-tag ubiquitin (1 μg) or NEDD8 (1 μg; ENZO Life Sciences) in TBS was captured for 1 h at room temperature (RT). After 5 washes with TBST, the wells were then incubated for 1 h rocking with 10^{11} phages in TBST. These phages were pre-incubated rocking for 1 h at RT in an empty well. After 10 quick washes (1 min or low stringency) or 5 long washes (25 min or high stringency) with TBST, the phage particles were eluted by incubation with 100 μl of 0.2 M Glycine pH 2.2, 1 mg/ml of BSA, with gentle rocking for 10 min, and neutralized with 15 μl of 1 M Tris pH 9.1. Part of the eluted (not amplified) phages were stored at 4°C for up to 1 week. The majority of the eluted phages (90%) were then amplified by infection of ER2378 cells and the phage particles were precipitated. The biopanning procedure was repeated two times and approximately 10^{11} pfu of the first or second round amplified eluate were used as input phage. In addition, the concentration of Tween-20 in buffers was increased to 0.3% (*v*/*v*) in the second round and to 0.5% in the final round of biopanning.

PCR on phage

Phage-eluted (not amplified) or amplified phages were used directly in PCR reactions. Primers (Sigma Genosys) used were designed with adaptors for Roche 454 sequencing and are shown in Figure 1. Forward primers were ordered as PAGE purified and reverse primers were HPLC purified. Each sequencing read starts from the amplicon bar code (Figure 1). The insert target sequence length is 36 bp for 12-mer library. The size of the PCR amplicon is therefore 429 bp. Wild-type phage contamination would generate a 381 bp amplicon (see Figure 1). PCR reactions were set up on ice in nuclease-free tubes with final concentrations 1x Herculase II Fusion DNA polymerase buffer (supplied with polymerase), 0.5 M Betain solution for PCR (Sigma), 13.2 mM Trehalose, 500 μM dNTP, 500 nM of each forward and reverse primers, 1 to 10 μl of eluted or amplified phage sample, 1 μl of Herculase II Fusion DNA polymerase (Agilent Technologies), nuclease-free water up to 50 μl. Thermal cycling conditions were: 95°C for 1 min; 95°C for 15 s, 55°C for 20 s, 70°C for 1 min, repeated 30x; 70°C for 3.5 min.

DNA purification and quantification

The entire PCR reaction was loaded on a 2% (*w*/*v*) agarose gel. Amplicons at the correct size were extracted and purified from 2% (*w*/*v*) agarose gel bands by using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions including the isopropanol addition step and additional wash with QG buffer. Purified DNA was eluted in 40 μl nuclease-free water. DNA concentration was accurately quantified based on an ultrasensitive fluorescent nucleic acid stain for double-stranded DNA in solution using Quant-iT PicoGreen dsDNA kit (Life Technologies) following the manufacturer's instructions. Equal quantities of each amplicon sample were pooled together in order to prepare 100 ng (Supplementary Table 1). Exact pool concentration was re-assessed with PicoGreen.

Next,-generation sequencing and DNA sequencing data extraction. Pooled amplicons from each sample were processed by capturing on beads, amplifications using emulsion PCR, and sequenced using the Roche454 Junior according to the manufacturer's protocols. Data were extracted in Excel using a program made in java (java SE platform) in which each sequence was associated with its iteration number. The program reads .fasta and .qual files simultaneously and for each sequence, the following steps are performed: (1) Length filtering: sequences too short to contain the 36-nucleotide variable region are discarded; (2) The beginning of the read is searched for an exact match with small subsequences of forward and reverse primers. If there is a match, we know this sequence is a forward/reverse read. The bar code is then extracted. If there is no match or the bar code is incomplete, the read is discarded; (3) The 36-nucleotide insert of the read is searched for exact matches with short sequences that should be located at the borders of the variable region. The insert sequence is extracted. If this was reverse read, this sequence is now transformed in reverse and complementary form; (4) If this sequence has 36 nucleotides, it is validated to confirm that all codons are correct (as some codons are not permitted, according to the Ph.D. library manual, there also should be no stop codons). If the sequence has less than 36 nucleotides, it is discarded. If it has more than 36 nucleotides, the program attempts a repair by checking the .qual file and discarding nucleotides of the lowest quality until there are only 36 nucleotides left. Then the program performs a codon check; (5) If a sequence passes the codon check, it is translated; 6) Translated sequences are listed in a table and exported in .csv format. Sequences are then processed, using Macros in Excel (Supplementary Table 2A). To remove existing biases, each sequence that is observed more than twice in the peptide library is removed in all the samples, as they could reflect phage amplification advantage in bacteria. Moreover, all the sequences containing a minimal number of HHH motifs are discarded so as to remove phage binding to nickel-coated plates. Next, data were sorted into tables (Supplementary Tables 2B,C). A short list of synthetic peptides was ordered from Chiron Mimotopes (Australia; Supplementary Table 3) with either format: (i) PEPTIDE-three amino acid space GSG-Lys(Biotin)-amide or an N-terminal biotin is added instead of a C-terminal biotin to test activity with an N-terminal tag, (ii) biotin-conjugated to a four amino acid SGSG spacer followed by a 15-amino acid sequence with a C-terminal amide.

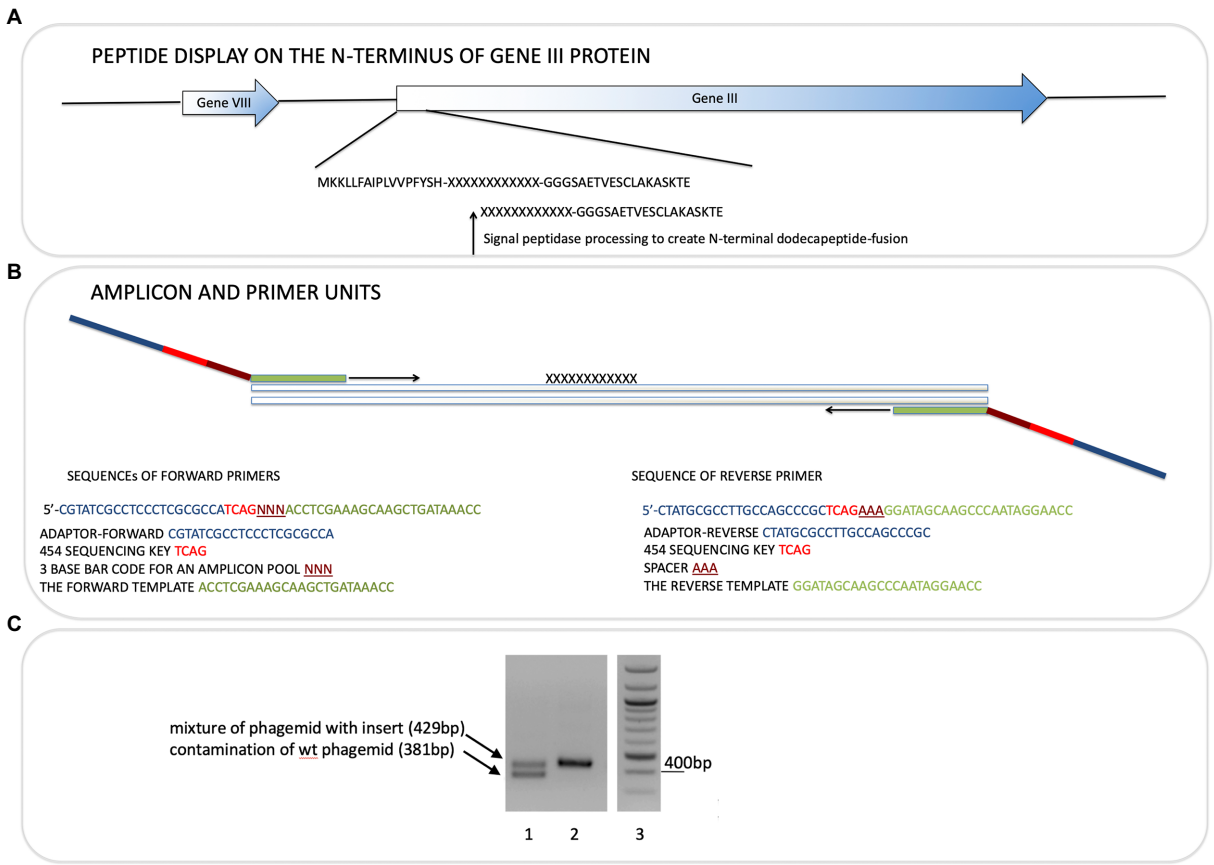


FIGURE 1
Next-generation sequencing of peptide-phage pools. His-tagged Ubiquitin or NEDD8 proteins were captured onto a nickel coated solid phase. After selection and elution of peptide phage, the phage DNA was amplified using PCR primer sets that capture the sequences flanking the peptide insert. **(A)**. The peptide sequence is at the N-terminus of the gIII M13 coat protein so that the dodecapeptide (X_{12}) would have a free N-terminus after signal peptide cleavage and the C-terminus of the peptide is linked to a GGG linker fused to gIII protein (Gene III). **(B)**. The primers used and bar codes for each PCR amplicon are highlighted, and include phages that were amplified or not amplified prior to DNA isolation for sequencing (Table 1). Pooling of all phage into deep sequencing reactions can be done with subsequent deconvolution using the trinucleotide "bar code" (NNN) whose position in the primer is indicated (Supplementary Table 2A; Figure 4 show examples of bar code stratification of peptide counts). **(C)**. An example of the amplification of peptide-phage library pools with primer pairs. The figure highlights a phage pool with a mixture of WT and peptide-containing phage (lane 1), relative to a representative peptide pool containing only peptide insert phage (lane 2) and DNA molecular mass markers (lane 3).

Peptide-binding assays

Peptide-binding ELISA was carried out in a white 96-well microtiter plate (Costar), which was coated with 1 µg of ubiquitin or 1 µg of NEDD8 (Bio-techne R&D systems) and incubated overnight at 4°C. The plate was blocked using BSA 3% (Sigma) for 1 h at RT. From 0-100 ng of biotinylated peptide-37 (that is H-FIPAQLHFHWRSGSG(LysBiotin) -NH2, in the format Amine-PEPTIDE-GSG-Lys(Biotin)-amide) was added and incubated for 1 h at 37°C (Figure 2C). In another set of experiments, 100 ng of peptides was incubated for 1, 5, 20, and 60min at 37°C (Figure 2D). Binding was detected using streptavidin-HRP-conjugated protein coupled to chemiluminescence on Varioskan Lux Reader (Thermo Fisher Scientific).

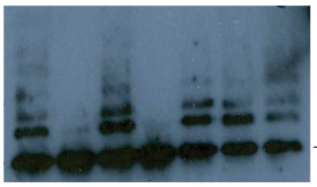
In vitro ubiquitination assay

Ubiquitination reaction had the following components: 25 mM HEPES pH 8, 6 mM MgCl2, 0.05% (*v*/*v*) Triton X-100, 3 mM ATP, 0.5 mM DTT, 1 mM benzamidine, 10 µM Ubiquitin (Boston Biochem), 100 nM E1 (UBE1, Boston Biochem), 1 µM E2 (His-Ubch5a, Boston Biochem), 30 µM creatine phosphate, 1.2 µM creatine kinase), 46 nM p53 (50 ng), 37 nM MDM2 (50 ng), the RING domain of MDM2 (0.55 µM, 100 ng), or his-tagged CHIP (12 nM, 10 ng) and the indicated amount of synthetic peptide [1.25 µg (or approximately 25–35 µM of each peptide depending upon precise mass), unless otherwise indicated], and DMSO control to a final concentration that matches the volume of the peptide added (peptides are resuspended in 100% DMSO). The synthetic peptides (from

A

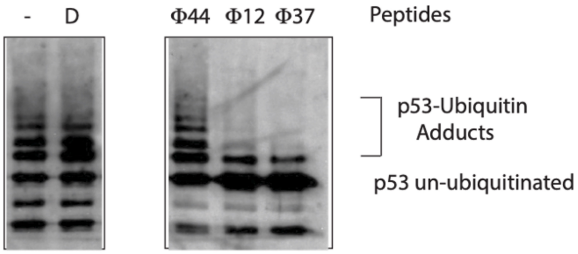
MDM2 catalyzed p53 ubiquitination

D 12 29 37 44 48 60 Individual ubiquitin binding peptides



B

CHIP catalyzed p53 ubiquitination



C

Binding specificity of Φ37 over 1 hour incubation at 37°C with protein (ubiquitin or NEDD8) captured on ELISA plate.

	Ubiquitin 1 ug		NEDD8 1 ug		Ctrl (no protein)	
Φ37	Average OD	**Corrected**	Average OD	**Corrected**	Average OD	**Corrected**
100 ng	92040	**70730**	68870	**68446,85**	47545	**47299,8**
50 ng	54355	**33045**	30075	**29651,85**	24125	**23879,8**
25 ng	38525	**17215**	14725	**14301,85**	11860	**11614,8**
12.5 ng	31760	**10450**	6119	**5695,85**	5017,5	**4772,3**
6.25 ng	27040	**5730**	2652	**2228,85**	1819,5	**1574,3**
0	21310	**0**	423,15	**0**	245,2	**0**

D

Binding specificity of Φ37 over time course: 1, 5, 10, 20 and 60 minutes incubated at RT with protein (ubiquitin or NEDD8) captured on ELISA plate.

		Ubiquitin 1 ug		Nedd 1 ug		Ctrl (no protein)	
Φ37 (ng)	Incubation (min)	Average OD	**Corrected**	Average OD	**Corrected**	Average OD	**Corrected**
100	1	21270	**-3090**	1977	**1030,4**	1047,35	**481,85**
100	5	32030	**7670**	7630,5	**6683,9**	4274	**3708,5**
100	10	31490	**7130**	17140	**16193,4**	6476	**5910,5**
100	20	37085	**12725**	15750	**14803,4**	10206	**9640,5**
100	60	52765	**28405**	31120	**30173,4**	23190	**22624,5**
0	-	24360	**0**	946,6	**0**	565,5	**0**

FIGURE 2
Identification of peptides with bioactivity in ubiquitin assays. **(A)** Ubiquitin assays driven by the E3 ubiquitin ligase MDM2 were assembled with individual peptides from the reactive pools derived from the 60 peptides in Supplementary Table 3. Representative assay showing that specific peptides such as 12 and 37 reproducibly inhibited MDM2 catalyzed ubiquitination (from the left, lanes 2 and 4 vs. lane 1 DMSO-only control). **(B)** Ubiquitin assays driven by the E3 ubiquitin ligase CHIP were assembled with individual peptides (labeled Φ) that were active in inhibiting MDM2-driven reactions (12 and 37) and a non-active ubiquitin-binding peptide 44 (Figure 2B). D represents DMSO-only control. Reactions were processed by immunoblotting to detect changes in p53 ubiquitination as indicated in the methods. **(C,D)** Analysis by ELISA of peptide-37 activity with a C-terminal biotin tag. Ubiquitin or NEDD8 was coated on the solid-phase to mimic the original peptide-phage screen. Peptides were added and then either titrated in duplicates as indicated from 6.25–100 ng for 60min **(C)** or else fixed peptide levels (100 ng) were added in duplicates as a time course from 1 to 60 min **(D)**. The data are tabulated as binding in ECL units.

Chiron Mimotopes, Australia) had a N-terminal biotin followed by an SGSG spacer, then the peptide sequence containing a C-terminal amide (See Supplementary Table 3). The MDM2- or CHIP-dependent ubiquitination reactions were carried out at 30°C for 12 min, as described previously (Fraser et al., 2015; Narayan et al., 2015). To stop the reaction 22 µl of 2× sample

buffer (5% (*w*/*v*) SDS, 25% (*v*/*v*) glycerol, 200 mM DTT, 0.3M Tris, pH 6.8) was added and the samples were analyzed by 4–12% NuPAGE Gels followed by immunoblotting using the p53 antibody DO1 antibody to measure p53 ubiquitination.

C18-reverse phase chromatography and deuterium exchange mass spectrometry

The ubiquitin protein either free or in complex with peptide-12 (Biotin-SGSG-AKFDMHIATRLS-amide) or peptide-37 (Biotin-SGSG-FIPAQLHFHWRS-amide) used in Ubiquitination reactions were initiated by dilution into deuterated water with 0.1% DMSO to a final concentration of 2 µM. The molar ratios between ubiquitin and peptide-12 or peptide-37 were 1:2 and 1:5. The incubation was carried out at room temperature and was quenched by the addition of 0.875MHCl in 1 M glycine at 1 min, followed by rapid freezing in liquid nitrogen. Each sample prepared for global HDX-MS analysis was thawed and injected onto trap column and desalted on-line on a microtrap (Michrom Bioresources, Auburn, CA) for 2 min at flow rate 20 µl/min. Next, the protein was eluted onto an analytical column (Jupiter C18, 1.0 × 50 mm, 5 µm, 300 Å, Phenomenex, CA) and separated using a linear gradient elution of 10%B in 2 min, followed by 31 min isocratic elution at 40%B. Solvents were: A, 0.1% formic acid in water; B, 80% acetonitrile/0.08% formic acid. The trap cartridge and the analytical column were kept at 1°C. Mass spectrometric analysis was carried out using an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) with ESI ionization on-line connected with a robotic system based on the HTS-XT platform (CTC Analytics, Zwingen, Switzerland). Analysis of deuterated samples was done in HPLC-MS mode with ion detection in the orbital ion trap, the data were processed manually, and deconvolution was done with *MagTran* software (Zhang and Marshall, 1998).

NMR analysis of the ubiquitin-binding peptides

Nuclear magnetic resonance (NMR) measurements were performed with a Varian Inova 500 NMR spectrometer operating at 11.7 T (^{1}H resonance frequency 500.606 MHz) equipped with a triple-resonance ^{1}H/^{13}C/^{15}N probe head and z-gradient Performa IV unit. N^{15}-labeled Ubiquitin was purchased from Asla Biotech (Latvia). The ^{1}H-^{15}N HSQC spectra were recorded at 298 K on the sample containing 0.5 mM uniformly labeled ^{15}N-human ubiquitin in 10 mM sodium phosphate buffer (92%/8% H_2O/D_2O) at pH 5.0 in the absence and presence of the peptides. Non-tagged peptide-12 (AKFDMHIATRLS) or Peptide-37 (FIPAQLHFHWRS) in this case were non-biotinylated with a C-terminal NH_2. The reason a C-terminal amide was used instead of a carboxylic acid is that the original peptide when fused to gIII protein in the phage would not have an acidic C-terminus. The ubiquitin-peptide complexes were obtained by dissolving 0.64 mg (peptide 12) or 0.7 mg (peptide 37) of lyophilized powder in 0.55 ml of the ubiquitin sample. The spectra were referenced indirectly with respect to external DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) with the Ξ coefficient equal to 0.101329118 for ^{15}N nuclei (Wishart et al., 1995). All recorded spectra were processed by NMRPipe (Delaglio et al., 1995) and analyzed using Sparky software (Lee et al., 2015). The assignments of the amide peaks for the human ubiquitin were based on BMRB databank (BMRB 4493) and our previous studies. The ^{15}N relaxation measurements were conducted with pulse sequences included in Agilent BioPack software (Agilent Inc. Palo Alto, USA) which was written on the base of previously published experiments (Farrow et al., 1994). The ^{15}N R_1 relaxation rates were obtained on the basis of eight experimental delays (10, 90, 170, 290, 410, 550, 690, and 850 ms). The ^{15}N R_2 relaxation rates were measured with the Carr–Purcell–Meiboom–Gill (CPMG) pulse train using 650 µs for refocusing time. The values were extracted from nine delays (10, 30, 50, 70, 90, 110, 130, 170, and 210 ms). The 3.0 s relaxation delay was employed in both experiments. ^{15}N R_1 and R_2 relaxation rates were determined by non-linear least-squares fit using peak amplitudes to a single exponent curve. Errors were extracted from the covariance matrix. The ^{1}H-^{15}N NOE experiments were acquired with a recycling delay in 6 s. The ^{1}H–^{15}N NOE values were obtained as a ratio between peak heights in reference and saturated spectra. Errors were estimated on the base of signal-to-noise ratio in both spectra. Analysis of the evaluated relaxation data was performed according to Spectral Density Mapping (SDM) approach, which provides the spectral density function (J(ω)) at three independent frequencies –0, ω_N and $0.87\omega_H$ (Farrow et al., 1995). Additional analysis was performed by determining the R_1 and R_2 product for selected residues exhibiting slow (µs–ms) dynamic motions (Kneller et al., 2002). The 2D ^{1}H–^{15}N HSQC spectrum collected for human ubiquitin is corresponding fully with the literature data. The sequence-specific assignments were performed on the base of BioMagnetic Resonance Databank (BRMB 4493) and previous 2D and 3D NMR experiments performed in our laboratory on ^{13}C^{15}N-uniformly labeled samples (Supplementary Figure 3). The recorded ^{15}N relaxation data (R_1, R_2, and ^{1}H-^{15}N NOE) acquired at magnetic field 11.7T (Supplementary Figure 4) are in line with available the data previously recorded in another group (BRMB 6470). They confirmed that during the performed experiments, human ubiquitin occurred in properly folded state with a very stable 3D structure.

Results

Isolation of ubiquitin-specific peptide aptamers through deep DNA sequencing of combinatorial phage peptide library pools

Recombinant human ubiquitin (with an N-terminal poly-Histidine tag) was immobilized on a nickel-coated plate in order to present a native or folded conformation in solution to the

12-mer combinatorial peptide-phage library. A control included the ubiquitin-like protein NEDD8 whose use would measure the specificity in the isolation of peptides to a structurally homologous protein but one with a different sequence and surface charge landscape (Figure 3). Three rounds of phage-peptide selection toward the two target proteins involved: (i) incubating the immobilized target proteins with the 12-mer peptide library for 1 h at 21°C; (ii) washing away non-specific or weakly bound peptides using two types of washes (either a 1 min (fast), low stringency wash or a 25 min (slow), high stringency wash to isolate peptides with a relatively high or low-off rate, respectively); (iii) eluting the specifically bound peptide-phage (unamplified phage-peptides); (iv) propagating the phage in bacteria (amplified phage-peptides), and (v) processing the amplified and unamplified phage DNA using RT-PCR DNA for next-generation DNA sequencing. Amplification of phages in bacteria after each round is necessary for multiplying the copy number of each clone to generate an enriched library that can be bio-panned again. Published results showed that amplification can enrich a subset of bacteriophage clones and thus identified a collapse of diversity after a single round of bacteriophage replication in bacteria (Matochko et al., 2012). As such, we focused our data analysis on next-generation 'deep' sequencing of peptide-phage processed directly from eluates (unamplified) to minimize amplified-based bias in peptide identification and/or motif discovery.

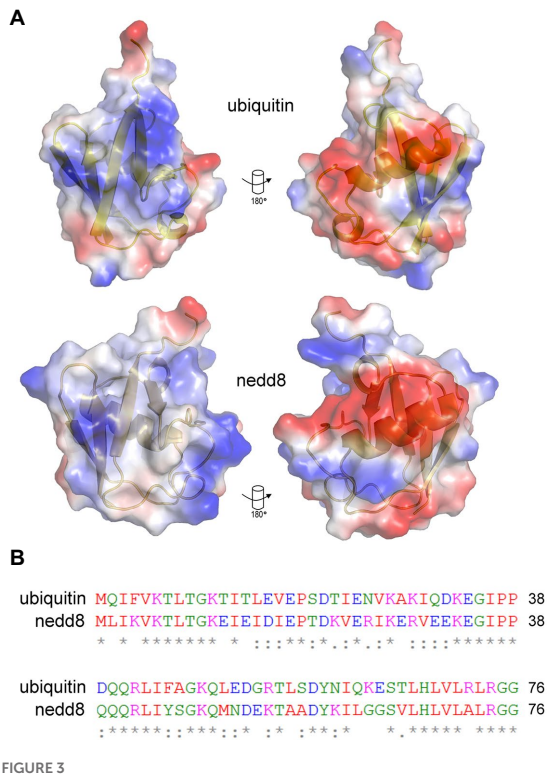


FIGURE 3
Similarities and differences between NEDD8 and ubiquitin proteins. **(A)** Charge density and structural similarities between ubiquitin and its ortholog NEDD8. **(B)** The amino acid sequence of NEDD8 and ubiquitin highlights a relatively high degree of divergence at the primary amino acid level. NEDD8 was used as a control to determine whether ubiquitin-specific peptides could be isolated using next-generation sequencing of peptide-phage library pools.

The bacteriophage DNA was extracted from each non-amplified or amplified M13 peptide-phage pool corresponding to each round and was then subjected to RT-PCR using DNA primers containing Roche 454 J adaptors as summarized in Figures 1A,B. Each PCR primer targeted a specific peptide phage pool and had a distinct internal triplet bar code that tags the DNA sequences to be extracted using Roche 454 next-generation sequencing datasets (Table 1; Figure 1B). A typical PCR product with an insert of the correct size of 429 bp is shown in Figure 1C (lane 2). An example of a PCR product with both insert (429 bp) and wild type (WT) phage (381 bp) is also shown (Figure 1C, lane 1). If WT-phage without insert contaminated a PCR reaction (as in Figure 1C, lane 1), then the upper band was excised for next-generation sequencing. Once all PCR products (from either Ubiquitin or NEDD8, rounds 1–3, low and high stringency washes, and included non-amplified and amplified peptide phage) were purified, then either 2.5 or 5 ng of gel purified DNA was pooled into one final sequencing pot containing 100 ng for next-generation sequencing (Supplementary Table 1). Unamplified phage isolated after elution became the major focus of analysis since these will not be biased by subsequent propagation in *E. coli*.

The DNA was processed by Next-Generation Sequencing methodologies according to 454 J protocols that can identify up to 70,000 sequences in the pooled amplicon series. Data were extracted from .fastq files where each sequence was associated with its sample identity (bar code is summarized in Table 1; Supplementary Table 2A–C summarize data for Ubiquitin and NEDD8, rounds 1–3, low and high stringency washes, and included non-amplified and amplified peptide phage, plus the parental library). In total, there were 40,451 total peptide sequences that were acquired from a theoretical total of approximately 70,000 sequences using the Roche 454 conjugation and amplification platform. The methodology is not confined to this sequencing platform as *Illumina* has also been used in phage display (Matochko et al., 2012). Deviations from the maximum using the Roche 454 platform relate to the presumed error in the calibration of the DNA concentrations to achieve approximately one oligonucleotide per bead coupled to the error in bead loss during the multiple washing steps. Of these 40,451 sequences, over 13,000 sequences were acquired from all rounds and washes using the Ubiquitin bar codes, and over 10,000 sequences were acquired from all rounds and washes using the NEDD8 bar codes (summarized in Supplementary Table 2C).

The output file containing the peptide sequences and the number of reads as a function of the bar codes was sorted as summarized in Supplementary Table 2A with a focused example in Figure 4. The peptide with the highest number of sequences identified (Figure 4, row 2, column C) was FIPAQLHFHWRS

TABLE 1 A summary of the sequences of the primers used for amplifying each phage peptide pool.

BAR-CODED PRIMER SETS FOR UBIQUITIN AND NEDD8		TIME; TARGET; ELUENT; ROUND
454-100-F	CGTATCGCCTCCCTCGCGCCATCAGAAAACCTCGAAAGCAAGCTGATAAACC	Parental 12 mer peptide library
454-101-F	CGTATCGCCTCCCTCGCGCCATCAGAATACCTCGAAAGCAAGCTGATAAACC	25 min wash; nedd8; amplified; round 1
454-102-F	CGTATCGCCTCCCTCGCGCCATCAGAAGACCTCGAAAGCAAGCTGATAAACC	25 min wash; nedd8; unamplified; round 1
454-103-F	CGTATCGCCTCCCTCGCGCCATCAGAACACCTCGAAAGCAAGCTGATAAACC	25 min wash nedd8 unamplified round 2
454-104-F	CGTATCGCCTCCCTCGCGCCATCAGATAACCTCGAAAGCAAGCTGATAAACC	25 min wash nedd unamplified round 3
454-105-F	CGTATCGCCTCCCTCGCGCCATCAGATTACCTCGAAAGCAAGCTGATAAACC	1 min wash ubiquitin unamplified round 1
454-106-F	CGTATCGCCTCCCTCGCGCCATCAGATGACCTCGAAAGCAAGCTGATAAACC	1 min wash nedd8 unamplified round 1
454-107-F	CGTATCGCCTCCCTCGCGCCATCAGATCACCTCGAAAGCAAGCTGATAAACC	25 min wash ubiquitin unamplified round 1
454-108-F	CGTATCGCCTCCCTCGCGCCATCAGAGAACCTCGAAAGCAAGCTGATAAACC	1 min wash ubiquitin unamplified round 2
454-109-F	CGTATCGCCTCCCTCGCGCCATCAGAGTACCTCGAAAGCAAGCTGATAAACC	1 min wash nedd unamplified round 2
454-110-F	CGTATCGCCTCCCTCGCGCCATCAGAGGACCTCGAAAGCAAGCTGATAAACC	25 min wash ubiquitin unamplified round 2
454-111-F	CGTATCGCCTCCCTCGCGCCATCAGAGCACCTCGAAAGCAAGCTGATAAACC	1 min wash ubiquitin unamplified round 3
454-112-F	CGTATCGCCTCCCTCGCGCCATCAGACAACCTCGAAAGCAAGCTGATAAACC	1 min wash nedd unamplified round 3
454-113-F	CGTATCGCCTCCCTCGCGCCATCAGACTACCTCGAAAGCAAGCTGATAAACC	25 min wash ubiquitin unamplified round 3
454-114-F	CGTATCGCCTCCCTCGCGCCATCAGACGACCTCGAAAGCAAGCTGATAAACC	25 min wash ubiquitin amplified round 1
454-115-F	CGTATCGCCTCCCTCGCGCCATCAGACCCACCTCGAAAGCAAGCTGATAAACC	1 min wash ubiquitin amplified round 1
454--Reverse	CTATGCGCCTTGCCAGCCCGCTCAGAAAGGATAGCAAGCCCAATAGGAACC	NA

The bar-coded primers (3 base bar code in purple) are highlighted as a function of time of wash (in minutes), the target (nedd8 or ubiquitin), whether the phage pool was amplified or not amplified, and the round of bio-panning (1–3).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	seq:	sum:	AAC	AAG	ACT	AAA	AGA	ATC	ACG	AGC	AGG	ACC	AGT	AAT	ATT	ATG	ACA	ATA
2	FIPAQLHFHWRS:	4717	0	2	2817	0	0	0	507	0	1366	0	0	0	1	1	3	1
3	YTDLTSDTLVRG:	4338	1	1	1	0	2	1	2	0	0	0	1	1	0	0	0	4
4	ALWPPNLHAWVP:	2377	306	90	0	98	172	67	212	20	62	282	380	234	54	95	123	166

FIGURE 4
Bar code stratification of peptide counts. An example of the peptide sequences enriched against ubiquitin taken from Supplementary Table 2A. The data summarize the top sequences that were detected after next-generation sequencing was applied to the combined phage pools. The data are segregated based on (i) the rounds (1, 2, or 3); (ii) whether short or long washes were used; and (iii) the number of peptide sequences extracted from each bar-coded sequencing pool (The bar-coded data can also be observed in Supplementary Table 2A). The bar codes are in Table 1.

with 4,717 reads. Columns D through S contain the bar codes that match a particular round (1–3), target protein (ubiquitin or NEDD8), and stringency method (short or long washes). The bar codes are listed in Table 1. For example, in Figure 4, column F (ACT bar code), row 2 (FIPAQLHFHWRS sequence) tabulates 2,817 peptide reads from non-amplified phage. This represents the ubiquitin target, round 3, with a high stringency wash. Reads from the ubiquitin target, rounds 1 and 2, with a high stringency wash, using non-amplified phage, are in Figure 4, column I (ATC bar code) and column L (AGG bar code), respectively. Interestingly, using the ubiquitin target, with a high stringency wash, but using amplified phage from round 1, 507 FIPAQLHFHWRS peptide reads were acquired (Figure 4, row 2, column J, ACG bar code). However, as we did not sequence any 'amplified' phage after round 1, we do not know whether this sequence would have been enriched in subsequent rounds 2 and 3 using amplified phage. Most other columns in row 2 that focus on the FIPAQLHFHWRS peptide showed: 0 reads (such as column H); 1 read (such as column P); 2 reads (such as column E); or 3 reads (such as column R). These data highlight the relative specificity of FIPAQLHFHWRS peptide enrichment using ubiquitin and high stringency washes.

An example of what we would call a "non-specific peptide" sequence (ALWPPNLHAWVP, Figure 4, row 4, column C, with 2,377 total reads) reveals that there are relatively high levels of reads in every condition, except for Figure 4, row 4, column F (ACT bar code), with zero reads. This is the bar code representing ubiquitin, round 3, high stringency wash which had over 99% of the reads containing FIPAQLHFHWRS (see Figure 5); therefore, it is consistent that ALWPPNLHAWVP would show 0 reads on this frame.

We next processed these data to remove possible background peptide-binding activity (Figure 5). First, we removed peptides with three or more histidine residues since such peptides were defined to be background-binding peptides to the nickel plate used in the selection toward the his-tagged ubiquitin or his-tagged NEDD8 control target (Figure 5A, in blue). In addition, any peptides that were present more than one time in the parental library were also removed. For example, some unique peptides

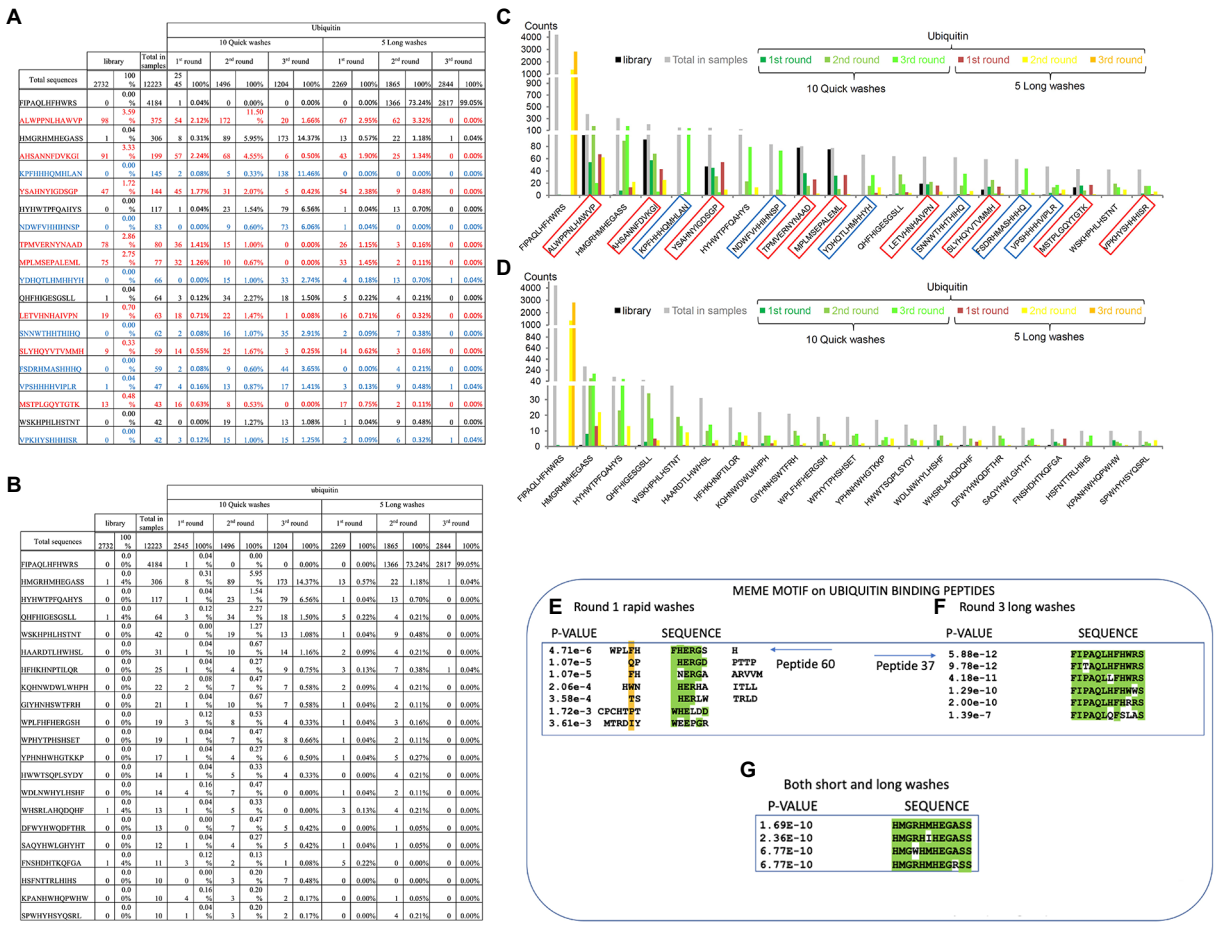


FIGURE 5
Specific peptide sequences enriched from Ubiquitin and NEDD8 peptide-library screens. **(A)** *Peptide sequences enriched against ubiquitin before removal of background peptides.* The data summarize the top sequences that were detected after next-generation sequencing was applied to the combined phage pools. The data are segregated based on **(I)** the rounds (1, 2, or 3); **(ii)** whether short or long washes were used; and **(iii)** the number of peptide sequences extracted from each bar-coded sequencing pool (The bar-coded data can also be observed in Supplementary Table 2A). Sequences in red reflect those peptides with a high representation in the parental library and are summarized as the total number of times the peptides were identified and as a percentage of the total sequences for that sequencing tag. For example, the sequence ALWPPNLHAWVP was identified 98 times out of 2,732 sequences with a parental library tag. The parental library bar code was AAA and this number of 98 counts of the sequence ALWPPNLHAWVP can be observed in Supplementary Table 2A. The sequences in red are defined as background due to their high abundance in the parental library and that they are not enriched during the selection process. The sequences in blue contain multiple histidine residues which were also removed to produce the final peptide lists. **(B)** *Peptide sequences enriched against ubiquitin after removal of background peptides.* The data summarize the top sequences that were detected after next-generation sequencing was applied to the combined phage pools and the background peptides (summarized in A) are removed. The data are segregated based on (i) the rounds (1, 2, or 3); (ii) whether short or long washes were used; and (iii) the number of peptide sequences extracted from each bar-coded sequencing pool. For example, from the long washes, the sequence FIPAQLHFHWRS was identified 0 times in round 1, 1,366 times in round 2, and 2,817 times in round 3. **(C,D)** The data from A and B are tabulated for visualization. The data are plotted as peptide count as a function of the indicated sequence. **(E–G)** The top 60 ubiquitin-selected peptides (from Supplementary Table 3) were evaluated using MEME (motif discovery; https://meme-suite.org/meme/tools/meme). As some examples: **(E)** Peptide 60 (motif collected from all peptides present in the Round 1 of Rapid washes) exhibited the core motif (W/F)HERG; **(F)** Peptide-37 (motif collected from all peptides present in Round 3 long washes) exhibited the core motif FIxAQLxFHWRS; and **(G)** Peptide 33 (from both short and long washes) exhibited the core motif HMGxHMHEGASS. In panels A-C, MEME derives a value of p based on parameters set in meme-suite (https://meme-suite.org/meme/tools/meme).

were found in higher copy numbers in the parental library itself such as ALWPPNLHAWVP which represented as much as 3.5% of the parental library (Figure 5A). Other peptides with high representation in the parental library include AHSANNFDVKGI, TPMVERNYNAAD, MPLMSEPALEML, or YSAHNYIGDSGP (Supplementary Tables 2A,B; Figure 5A). These peptides might be over-produced because they increase bacteriophage propagation in bacteria. As anticipated, some of these over-represented peptide sequences were specifically repressed in long washes and no significant enrichment observed using quick washes even if they remained abundant (Figure 5A). These same five peptides and other abundant peptides in the parental library are also observed in the NEDD8 screen (see Supplementary Table 2C).

Once such "background' peptides were removed, the final list of peptides acquired is for Ubiquitin and NEDD8, respectively. A tabulation of the most enriched Ubiquitin-specific peptides is

shown in Figure 5B. Notable was the peptide FIPAQLHFHWRS that showed enrichment in rounds 1, 2, and 3 of 0, 73, and 99%, respectively, of the total sequenced peptides from the long washes (Figure 5B). The data are quantified in graph form in Figures 5C,D. As control, the NEDD8 peptide enrichment (Supplementary Tables 2B vs. 2C) did not show significant similarity to the Ubiquitin peptide screen. One dominant peptide for NEDD8 binding was QHFHIGESGSLL with enrichment in each round seen only in the rapid washes (although binding was observed in the long washes; Supplementary Table 1). The peptide HYTHTHQYTYSM was also enriched in the long washes toward NEDD8. The ubiquitin-enriched peptide FIPAQLHFHWRS was observed in rounds 3 from NEDD8 selections but the penetrance was approximately two orders of magnitude lower than that observed when screening against Ubiquitin (Supplementary Table 1).

Validation of ubiquitin-binding peptides

The peptide sequences enriched toward the Ubiquitin target can be stratified in several ways, apart from simple abundance. The first approach we developed was using MEME to identify motifs that emerge between the filtered, enriched peptides.[1] The presence of motifs would suggest a binding specificity toward the target protein, as reported previously (Stevens et al., 2009; Mohtar et al., 2018). We focus here on two distinct groups of enriched peptides; (i) on peptides enriched toward ubiquitin in round-1_fast washes_not-amplified; (ii) on peptides enriched toward ubiquitin in both fast and slow washes_not-amplified; and (iii) on peptides enriched toward ubiquitin in round-3_long washes_not-amplified. These three states cover the three extremes in the peptide selection process; (i) lower affinities; (ii) intermediate affinities; and (iii) higher affinities, respectively. One peptide WPLFHFHERGSH that forms part of a cluster of peptides that derive a motif using MEME (Figure 5E) is relatively specific for enrichment in the fast washing stages (Figures 5B–D). Another peptide, FIPAQLHFHWRS, also forms a motif with additional peptides enriched in the screen with more stringent washes (Figure 5F). An example of another peptide that forms a top enriched peptide in either slow or fast washes toward Ubiquitin, HMGRHMHEGASS (peptide-33), also forms a consensus motif using MEME (Figure 5G).

Validation of peptide aptamers as tools to manipulate an *in vitro* reconstituted ubiquitination system

We next evaluated whether these synthetic peptides (Supplementary Table 3) were able to impact on *in vitro* ubiquitination reactions. The initial reconstituted system was composed of ubiquitin, E1, UBCH5a (E2), E3 (MDM2), the p53 substrate, and an ATP regeneration system. Sixty of the most abundant synthetic peptides (Supplementary Table 3), were pooled into groups of 10 so that the molar ratio of any one peptide:Ubiquitin was 3:1. The reaction was balanced with a DMSO control and six reactions revealed two pools of peptide with significant inhibition of ubiquitination of p53 (data not shown).

Each 10-plex peptide pool was then divided into individual peptides so we could identify the bioactive peptide mediating inhibition of ubiquitination (representative data in Figure 2A). The most potent, reproducible inhibitors of ubiquitination were synthetic peptides 12 and 37 (Figure 2A, lanes 2, 4, and 7 vs. lane 1). Interestingly, peptide 37 has the sequence FIPAQLHFHWRS, which was the most highly significantly enriched in stringent screens (Figure 5B). Peptide 44 was inactive as ubiquitin inhibitors (Figure 2A, lane 5); this was an important negative control as it showed that a ubiquitin-binding peptide aptamer does not necessarily impact functionally on the ubiquitin transfer cascade.

In order to determine if the ubiquitination inhibition was linked to the E3 (MDM2) we evaluated these peptides in ubiquitination reactions using an unrelated ubiquitin ligase, CHIP (Figure 2B). Peptides 12 and 37 remained active in mediating inhibition of substrate ubiquitination (Figure 2B, lanes 4 and 5). This suggests that the peptides might be inhibiting at an earlier stage; (i) either E1 ubiquitin conjugation; (ii) ubiquitin transfer to E2; and/or (iii) release of Ubiquitin from E2 to substrate. As a control peptide 44 was inactive (Figure 2B, lane 3 vs. lane 2) as an inhibitor, indicating that synthetic peptides are not generally toxic to CHIP or the ubiquitination cascade in general. Future work will be carried out to dissect at what stage the peptides inhibit ubiquitination for example at the E2 discharge step.

In order to further validate the Ubiquitin-binding peptides, they were titrated into reactions above and below stoichiometric levels to Ubiquitin (Supplementary Figure 1A). The titration of both peptide-12 and pepide-37 reveals that near-complete inhibition of MDM2-dependent ubiquitination is observed at a 2:1 molar ratio of peptide:ubiquitin (Supplementary Figure 1A, lanes 3 and 6 vs. 2). However, at a ratio of peptide:ubiquitin of 0.2:1, there was no inhibition (Supplementary Figure 1A, lanes 4 and 7 vs. 2). These data suggest that inhibition by both peptides is not catalytic but involves direct sequestration of ubiquitin from the ubiquitin transfer reaction. If, for example, the peptides were active at a molar ratio of peptide:ubiquitin of 1:100, this might suggest that the peptide was inhibiting the E1-ubiquitin conjugated stage because E1 is present at 100-fold lower molar levels to Ubiquitin. The peptides were also tested against a more non-specific ubiquitination reaction involving RING domain only from MDM2 which lacks the p53 docking sites. Using this ubiquitination reaction, it is interesting to note that only peptide-37 inhibits the reaction (Supplementary Figure 1B, lanes 4 vs. 1). This could suggest that peptides 12 and 37 bind to Ubiquitin by different mechanisms and that the RING-only ubiquitination reaction does not involve a perhaps more complex

1 https://meme-suite.org/meme/tools/meme

protein–protein interaction as does CHIP or MDM2, both of which are inhibited by peptide-12 (Figure 2). However, these data can also be interpreted that peptide-37 has a higher affinity than peptide-12 which is why it inhibits RING-only ubiquitination.

Orthogonal validation of ubiquitin-binding peptide-37 and peptide-12 using ELISA, NMR, and reverse phase chromatography

We next compared the relative binding of peptide-37 and peptide-12 to ubiquitin by ELISA using the same 'sandwich' format as the original screen which involved his-tagged ubiquitin added to the solid phase and peptide in the mobile phase. There is a weakness in the phage peptide screening concept that aims to acquire a useful tool. The phage peptide screen selects peptides binding when peptides are fused C-terminally to gIII protein of M13 and then the peptide has a free N-terminus after cleavage of the leader peptide (Figure 1). It is possible therefore that the peptide will only bind ubiquitin when it is fused to gIII because the gIII influences peptide conformation. This is not necessarily the case, as the bioactive peptides used in Ubiquitination assays had a, N-terminal tag and a C-terminal amide (Figure 2).

Nevertheless, we tested the binding of peptide-37 and peptide-12 in an ELISA format in which the peptides have a biotin-tag on the C-terminus of the peptide (Figure 2C). We realize that this C-terminal Biotin-tag does not mimic a gIII fusion, but it provides a useful tag in the event that is it active. The C-terminal biotin tag might mimic the peptide-phage in that the peptide is fused C-terminally to the tag of interest (Figure 1A). However, the final streptavidin-HRP:peptide complex might also deter binding by ELISA. Thus, it cannot be guaranteed that a peptide containing a tag on any orientation could even bind ubiquitin because of steric effects. Representative peptide-37 binding with C-terminal tag is shown in Figures 2C,D. In addition, as NEDD8 also has a similar structure to Ubiquitin, we tested peptide-37 binding by titration also in ELISA format and the data show that peptide-37 also binds NEDD8 but to a lower extent than Ubiquitin. For example, at lower peptide levels (12.5 ng) of peptide-37, corrected ECL units for Ubiquitin are 10,450 and for NEDD8 are 5,695 (Figure 2C). These data also highlight potential methodological problems in testing peptide binding by ELISA as an orthogonal assay. In this case, peptide-37 titration alone yields a relatively high background (Figure 2C, up to 47,545 avg. ECL units), while Ubiquitin alone also yields a relatively high background (up to 21,310 avg. ECL units). Peptide-12 did not bind by ELISA under these same conditions (data not shown). Similar results were obtained with a time course at fixed levels of peptide-37 (100 ng, Figure 2D). In this case, for example, by 60 min of binding, we detected 28,405 corrected ECL units for Ubiquitin and 30,173 corrected ECL units for NEDD8. Again, the background of Ubiquitin alone is relatively high at an average of 24,360 ECL Units (Figure 2D). Presumably, this indicates that there is some binding of the Ubiquitin to the Streptavidin peroxidase.

We finally attempted to employ other biophysical approaches to demonstrate whether peptides 37 or 12 bound to ubiquitin. We first attempted to define mode of binding of the peptides through analysis of peptic fragments using our previously published methods for measuring changes in ligand binding or mutational effects on MDM2 and CHIP E3 ubiquitin ligases using hydrogen-deuterium exchange mass spectrometry (Hernychova et al., 2013; Narayan et al., 2015). However, we found that ubiquitin was highly resistant to proteolysis by pepsin under these conditions (data not shown) and we were unable to acquire a comprehensive deuteration map of peptic ubiquitin peptides. As such, we sought to probe for global changes in overall deuteration of native ubiquitin in the absence and presence of each peptide. Deuterium exchange of ligand-free ubiquitin protein or in complex with peptide-12 or peptide-37 was initiated by dilution of ubiquitin into deuterated water. The molar ratios between ubiquitin and peptide-12 or peptide-37 were 1:2 and 1:5. Deuterium exchange was carried out for 1 min (Supplementary Figures 2A–F). The isotopic distribution of non-deuterated ubiquitin deconvoluted peak (calculated using *MagTran* (Zhang and Marshall, 1998) from charge states $z = 7–12$ and retention time 26.8–27.6 min) with an nominal average mass of 8564.85 Da (Supplementary Figure 2A). One minute of deuteration made a shift in global deuteration of ubiquitin resulted in an average mass of 8577.16 Da (Supplementary Figure 2B), together indicating an increasing in 12.31 deuterons. A reduction in the average mass of ubiquitin can be observed with increasing ratio of peptide-12:ubiquitin from 1:2 to 1:5 (Supplementary Figures 2C,D). Peptide-37 resulted in marginally elevated deuteration at peptide-37:ubiquitin ratio of 1:2, and an attenuated deuteration at a 1:5 ratio (Supplementary Figures 2E,F). These data together suggest that the peptides can bind to and have specific effects on the global deuteration of native ubiquitin. However, the data do not conclusively demonstrate a distinct and/or specific mode of action in the binding of peptide-12 or peptide-37 to Ubiquitin.

As such, NMR was used to determine whether we could detect binding of peptide-12 and peptide-37. The 2D $^{1}H–^{15}N$ HSQC spectrum collected for human ubiquitin corresponds fully with the literature and sequence-specific assignments were performed using the BioMagnetic Resonance Databank (BMRB 4493) and previous 2D and 3D NMR experiments performed on $^{13}C^{15}N$-uniformly labeled samples (Supplementary Figure 3). The recorded ^{15}N relaxation data (R_1, R_2, and $^{1}H–^{15}N$ NOE) acquired at magnetic field 11.7 T are in line with available data previously recorded (BMRB: 6470 BMRB 4493). The addition of the peptide-12 to the ubiquitin sample resulted in increased intensity of several cross peaks. Subsequently, characterizing the peptide-12 as selectively bound to ubiquitin along with substantial changes in peak heights detected for Thr9 and Gly75 residues (inserts on Supplementary Figure 3). A more in-depth analysis showed that these substantial changes

include the β-turn located between first and second β-sheets (Leu8 – Thr12) and C-terminal fragment of the 3D ubiquitin structure (Arg72–Gly75; Figure 6). It is interesting that we did not detect any changes in the chemical shifts for amide groups of the ubiquitin (Supplementary Figure 3).

Analysis of the ^{15}N relaxation data suggests an increase in the intensity of molecular dynamic processes at high frequency timescales (ns–ps) for residues located in C-terminal part of the protein, which is reflected in lower NOE values for residues starting from Arg72. The detected weak minimum for ^{15}N R_2 and NOE suggests an increasing amplitude of such a type of motions also in the fragment located between Thr9 – Thr12 (Supplementary Figure 4). The same residues exhibited changes in peak-height after peptide binding. On the other hand, there are two fragments of the protein that exhibited increased intensity of low-frequency motions (μs–ms) after binding peptide-12. Namely, the structural loop Ile23–Asn25 and the fragment around Ala46. The existence of the slow chemical exchange motions for the Ala46 was confirmed by the analysis of the R_1 and R_2 products (Supplementary Figure 5). That observation was further confirmed by analysis of the relaxation data using the SDM approach, revealing higher level of spectral density function observed for Ala46 at low and middle frequencies ($\omega = 0$ and ω_N; Supplementary Figure 6).

The addition of peptide-12 to ubiquitin resulted in a decrease in the intensity of molecular dynamic processes at low frequency time-frames (10^{-3} … 10^{-6} s) for the β-turn around Ala46 (Supplementary Figures 4–6). Another interesting observation involves the changes of the height of cross-peaks in ^{1}H-^{15}N HSQC spectra for the several residues grouped around Thr9 and Gly75 (Supplementary Figure 3). One interpretation of these data is that the increase in the peak intensities can be explained by binding the peptide-12 to selected residues in a fast-exchange regime, where the lifetime of the bound conformation is shorter than the lifetime of the unbounded (free) ubiquitin. The process of transition between the two states can cause the decrease in the exchange of amide protons with water for selected residues. The surface representation of the 3D structure of human ubiquitin reveals that all residues noted as possible binding sites for the peptides are located on one side of the protein (Figure 6). Our

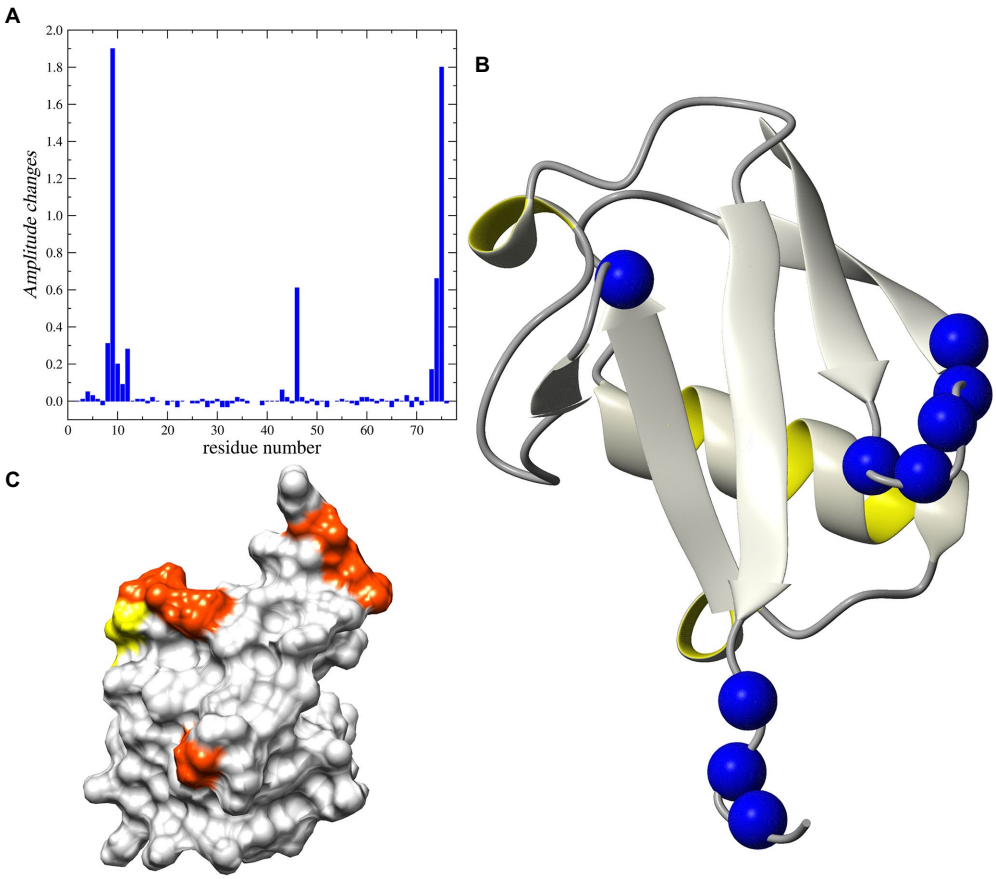


FIGURE 6
NMR analysis of Peptide-12 binding to ubiquitin. **(A)** Relative changes in intensity of ^{1}H-^{15}N HSQC cross-peaks induced by the addition of the peptide-12. **(B)** Ribbon representation of the 3D structure of human ubiquitin. The ^{15}N nuclei in the residues corresponding to the increased peak heights are shown as blue balls. **(C)** The surface presentation of human ubiquitin. The positions of the residues demonstrating the increased peak intensity are highlighted in orange and yellow.

observations suggest that the peptide-12 interacts with ubiquitin specifically in the regions of loop Leu8–Thr12 and C-terminus (Leu73–Gly75). The peptide binding leads to the decrease of slow motion intensity for Ala46 in, most probably, an allosteric manner.

By contrast to peptide-12, the addition of peptide-37 to the ubiquitin sample was immediately detected in the NMR data. The signals in $^{1}H–^{15}N$ HSQC spectrum were characterized by increase of linewidth and decrease of signal-to-noise ratio (Supplementary Figure 7). The analysis of observed changes suggests the existence of a strong oligomerization process of human ubiquitin after incubation with the peptide-37. Extremely low signal-to-noise ratio excludes the possibility to collect any high-quality ^{15}N relaxation data. Nevertheless, our estimations suggest that overall tumbling time (τ_m) increased twice (at least), which corresponded with the effective molecular mass of ubiquitin oligomers around 25–30 kDa.

We finally used reverse-phase HPLC (C18) as a methodology to determine whether Peptide-37 induces an aggregation-like property to ubiquitin as suggested by NMR. Reverse-phase HPLC has been used previously to identify conformational changes in a target protein. For example, insulin (mass of approximately 5,300 Da) from different species can be separated by reverse-phase HPLC when differing by only a serine and threonine residue (Rivier and McClintock, 1983). Interleukin mutant retention times on reverse phase-HPLC have been correlated to structural changes (Kunitani et al., 1986). The molar ratios between ubiquitin and peptide-12 or peptide-37 were 1:2 and 1:5. The total ion chromatograms (TIC) of ubiquitin-peptide complexes are shown in Figures 7A–E. TIC highlighted that three of the most abundant ion signals corresponded to the peptides (peptide-12 had a retention time (RT) of 20.47–21.51 min and peptide-37 had a RT of 25.48–28.86 min). The other two ion signals represent co-elution of ubiquitin and the respective peptides (RT 25.48–28.86 min and 33.34–34.34 min; mass spectral data not shown). The retention time of ubiquitin either alone or in complex with both peptides remains the same (RT 24.45–28.70 min). However, the absolute peak intensities of the eluted ubiquitin are different. In particular, the intensity of the ubiquitin signal is compressed when it is in complex with peptide-37 at a 1:2 or 1:5 molar ratio of protein: peptide (Figures 7D,E vs. Figure 7A), which is characteristic of protein loss during chromatography, perhaps due to 'aggregation'. The impact of peptide-37 on apparent aggregation or oligomerization of ubiquitin during chromatography is in agreement with NMR data. A model of peptide-12 and peptide-37 docking using molecular dynamics simulation (MDS), hydrogen bond (H-Bond) interaction map, and structural changes and amino acid fluctuations identified for the Ubiquitin-peptide-12/peptide-37 complex using RMSD (root-mean-square deviation) and RMSF (root-mean-square fluctuation) calculations are in Figure 8, which suggests different modes of binding.

Discussion

Discovering protein–protein interactions is an important goal in life sciences (Morelli and Hupp, 2012). The number of interacting proteins for any given eukaryotic protein can be relatively large. Classic examples include the ATM/ATR kinases which are reported to have over 700 substrates (Matsuoka et al., 2007). It is becoming apparent that one of the main drivers of such a diversity of protein–protein interactions are linear motifs that form weak, but specific protein–protein interactions important in cell signaling. Linear motifs form vast numbers of combinatorial protein–protein interaction interfaces that offer a rich source of signaling diversity (Tompa et al., 2014). Thus, advances in approaches for discovering functional linear peptide motifs is important for the protein science field.

Peptide combinatorial libraries could provide a powerful platform for high throughout linear peptide motif discovery. The power of massively diverse combinatorial peptide phage display lies in its ability for an efficient and rapid identification of peptide ligands from a phage population displaying millions of diverse surface peptides. In addition, the reliance on *in vitro* screening is also an advantage in that the bait conformation, buffers, and selection conditions can be applied selectively. *In vitro* selection previously conditions identified MDM2 protein interactors; pre-incubating MDM2 with its ligands including zinc or RNA can give rise to the isolation of distinct linear peptide motifs from phage-peptide populations (Shimizu et al., 2002; Burch et al., 2004). However, a major disadvantage of the phage-peptide pool has been the relatively low throughout in peptide identification by DNA sequencing. Typically, only dozens of peptides are identified from sequencing using standard Sanger sequencing platforms (Murray et al., 2007). Next-generation sequencing methods could create a platform in which the diversity of the peptide-phage library can be fully exploited through the sequencing of thousands or millions of peptide inserts. In this report, we set up a next-generation sequencing platform on a peptide-phage library screen that can sequence using Roche454J platforms up to 70,000 inserts from one reaction (Figure 1). Our data demonstrate that using next-generation DNA sequencing, peptide library diversity can be defined, background (non-specific) peptides can be identified, and enrichment can be quantified isolating peptides with a high degree of specificity. Illumina sequencing of peptide-phage populations was used previously to our study (Matochko et al., 2012), thus providing greater sequence depth to the platform we use in this report.

E3 ubiquitin ligases can select and ubiquitinate client proteins by complex mechanisms that have not been fully dissected. For example, allostery in the CHIP E3 ubiquitin ligase was revealed by biophysical characterization including hydrogen-deuterium exchange mass spectrometry that demonstrated conformational-inhibition-signals extend from the TPR-domain to the U-box (Narayan et al., 2015). The MDM2 E3 ubiquitin ligase exhibits complex allosteric effects that exploit multiple protein–protein docking sites on the p53 substrate and the co-factor, E2 (Wallace

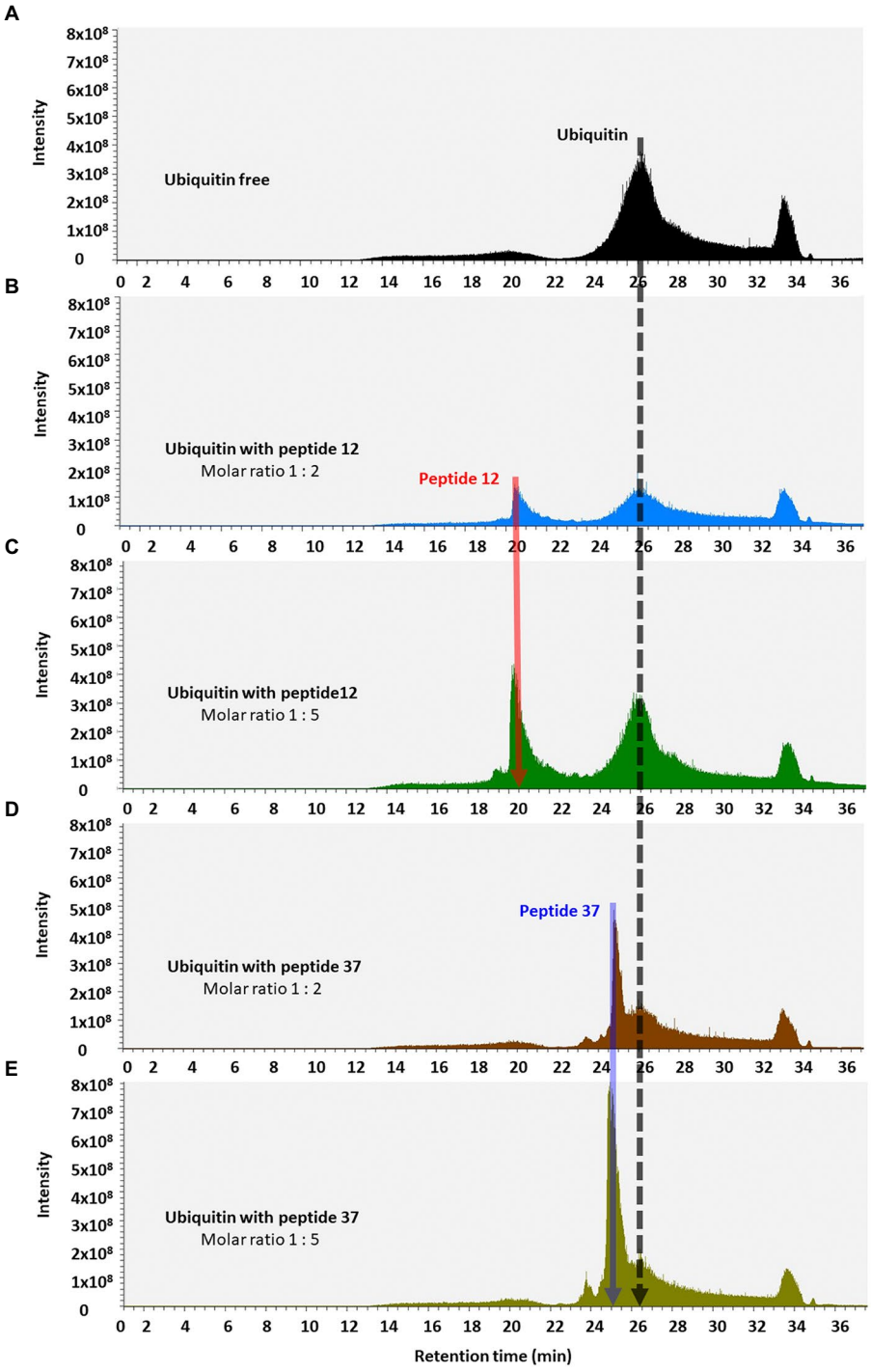


FIGURE 7
Total ion chromatograms of ubiquitin-peptide complexes. Ubiquitin was incubated alone **(A)** or with the indicated peptides **(B–E)**, diluted in D_2O with 1% DMSO, and separated on a reverse phase C18 column according to the approach described in Methods. Total ion chromatogram was plotted as the relative abundance of the signal observed at a chosen m/z value as a function of RT. Material eluted at highlighted position and RT were confirmed by MS/MS analysis (Orbitrap Elite mass spectrometer) as the peptide-12 (red line, RT 20.25–22.00 min); the peptide-37 (blue line, RT 25.28–25.67 min), and ubiquitin (black line, RT 24.45–28.70 min).

et al., 2006; Wawrzynow et al., 2009; Fraser et al., 2015). The stoichiometry and dynamics of how the E1–E2–E3 conjugation system operates is not well understood; for example, in the case of MDM2 it is not clear if substrate docking involves the formation of a dimeric MDM2 structure (Robson et al., 2012). How ubiquitin itself is detected, orientated, and conjugated is not completely defined and it is likely to be highly combinatorial (Ernst et al., 2013). Conjugation of mono-ubiquitin to a transcription factor

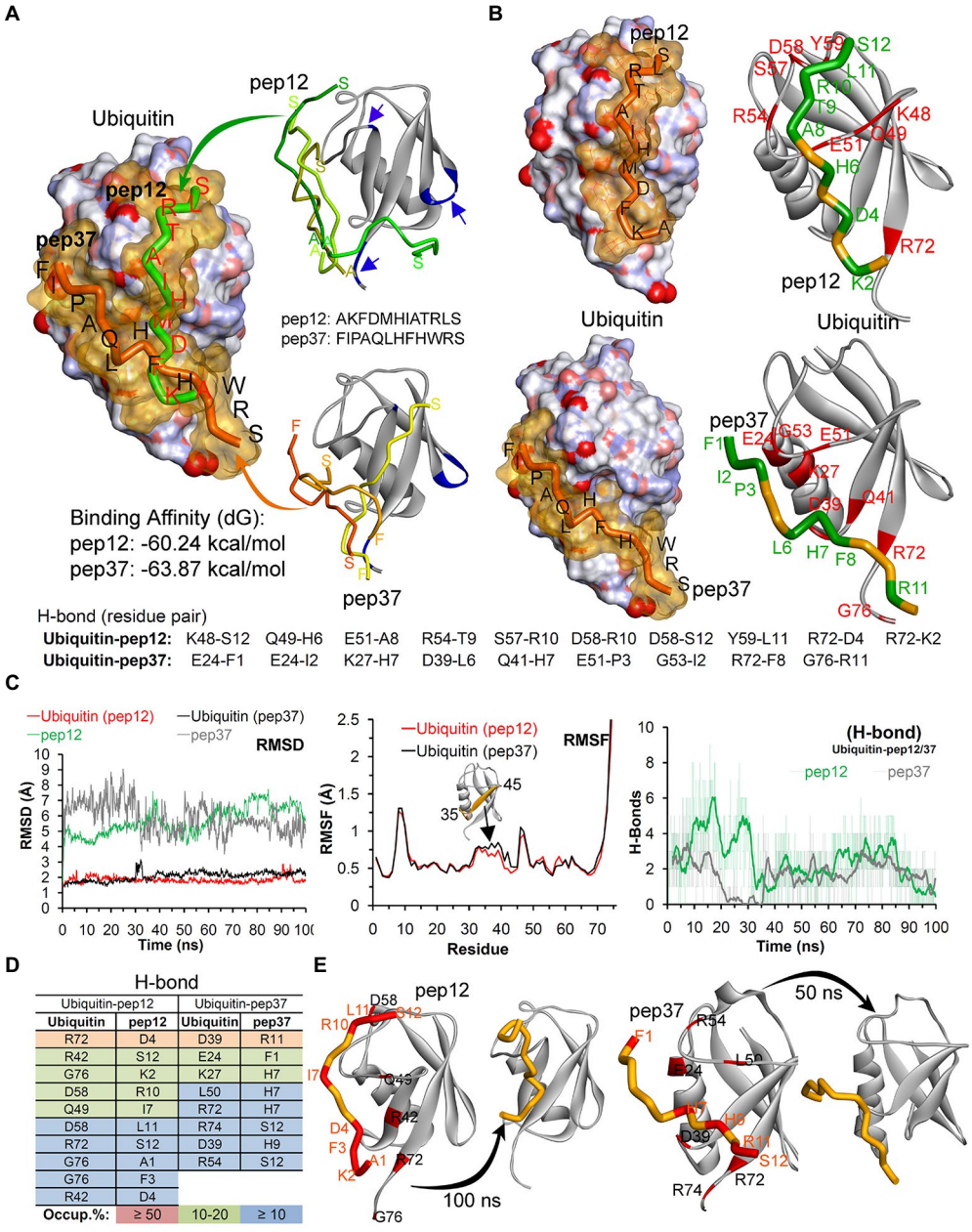


FIGURE 8
Model of peptide-12 and peptide-37 mediated inhibition of ubiquitination. **(A)** Peptides docked with ubiquitin having highest binding affinity are shown (blue arrows/region represent peptide-binding region traced in our NMR study). Although we were unable to determine whether the ubiquitin-binding peptides-12 or 37 bound to different or identical regions on ubiquitin, we analyzed potential conformations of peptide-12 and peptide-37 with ubiquitin by molecular dynamics simulation (MDS) and docking approaches. **(B)** The hydrogen bond (H-Bond) interaction map between amino acids of ubiquitin and peptides. **(C)** Structural changes and amino acid fluctuations identified for the Ubiquitin-peptide-12/peptide-37 complex using RMSD (root-mean-square deviation) and RMSF (root-mean-square fluctuation) calculations. **(D)** The stable intermolecular H-bonds between ubiquitin-peptide-12/peptide-37 over 100 ns of MDS, and residues pair with occupancy ≥5% (from 100ns) are listed. **(E)** Structural dynamics of peptide-12 and peptide-37 observed during MDS. *In this figure, the ubiquitin crystal structure PDB id.: 1UBQ (60) in apo/inactive form was used, and ubiquitin is represented as atom charge surface/solid ribbon and pep12/pep37 are presented as tubes. Ubiquitin has a compact and rigid structure that ends in a flexible C-terminal tail. This rigid conformation allows only certain residues to have limited flexibility to permit a large diversity of protein–protein interactions (Komander and Rape, 2012; Ernst et al., 2013). It contains a hydrophobic core formed divided in three areas: the first one around Ile44, Leu8, Val70 and His68 is involved in most of the interactions with ubiquitin-binding proteins (UBP) and the proteasome (Sloper-Mould et al., 2001; Dikic et al., 2009); the second patch, around the residues Ile36, Leu71 and Leu73 is important for recognition by some HECT E3 ligases (Kamadurai et al., 2009). Finally, the area around Gln2, Phe4 and Thr14 interacts with certain proteins including some DUBs (Hu et al., 2002).

can present a face of ubiquitin that can interact with DNA (Landre et al., 2017). As such, we applied our next-generation peptide-screening platform to human ubiquitin, a core substrate in the ubiquitin conjugation reaction, to determine if peptides can be isolated that bind to and impact on ubiquitination. We identified using deep-sequencing a small number of specific peptide aptamers that were useful as tools to begin to explore steric constraints in the ubiquitin conjugation reaction. For example, two peptides we identified were studied (named 12 and 37). Analysis of these two peptides revealed that, although they both inhibit ubiquitination at similar molar ratios (Supplementary Figure 1A), they differ in these orthogonal assays; (i) ELISA's show difference in binding to ubiquitin; (ii) RING domain-only ubiquitination is specific for peptide 37; (iii) HPLC analysis of peptide-ubiquitin complexes reveal a difference in ubiquitin peak intensity; and (iv) NMR suggests different binding modes with peptide-37 stimulating apparent aggregation. Further biophysical studies would be required to determine if the peptide-12 and peptide-37 bind by different mechanisms to Ubiquitin or whether these differences reflect changes in affinity to the target. Nevertheless, these orthogonal assays are presented to provide readers, of this methodological study, some ideas on orthogonal biochemical assays that can be used to test synthetic peptides, derived from phage display, when the peptides are not fused to gIII of M13 phage.

Previous research has highlighted the conformational dynamism of the ubiquitin conjugation cascade and the importance of weak ubiquitin interfaces (Haririnia et al., 2008). The E2 ubiquitin transfer proteins use a reactive cysteine that forms a covalent adduct with ubiquitin through a thioester bond. The stability of the E2-ubiquitin conjugate is maintained by conserved asparagine that stabilizes the oxyanion transition state (Wenzel et al., 2011). A low-affinity protein–protein interaction between ubiquitin and E2 maintains steric occlusion of the active site to permit nucleophile attack of the thioester bond and orientates the incoming ubiquitin to catalyze ubiquitin chain linkage (Saha et al., 2011). A small chemical molecule (CC0651) was identified that binds the E2 conjugation protein CDC42 and this molecule blocks ubiquitination transfer by virtue of trapping a weak interaction between ubiquitin and the E2 (CDC42) ubiquitin-binding site (Huang et al., 2014). The data suggested that stabilization of other weak interactions between ubiquitin and ubiquitin conjugation enzymes by small molecules could drive drug-discovery in the UPS system. Indeed, it is interesting that one of the peptides identified in our study (peptide-37) also attenuated ubiquitin-E2 hydrolysis (data not shown) suggesting that this thioester conjugate stability can be targeted by ligands that bind E2 or ubiquitin. We have also identified natural product extracts that we also show can mimic the peptides and impact on the E2-ubiquitin discharge (data not shown) and further suggests that the E2-ubiquitin complex can be manipulated by several types of bio-chemical interventions.

A final utility of the deep sequencing of linear peptide libraries is the opportunity to produce a consensus site that drives discovery of authentic and novel protein–protein interactions. The use of motif search engines can result in the identification of novel protein–protein interfaces that can be validated using emerging robust platforms for *in situ* measurement of endogenous protein–protein interactions (Dinkel et al., 2014; Gibson et al., 2015). Using the consensus site of peptide-37 (Figure 5), we can similarly identify several proteins in the human proteome with this motif (data not shown) and some of these targets are currently being validated as novel ubiquitin-binding adaptor proteins (unpublished data). Thus, deep DNA sequencing of combinatorial peptide libraries can therefore be used to (i) identify peptide aptamers to a target protein for use as tools to manipulate and study protein–protein interaction dynamics and (ii) discover new protein–protein interfaces of potential biological relevance.

Conclusion

This study provides a methodological blueprint for using next-generation sequencing of peptide-phage library pools screened against a target protein. In addition, orthogonal assays that are presented to characterize enriched peptides include (i) synthesis of peptides with an N-terminal or C-terminal Biotin tag to test activity in binding to target protein by ELISA; (ii) testing peptide activity in an enzymatic assay; (iii) measuring the effects of peptides on target protein binding by NMR; (iv) examination of the effects of peptides on target protein deuteration; and (v) possible mode of binding of peptides onto target using MDS.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ML: investigation and validation. FL: data curation, investigation, and writing—original draft. MG-M: investigation and data curation. A-SH, LH, and AK: methodology. LW: writing—review and editing. PM, BV, RK, IZ, and PJ: resources. SR-M: conceptualization and writing—review and editing. KB: supervision. TH: funding acquisition, supervision, and conceptualization. UK: software, visualization, and supervision. ZT: investigation. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the European Regional Development Fund—Project ENOCH (No. CZ.02.1.01/0.0/0.0

/16_019/0000868), by the Ministry of Health Development of Research Organization, MH CZ—DRO (MMCI, 00209805). BV, PM, LH, and RK were supported by Czech Science Foundation (No: 19-03796S). MG-M was funded by Scottish Power. A-SH was funded by Cancer Research UK (C483/A10706). LW was supported by BBSRC (BB/K011278/1). UK was supported by grant 2020/36/C/NZ2/00108 from The National Science Centre, Poland (Narodowe Centrum Nauki, Krakow, Poland). The International Centre for Cancer Vaccine Science (ICCVS) project is carried out at the University of Gdańsk within the International Research Agendas programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund (FL, ML, and UK).

Acknowledgments

The authors would like to thank the PL-Grid Infrastructure, Poland, for providing their hardware and software resources.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.875556/full#supplementary-material

References

Bailly, A., Perrin, A., Bou Malhab, L. J., Pion, E., Larance, M., Nagala, M., et al. (2016). The NEDD8 inhibitor MLN4924 increases the size of the nucleolus and activates p53 through the ribosomal-Mdm2 pathway. *Oncogene* 35, 415–426. doi: 10.1038/onc.2015.104

Bialas, J., Groettrup, M., and Aichem, A. (2015). Conjugation of the ubiquitin activating enzyme UBE1 with the ubiquitin-like modifier FAT10 targets it for proteasomal degradation. *PLoS One* 10:e0120329. doi: 10.1371/journal.pone.0120329

Bottger, V., Bottger, A., Howard, S. F., Picksley, S. M., Chene, P., Garcia-Echeverria, C., et al. (1996). Identification of novel mdm2 binding peptides by phage display. *Oncogene* 13, 2141–2147. PMID: 8950981

Burch, L., Shimizu, H., Smith, A., Patterson, C., and Hupp, T. R. (2004). Expansion of protein interaction maps by phage peptide display using MDM2 as a prototypical conformationally flexible target protein. *J. Mol. Biol.* 337, 129–145. doi: 10.1016/j.jmb.2004.01.017

Ceccarelli, D. F., Tang, X., Pelletier, B., Orlicky, S., Xie, W., Plantevin, V., et al. (2011). An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cells* 145, 1075–1087. doi: 10.1016/j.cell.2011.05.039

Ciechanover, A. (2015). The unravelling of the ubiquitin system. *Nat. Rev. Mol. Cell Biol.* 16, 322–324. doi: 10.1038/nrm3982

Das, R., Liang, Y. H., Mariano, J., Li, J., Huang, T., King, A., et al. (2013). Allosteric regulation of E2:E3 interactions promote a processive ubiquitination machine. *EMBO J.* 32, 2504–2516. doi: 10.1038/emboj.2013.174

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293. doi: 10.1007/BF00197809

Derda, R., Tang, S. K., Li, S. C., Ng, S., Matochko, W., and Jafari, M. R. (2011). Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* 16, 1776–1803. doi: 10.3390/molecules16021776

Derda, R., Tang, S. K., and Whitesides, G. M. (2010). Uniform amplification of phage with different growth characteristics in individual compartments consisting of monodisperse droplets. *Angew. Chem. Int. Ed. Engl.* 49, 5301–5304. doi: 10.1002/anie.201001143

Deshaies, R. J., and Joazeiro, C. A. (2009). RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* 78, 399–434. doi: 10.1146/annurev.biochem.78.101807.093809

Dias-Neto, E., Nunes, D. N., Giordano, R. J., Sun, J., Botz, G. H., Yang, K., et al. (2009). Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PLoS One* 4:e8338. doi: 10.1371/journal.pone.0008338

Dikic, I., Wakatsuki, S., and Walters, K. J. (2009). Ubiquitin-binding domains—from structures to functions. *Nat. Rev. Mol. Cell Biol.* 10, 659–671. doi: 10.1038/nrm2767

Dinkel, H., Van Roey, K., Michael, S., Davey, N. E., Weatheritt, R. J., Born, D., et al. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 42, D259–D266. doi: 10.1093/nar/gkt1047

Ernst, A., Avvakumov, G., Tong, J., Fan, Y., Zhao, Y., Alberts, P., et al. (2013). A strategy for modulation of enzymes in the ubiquitin system. *Science* 339, 590–595. doi: 10.1126/science.1230161

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., et al. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* 33, 5984–6003. doi: 10.1021/bi00185a040

Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995). Spectral density function mapping using 15N relaxation data exclusively. *J. Biomol. NMR* 6, 153–162. doi: 10.1007/BF00211779

Fraser, J. A., Worrall, E. G., Lin, Y., Landre, V., Pettersson, S., Blackburn, E., et al. (2015). Phosphomimetic mutation of the N-terminal lid of MDM2 enhances the polyubiquitination of p53 through stimulation of E2-ubiquitin thioester hydrolysis. *J. Mol. Biol.* 427, 1728–1747. doi: 10.1016/j.jmb.2014.12.011

Fuchs, S., Nguyen, H., Phan, T., Burton, M. F., Nieto, L., Leeuwen, I., et al. (2013). Proline primed helix length as a modulator of the nuclear receptor-coactivator interaction. *J. Am. Chem. Soc.* 135, 4364–4371. doi: 10.1021/ja311748r

Gibson, T. J., Dinkel, H., Van Roey, K., and Diella, F. (2015). Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun. Signal* 13:42. doi: 10.1186/s12964-015-0121-y

Gong, P., Canaan, A., Wang, B., Leventhal, J., Snyder, A., Nair, V., et al. (2010). The ubiquitin-like protein FAT10 mediates NF-kappaB activation. *J. Am. Soc. Nephrol.* 21, 316–326. doi: 10.1681/ASN.2009050479

Haririnia, A., Verma, R., Purohit, N., Twarog, M. Z., Deshaies, R. J., Bolon, D., et al. (2008). Mutations in the hydrophobic core of ubiquitin differentially affect its recognition by receptor proteins. *J. Mol. Biol.* 375, 979–996. doi: 10.1016/j.jmb.2007.11.016

Heride, C., Urbe, S., and Clague, M. J. (2014). Ubiquitin code assembly and disassembly. *Curr. Biol.* 24, R215–R220. doi: 10.1016/j.cub.2014.02.002

Hernychova, L., Man, P., Verma, C., Nicholson, J., Sharma, C. A., Ruckova, E., et al. (2013). Identification of a second Nutlin-3 responsive interaction site in the N-terminal domain of MDM2 using hydrogen/deuterium exchange mass spectrometry. *Proteomics* 13, 2512–2525. doi: 10.1002/pmic.201300029

Hu, Z., and Crews, C. M. (2021). Recent developments in PROTAC-mediated protein degradation: from bench to clinic. *Chembiochem*. 23:e202100270. doi: 10.1002/cbic.202100270

Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J. W., et al. (2002). Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cells* 111, 1041–1054. doi: 10.1016/S0092-8674(02)01199-6

Huang, H., Ceccarelli, D. F., Orlicky, S., St-Cyr, D. J., Ziemba, A., Garg, P., et al. (2014). E2 enzyme inhibition by stabilization of a low-affinity interface with ubiquitin. *Nat. Chem. Biol.* 10, 156–163. doi: 10.1038/nchembio.1412

Kamadurai, H. B., Souphron, J., Scott, D. C., Duda, D. M., Miller, D. J., Stringer, D., et al. (2009). Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol. Cell* 36, 1095–1102. doi: 10.1016/j.molcel.2009.11.010

Ketscher, L., and Knobeloch, K. P. (2015). ISG15 uncut: dissecting enzymatic and non-enzymatic functions of USP18 *in vivo*. *Cytokine* 76, 569–571. doi: 10.1016/j.cyto.2015.03.006

Kneller, J. M., Lu, M., and Bracken, C. (2002). An effective method for the discrimination of motional anisotropy and chemical exchange. *J. Am. Chem. Soc.* 124, 1852–1853. doi: 10.1021/ja017461k

Komander, D., and Rape, M. (2012). The ubiquitin code. *Annu. Rev. Biochem.* 81, 203–229. doi: 10.1146/annurev-biochem-060310-170328

Kunitani, M., Johnson, D., and Snyder, L. R. (1986). Model of protein conformation in the reversed-phase separation of interleukin-2 muteins. *J. Chromatogr.* 371, 313–333. doi: 10.1016/S0021-9673(01)94716-8

Labbadia, J., and Morimoto, R. I. (2015). The biology of proteostasis in aging and disease. *Annu. Rev. Biochem.* 84, 435–464. doi: 10.1146/annurev-biochem-060614-033955

Landre, V., Revi, B., Mir, M. G., Verma, C., Hupp, T. R., Gilbert, N., et al. (2017). Regulation of transcriptional activators by DNA-binding domain ubiquitination. *Cell Death Differ.* 24, 903–916. doi: 10.1038/cdd.2017.42

Lee, W., Tonelli, M., and Markley, J. L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31, 1325–1327. doi: 10.1093/bioinformatics/btu830

Matochko, W. L., Chu, K., Jin, B., Lee, S. W., Whitesides, G. M., and Derda, R. (2012). Deep sequencing analysis of phage libraries using Illumina platform. *Methods* 58, 47–55. doi: 10.1016/j.ymeth.2012.07.006

Matsuoka, S., Ballif, B. A., Smogorzewska, A., Mcdonald, E. R. 3rd, Hurov, K. E., Luo, J., et al. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316, 1160–1166. doi: 10.1126/science.1140321

Metzger, M. B., Pruneda, J. N., Klevit, R. E., and Weissman, A. M. (2014). RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim. Biophys. Acta* 1843, 47–60. doi: 10.1016/j.bbamcr.2013.05.026

Middleton, A. J., Budhidarmo, R., and Day, C. L. (2014). Use of E2~ubiquitin conjugates for the characterization of ubiquitin transfer by RING E3 ligases such as the inhibitor of apoptosis proteins. *Methods Enzymol.* 545, 243–263. doi: 10.1016/B978-0-12-801430-1.00010-X

Mohtar, M. A., Hernychova, L., O'neill, J. R., Lawrence, M. L., Murray, E., Vojtesek, B., et al. (2018). The sequence-specific peptide-binding activity of the protein sulfide isomerase AGR2 directs its stable binding to the oncogenic receptor EpCAM. *Mol. Cell. Proteomics* 17, 737–763. doi: 10.1074/mcp.RA118.000573

Morelli, X., and Hupp, T. (2012). Searching for the holy grail; protein-protein interaction analysis and modulation. *EMBO Rep.* 13, 877–879. doi: 10.1038/embor.2012.137

Murray, E., Mckenna, E. O., Burch, L. R., Dillon, J., Langridge-Smith, P., Kolch, W., et al. (2007). Microarray-formatted clinical biomarker assay development using peptide aptamers to anterior gradient-2. *Biochemistry* 46, 13742–13751. doi: 10.1021/bi7008739

Narayan, V., Landre, V., Ning, J., Hernychova, L., Muller, P., Verma, C., et al. (2015). Protein-protein interactions modulate the docking-dependent E3-ubiquitin ligase activity of Carboxy-terminus of Hsc70-interacting protein (CHIP). *Mol. Cell. Proteomics* 14, 2973–2987. doi: 10.1074/mcp.M115.051169

Rivier, J., and Mcclintock, R. (1983). Reversed-phase high-performance liquid chromatography of insulins from different species. *J. Chromatogr.* 268, 112–119. doi: 10.1016/S0021-9673(01)95395-6

Robson, A. F., Hupp, T. R., Lickiss, F., Ball, K. L., Faulds, K., and Graham, D. (2012). Nanosensing protein allostery using a bivalent mouse double minute two (MDM2) assay. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8073–8078. doi: 10.1073/pnas.1116637109

Saha, A., Lewis, S., Kleiger, G., Kuhlman, B., and Deshaies, R. J. (2011). Essential role for ubiquitin-ubiquitin-conjugating enzyme interaction in ubiquitin discharge from Cdc34 to substrate. *Mol. Cell* 42, 75–83. doi: 10.1016/j.molcel.2011.03.016

Scheffner, M., and Kumar, S. (2014). Mammalian HECT ubiquitin-protein ligases: biological and pathophysiological aspects. *Biochim. Biophys. Acta* 1843, 61–74. doi: 10.1016/j.bbamcr.2013.03.024

Shimizu, H., Burch, L. R., Smith, A. J., Dornan, D., Wallace, M., Ball, K. L., et al. (2002). The conformationally flexible S9-S10 linker region in the core domain of p53 contains a novel MDM2 binding site whose mutation increases ubiquitination of p53 *in vivo*. *J. Biol. Chem.* 277, 28446–28458. doi: 10.1074/jbc.M202296200

Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M., and Hicke, L. (2001). Distinct functional surface regions on ubiquitin. *J. Biol. Chem.* 276, 30483–30489. doi: 10.1074/jbc.M103248200

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317. doi: 10.1126/science.4001944

Stevens, C., Lin, Y., Harrison, B., Burch, L., Ridgway, R. A., Sansom, O., et al. (2009). Peptide combinatorial libraries identify TSC2 as a death-associated protein kinase (DAPK) death domain-binding protein and reveal a stimulatory role for DAPK in mTORC1 signaling. *J. Biol. Chem.* 284, 334–344. doi: 10.1074/jbc.M805165200

Tompa, P., Davey, N. E., Gibson, T. J., and Babu, M. M. (2014). A million peptide motifs for the molecular biologist. *Mol. Cell* 55, 161–169. doi: 10.1016/j.molcel.2014.05.032

Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., et al. (2004). *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303, 844–848. doi: 10.1126/science.1092472

Wallace, M., Worrall, E., Pettersson, S., Hupp, T. R., and Ball, K. L. (2006). Dual-site regulation of MDM2 E3-ubiquitin ligase activity. *Mol. Cell* 23, 251–263. doi: 10.1016/j.molcel.2006.05.029

Wawrzynow, B., Pettersson, S., Zylicz, A., Bramham, J., Worrall, E., Hupp, T. R., et al. (2009). A function for the RING finger domain in the allosteric control of MDM2 conformation and activity. *J. Biol. Chem.* 284, 11517–11530. doi: 10.1074/jbc.M809294200

Wenzel, D. M., Stoll, K. E., and Klevit, R. E. (2011). E2s: structurally economical and functionally replete. *Biochem. J.* 433, 31–42. doi: 10.1042/BJ20100985

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., et al. (1995). 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6, 135–140. doi: 10.1007/BF00211777

Zhang, Z., and Marshall, A. G. (1998). A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* 9, 225–233. doi: 10.1016/S1044-0305(97)00284-5

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact