D560–D570 *Nucleic Acids Research, 2022, Vol. 50, Database issue* *Published online 19 October 2021*
https://doi.org/10.1093/nar/gkab926

SYNBIP: synthetic binding proteins for research, diagnosis and therapy

Xiaona Wang[1,†], Fengcheng Li[2,†], Wenqi Qiu[3,†], Binbin Xu[1], Yanlin Li[1], Xichen Lian[2], Hongyan Yu[1], Zhao Zhang[1], Jianxin Wang[4], Zhaorong Li[5], Weiwei Xue[1,*] and Feng Zhu[1,2,5,*]

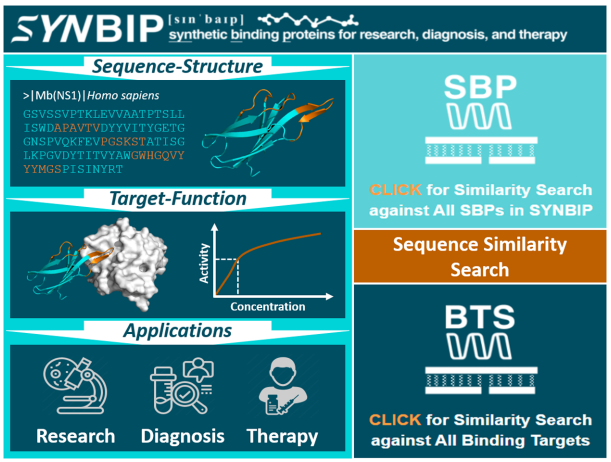
[1]School of Pharmaceutical Sciences, Chongqing University, Chongqing 401331, China, [2]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China, [3]Department of Surgery, HKU-SZH & Faculty of Medicine, The University of Hong Kong, Hong Kong, China, [4]School of Computer Science and Engineering, Central South University, Changsha 410083, China and [5]Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 330110, China

Received August 10, 2021; Revised September 13, 2021; Editorial Decision September 22, 2021; Accepted October 14, 2021

ABSTRACT

The success of protein engineering and design has extensively expanded the protein space, which presents a promising strategy for creating next-generation proteins of diverse functions. Among these proteins, the synthetic binding proteins (SBPs) are smaller, more stable, less immunogenic, and better of tissue penetration than others, which make the SBP-related data attracting extensive interest from worldwide scientists. However, no database has been developed to systematically provide the valuable information of SBPs yet. In this study, a database named 'Synthetic Binding Proteins for Research, Diagnosis, and Therapy (SYNBIP)' was thus introduced. This database is unique in (a) comprehensively describing thousands of SBPs from the perspectives of scaffolds, biophysical & functional properties, etc.; (b) panoramically illustrating the binding targets & the broad application of each SBP and (c) enabling a similarity search against the sequences of all SBPs and their binding targets. Since SBP is a human-made protein that has not been found in nature, the discovery of novel SBPs relied heavily on experimental protein engineering and could be greatly facilitated by *in-silico* studies (such as AI and computational modeling). Thus, the data provided in SYNBIP could lay a solid foundation for the future development of novel SBPs. The SYNBIP is accessible without login requirement at both official (https://idrblab.org/synbip/) and mirror (http://synbip.idrblab.net/) sites.

GRAPHICAL ABSTRACT



INTRODUCTION

The success of protein engineering and design has extensively expanded the protein space (1–3), which presents a promising strategy for developing next-generation proteins of diverse functions, such as binders (4,5), enzymes (6,7), biosensors (8–10), etc. Protein engineering was first applied to design antibodies (11) that were used as an essential tool for virtually every biological research discipline (12). However, various serious attributes of antibodies (such as large size, poor folding, & stability issue) limit their application in living systems (13). Therefore, the idea of synthesizing binding protein using scaffold from antibody fragments (e.g. nanobody (14–16)) or non-antibody (e.g. designed ankyrin repeat protein (17)) has recently emerged as a popular technique (18–20), which opens up exciting opportunity to

*To whom correspondence should be addressed. Tel: +86 187 0236 4293; Fax: +86 023 6567 8450; Email: xueww@cqu.edu.cn
Correspondence may also be addressed to Feng Zhu. Email: zhufeng@zju.edu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

develop numerous synthetic binding proteins (SBPs, that are tailored to bind to a particular molecular target of interest) (21–28).

Compared with classical antibodies, the SBPs are smaller, and most of them are more stable, less immunogenic, and better of tissue penetration (24,29–31), which makes them hold great promise for tackling biomedical challenges, such as COVID (32–37), cancers (38–41), and CNS disorders (42,43). Moreover, many SBP-based biological therapies exhibit their clinical implications with some drugs approved (e.g. *Ecallantide* (44)) and others in clinical trials/preclinical investigations (45,46). Due to the great importance, SBP-related data attract extensive interests from worldwide scientists (47–54). Such data include (i) the appropriate scaffolds that shape the starting point of rational SBP design (47,48), (ii) the biophysical (e.g. thermal stability) & functional (e.g. binding conformation variation) properties of SBP that determine target binding potency (49,50,55) and (iii) the structure and sequence properties of privileged SBPs that can facilitate the design of new SBPs of minimized off-target interaction (52–54). In other words, these data are essential for the fields of protein engineering and the development of next-generation proteins (54,56).

So far, several SBP-related databases have been developed and are currently active, the majority of which focus on providing intact data of antibody (e.g. ABCD (57) and Yvis (58)) or nanobody (e.g. Thera-SAbDab (59) and sdAb-DB (60)), and another of which are specialized in describing structural classification of diverse antibodies (e.g. PyIgClassify (61)). Moreover, some reputable databases (e.g. STRING (62), BioGRID (63), DifferentialNet (64), HPRD (65), and IntAct (66)) and tools (e.g. iLearnPlus (67), and DeepCleave (68,69)) demonstrating a wealth of information of protein–protein interactions and convenient analyses of protein sequences have been available. However, no database has been constructed yet to systematically describe the SBPs' information of their scaffold, sequence, structure, biophysical/functional property, and so on.

Herein, a new database, synthetic binding proteins for research, diagnosis, and therapy (SYNBIP) was therefore introduced. First, comprehensive literature reviews on SBPs were conducted, and thousands of unique SBPs binding specifically to physiologically relevant targets were collected. These SBPs were from diverse scaffolds, such as affibodies (70), anticalins (71), DARPins (17), i-bodies (72), monobodies/adnectins (73), nanobodies (14), repebodies (74), scFabs (75), scFvs (76) and vNARs (77). Second, based on the collected SBPs, their binding targets were manually curated from literatures, and the panoramic view of their binding profile and application (therapy, diagnosis and/or research) was provided. Third, the sequence-based similarity search against all SBPs and their binding targets was enabled to facilitate the design of novel SBPs and application to new research directions. All these efforts contributed to the unique characteristics of SYNBIP (described in Figure 1). Since the SBP is a human-made protein that has not been found in nature, the discovery of novel SBPs relied heavily on the experimental protein engineering (78) and can be greatly facilitated by *in silico* studies (e.g. AI technique (2,49,79) and computational modeling (3,80–82)). Thus, the unique data and functions provided in SYNBIP (https://idrblab.org/synbip/) laid a solid foundation for the future development of novel SBPs.

FACTUAL CONTENT AND DATA RETRIEVAL

Data collection and SBP classification & scaffolds

The SBPs and their scaffolds were collected by the literature review in PubMed (83). Frist, using the keyword combinations of 'synthetic + binding protein + scaffold', 'non-antibody + scaffold', 'engineered + binding protein + scaffold', etc., a total of 68 SBP scaffolds were identified. Then, 2074 SBPs were collected by the keyword searching of SBP scaffold names and their synonyms in PubMed (83). Third, detailed information of each SBP was further obtained from KEGG (84), PDB (85), UniProt (86) and additional literature review. The resulting information included SBP name, sequence, structure, molecular weight, expression system, function, applications, research organizations and thermal denaturation temperature. Fourth, several reputable clinical databases (e.g. ClinicalTrials.gov, ChiCTR, and EU-CTR) and the official websites of many pharmaceutical enterprises (e.g. AffibodyAB, NavigoProteins, and Bicycle Therapeutics) were scanned, and the highest clinical development status for each SBP was therefore confirmed. Finally, the additional SBP affiliated data were reviewed and collected from literatures, which included binding targets (affinity, mechanism of action, etc.), atomic details (nonstandard amino acids, connections in the sequences, etc.) and experimental details (expression system, *in vitro* method, etc.).

There are two types of SBPs in SYNBIP: non-antibody and antibody fragment (87,88). As shown in Figure 2, a table of scaffolds for all SBPs collected in SYNBIP was described, which was the key starting points for the engineering of novel SBPs in current researches (24,89). The numbers of non-antibody scaffolds and antibody fragments were 57 and 11, respectively. The scaffolds in the same column of Figure 2 were engineered within the same type of regions. There were seven region types for non-antibody scaffold (such as loops, α-helixes, and β-sheets) and two types for antibody fragment (single domain & multi-domains). The scaffolds within the same column were ordered based on the molecular weight (decreasing from the bottom to the top). In this study, the cyclic peptides were considered as SBP scaffolds due to the following reasons. First, similar to those known SBPs, cyclic peptides have small molecular weight, relative high stability, and with the region of protein engineering in the loops on ring-shaped structures (90). Second, the binding property and function of cyclic peptides to their targets were highly similar to those known SBPs (91). As a result, Figure 2 provided comprehensive description on all SBP scaffolds collected in SYNBIP, such as scaffold name, typical 3D structure, the ranges of molecular weight and melting temperature, and so on. These data were essential for exploring the thermal stability (49,50) and off-target interactions (52–54) of SBP, and were thus key for the rational design of new SBP (48). Moreover, the clas-

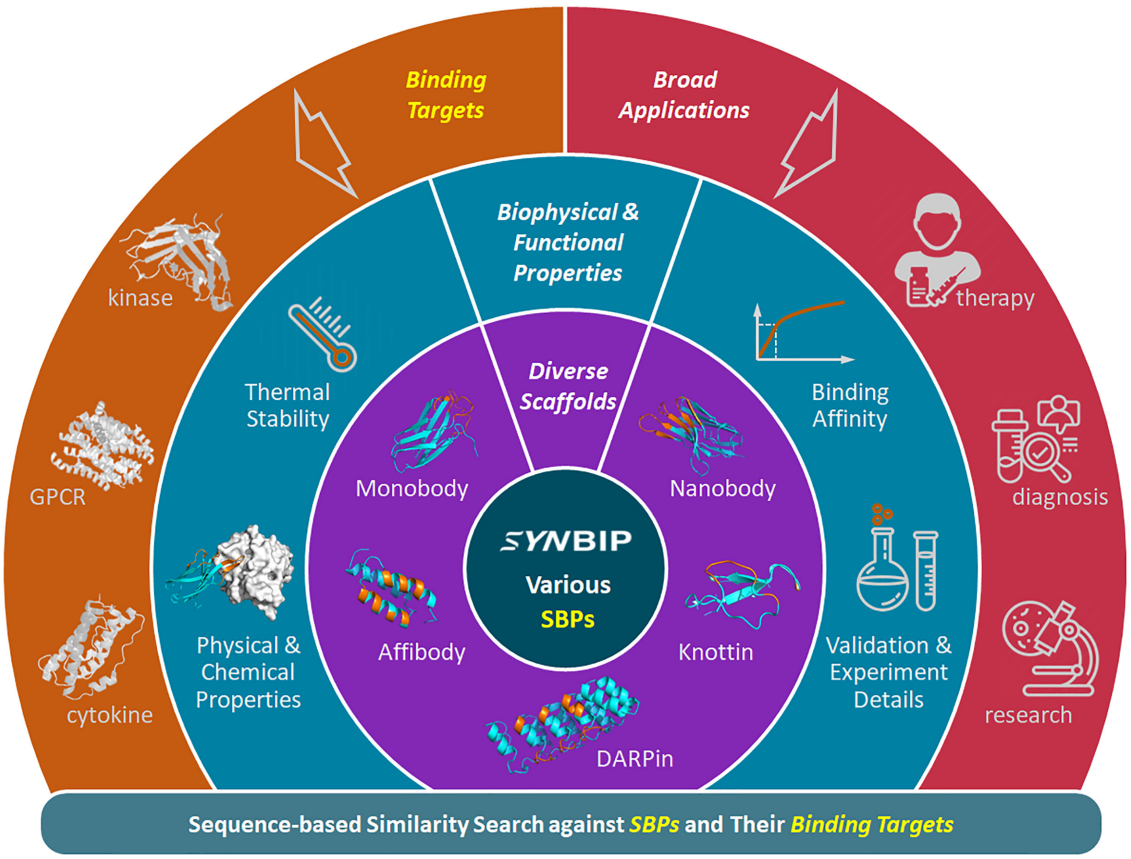


Figure 1. The unique characteristics of SYNBIP. Extensively describing a comprehensive set of synthetic binding proteins (SBPs) from the perspectives of scaffolds, biophysical and functional properties, *etc*. (shown in the inner three layers); panoramically illustrating the binding target & and the broad application of each SBP (presented in the outermost layer); enabling the sequence-based similarity search against SBPs and their binding targets (provided at the bottom).

sification (class/scaffold) and the amount of SBPs in each class/scaffold were described in Supplementary Figure S1. As shown, those top-5 scaffolds of the most SBPs were: scFv (343 SBPs), nanobody (271 SBPs), monobody (234 SBPs), DARPin (182 SBPs) and Fab (114 SBPs).

It was worth mentioning that SYNBIP mainly focused on providing the human-made 'synthetic' proteins by excluding the natural-occurring ones such as native protein binder. This was different from SAbDab (59,92), a database previously featured in NAR to provide valuable nanobody data. Particularly, the majority (~90%) of the nanobodies in the SAbDab were the native protein. Since the SYNBIP had collected and described 271 'synthetic' nanobodies, it could be adopted as an important complement to those available databases, such as SAbDab (59,92).

Biophysical, structural & functional data of SBPs

Low molecular weight. Compared with the classical antibodies, the molecular weights (MWs) of SBPs were much lower. As provided in Figure 2, the MWs of 65.7% SBPs collected in SYNBIP were between 2 and 20 kDa, which demonstrated that the size was a major feature of SBPs as next-generation proteins. Owing to its low MW, the SBP showed the advantages of efficient tissue delivery and penetration (30), which are well-suited for generating bi-/multi-specific molecules (88).

High thermal stability. Thermal stability (measured by thermal denaturation temperature, Tm) of the starting scaffold is frequently considered in protein engineering (24,89). As illustrated in Figure 2, except for some SBPs from the scaffold of *i*-body, beta roll domain, EVH1 domain, beta-hairpin mimetic, abdurin, and diabody, the Tm values of the majority (72%) of the SBP scaffolds were within the range of 37–120°C. This indicated that most of the SBPs in SYNBIP were stable at high temperature, and were therefore relatively easy and cheap for production in bacteria, yeast or even by chemical synthesis (30). Moreover, those SBPs were reported to remain remarkable stabilities and binding activities after long-term (years) storage at room temperature (30).

In addition to those low molecular weight and high thermal stability, the solubility and expression yield were essential for SBP's applications (24). However, only a few relevant data (<100 entries) could be obtained, due to the limited number of related publications. With the rapid advances in these promising fields, we would expect an explosion of such valuable data, which will be timely collected to SYNBIP

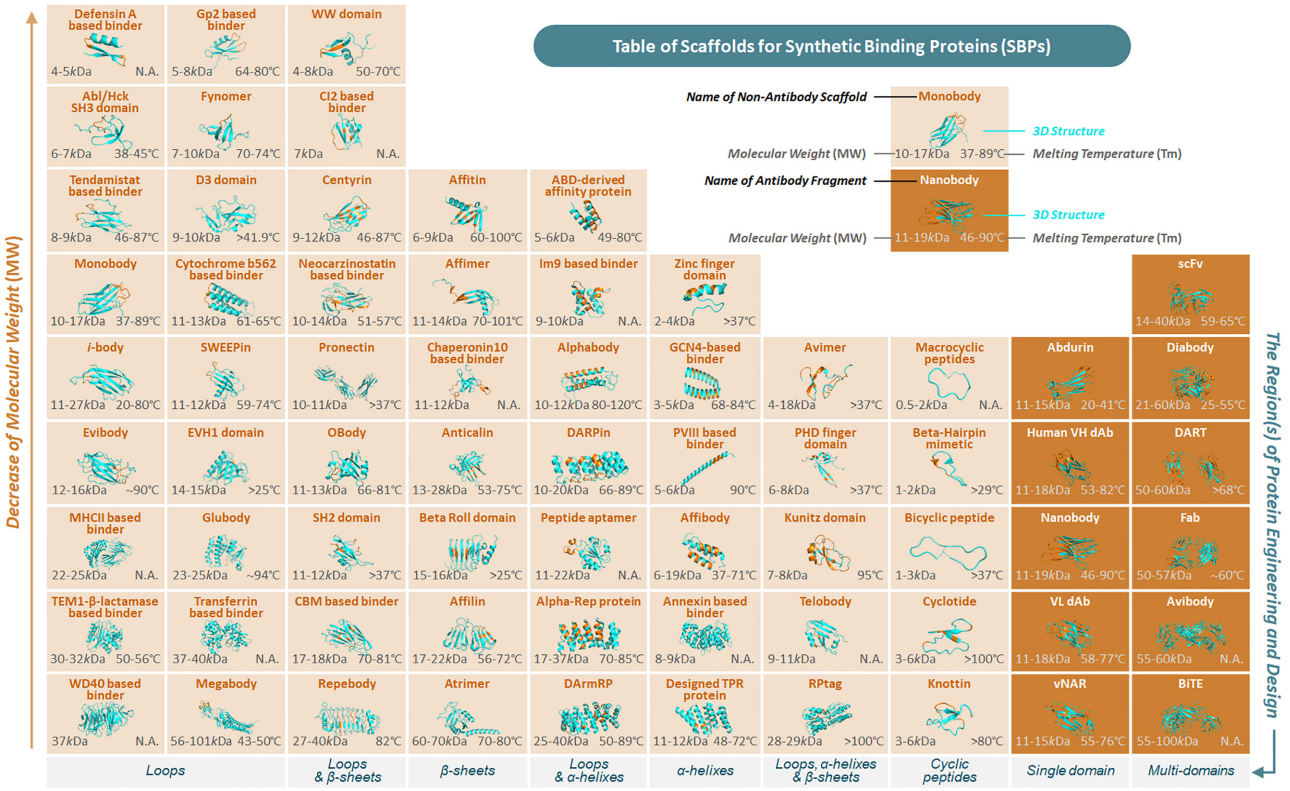


Figure 2. Table of scaffolds for all synthetic binding proteins (SBPs) collected in SYNBIP. The backgrounds of non-antibody scaffolds and antibody fragments were colored in light orange and orange, respectively. The scaffolds in the same column indicated the same region(s) of protein engineering and design. There were nine types of region(s), seven of which were for non-antibody scaffolds (loops; loops & β-sheets; β-sheets; loops & α-helixes; α-helixes; loops, α-helixes & β-sheets; and cyclic peptides) and the remaining two of which were for antibody fragments (single domain; and multi-domains). The scaffolds in the same column were ordered according to their molecular weights (decreasing from the bottom to the top). Within each cell, the details of a particular scaffold were provided, which included scaffold name, representative structure, the range of molecular weights (MWs), and the range of melting temperature (Tm).

for facilitating the further advancement for this research direction.

Sequence & structure. There were 1359 SBPs in SYNBIP with full-length sequence information, which accounted for most (65.5%) of those 2074 collected SBPs. Such sequence data were frequently adopted in the site-directed mutagenesis study and design of scaffold-based library, which greatly facilitates the engineering of SBP (2,48). Moreover, based on literature review, the structures of 246 SBPs were resolved by the technique of nuclear magnetic resonance (NMR), X-ray crystallography (X-ray) or cryogenic electron microscopy (Cryo-EM) and thus collected to the SYNBIP, which had great implications for structure-guided engineering of critical protein regions (38,93–96).

Apart from the experimentally validated structures, the computationally modelled SBP structures were found to be capable of modelling 3D structures (from protein sequences), which extensively facilitated the rational design of SBP (97–100). In SYNBIP, the 3D structure of SBP without any experimentally validated structural information was therefore modelled using a well-established protocol *trRosetta* (101,102), which combined algorithms of deep residual network and *Rosetta*-constrained energy minimization, and had been widely used to the rapid and accurate prediction of protein structure (102). As a result, the 3D structures of 1083 SBPs without an experimentally validated structure were modelled by *trRosetta* (101,102), and the confidence estimation scores (TM-scores) of all predicted SBP structures were higher than 0.7, indicating a correctly modeled topology (101). Although the modelled structures may not be completely identical to SBP's 3D conformation (98), they can be adopted as references for guiding the rational design of SBP (98). To distinguish the experimentally validated SBP structures from those computationally modelled ones, these two types of structure were therefore labelled in SYNBIP website by '*Experimentally Validated Structure*' and '*Computationally Modelled Structure*', respectively. Both types of SBP structure data together with the SBP sequence information can be fully downloaded directly from the official (https://idrblab.org/synbip/) and mirror (http://synbip.idrblab.net/) sites of SYNBIP.

Binding target & affinity. The binding targets of SBPs were collected by literature reviews or from the official websites of many pharmaceutical enterprises. As a result, the binding targets of all SBPs (2074 in total) were identified, which resulted in 423 protein targets, 28 small molecular targets, and 15 other targets (such as carbohydrates, RNAs, DNAs, *etc.*). As shown in Supplementary Figure S2, the

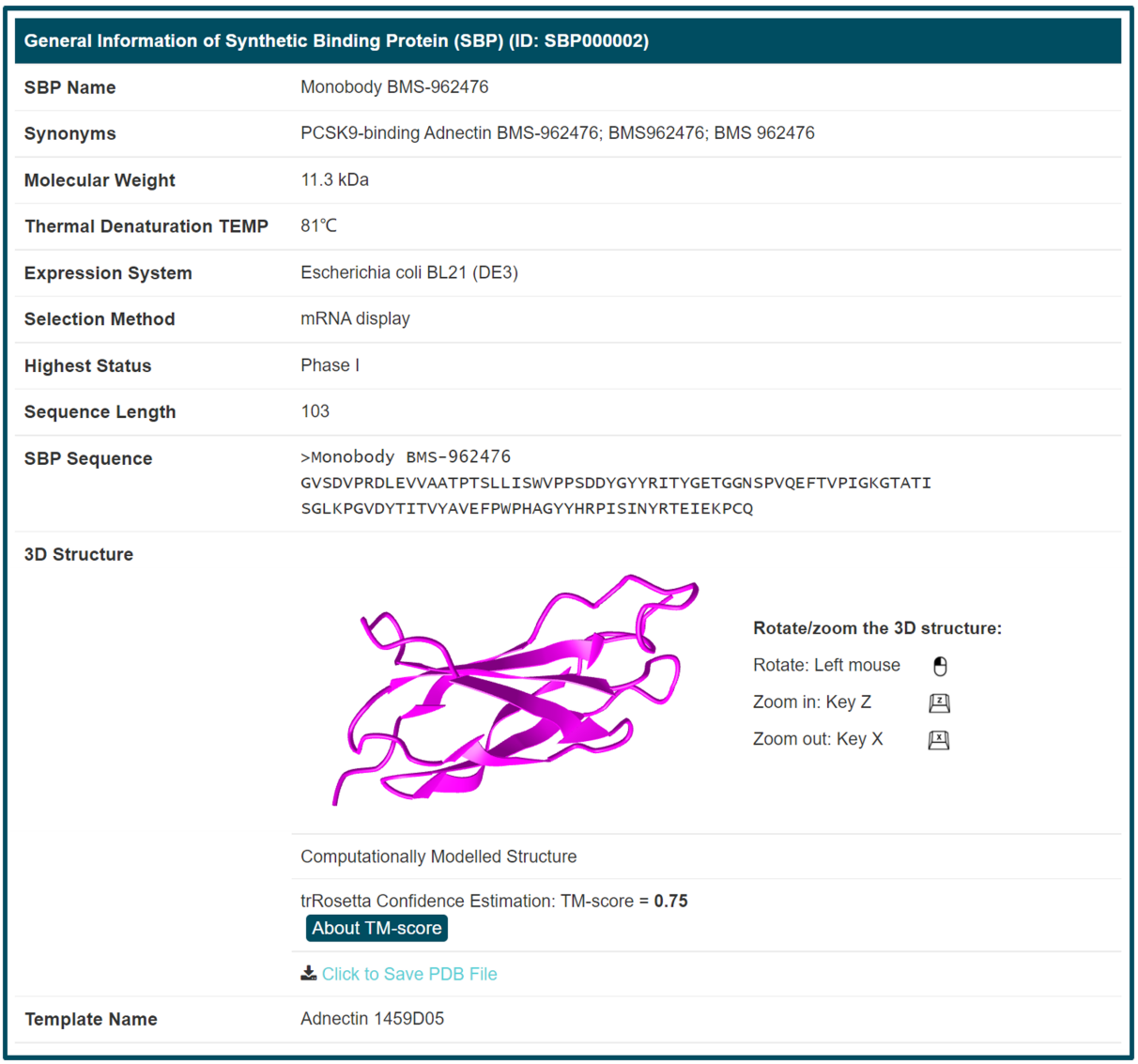


Figure 3. A typical page in SYNBIP describing the SBP (monobody BMS-962476, SBP000002). Monobody BMS-962476 was explicitly described as one SBP with molecular weight of 11.3kDa, thermal denaturation temperature of 81°C, and highest clinical trial status in Phase I. Meanwhile, its sequence and structure were fully provided and could be directly download from this page.

targets were from very broad origins. Particularly, 317 targets were from very diverse metazoan species, such as human, mouse, bovine, jellyfish, scorpion, etc.; 106 targets were from various microorganisms, such as *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Mycobacterium tuberculosis*, *Escherichia coli*, *Plasmodium falciparum*, *Streptomyces clavuligerus*, etc.; and 43 targets were from plant species, such as *Lolium perenne*, *Ricinus communis*, *Chlamydomonas reinhardtii*, etc.

Moreover, 1860, 192 and 22 out of all 2074 SBPs collected in SYNBIP were identified with 1, 2 and ≥3 binding targets, respectively. Particularly, 1384 out of all the collected SBPs (∼66.7%) were with binding affinities reported, the value of which was measured by dissociation constants (K_d), inhibition constants (*Ki*), half maximal inhibitory concentrations (*IC50*), and so on. Among these 1384 SBPs, 1087 (78.5%), 81 (5.9%) and 216 (15.6%) were found with binding affinities against 243 protein, 19 small molecular, and 3 other targets, respectively. In the meantime, 16.8%, 24.1%, 24.7%, 16.4% and 18.1% of all the affinities collected in SYNBIP were <1 nM, 1–10 nM, 10–100 nM, 100 nM–1 μM and >1 μM, respectively. In other words, most (65.4%) affinities were <100 nM, which may benefit from the highly-specific molecular recognition of SBPs (89).

SBPs' applications in research, diagnosis & therapy

Due to the rapidly-growing interest in SBP design based on the protein scaffold of low molecular weight and high thermal stability, significant advances have been made not only in the design of new binders but also in their applications to various directions of research, diagnosis, and therapy (22). In SYNBIP, the detailed applications together with the cur-

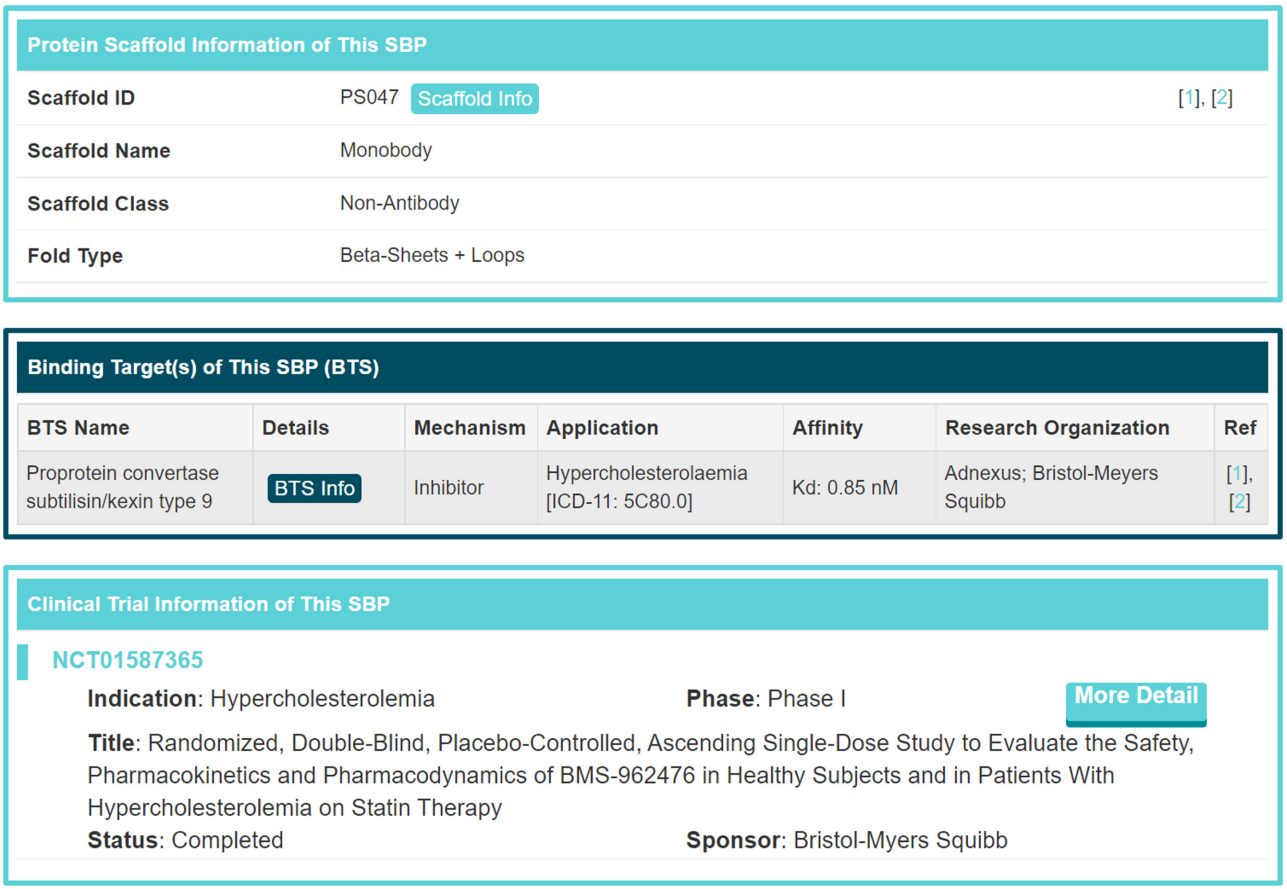


Figure 4. The additional information of the SBP provided in SYNBIP. Those provided additional data (for this particular SBP: monobody BMS-962476) included: SBP scaffold (e.g. monobody), binding target (e.g. proprotein convertase 9), and clinical development status (Phase I for treating hypercholesterolemia). Detailed clinical information was also provided at the bottom.

rent clinical/investigative status were described in the corresponding webpage of a specific SBP.

Research. There were 1189 SBPs in SYNBIP (from 56 scaffolds provided in Figure 2) reported as powerful *research* tools for bridging the functional investigations with the structural ones (22), monitoring the localization of endogenous proteins in living system (103), stabilizing the protein structure for capturing specific crystalized conformation (104), and so on.

Diagnosis. 139 SBPs (covering 20 scaffolds in Figure 2) had been tested or applied as *diagnosis* tools for monitoring the *in-vivo* sites of disease's occurrence (89), imaging the disease-associated molecular targets (105,106), and so on.

Therapeutics. 746 SBPs (belonging to 44 scaffolds in Figure 2) were engineered as *therapeutics* for the treatment of various diseases, especially for the complex indications like cancer, infection, CNS disorders, *etc*. Among these clinically important SBPs, 66 of them (belonging to 17 protein scaffolds) had been tested in clinical trials, and 9 of them had been approved by FDA. Particularly, 71 SBPs (belonging to nine protein scaffolds) were clinically adopted for dealing with the pandemic of COVID-19, and two representative SBPs (*Glenzocimab* and *Ensovibep*) were clinically tested in Phase II and Phase III, respectively. All in all, the detailed descriptions of the applications in research, diagnosis, and therapy were provided in the corresponding page of each SBP.

Similarity-based identification of SBP from SYNBIP

In addition to the keyword search, the sequence-based similarity search against SBPs in SYNBIP was realized, which might facilitate the design of SBP and its application to novel research fields. The level of similarity between the sequence of an input protein and that of those SYNBIP SBPs were evaluated using BLAST (107), and with the sequence identities listed out in the order from the highest to the lowest. Using the sequence of 'monobody anti-KRas/HRas NS1' as one query, a total of 226 SBPs could be identified. For the identified SBPs showing high sequence similarity (e.g. monobody anti-KRas 12VC1), they were generated from the same/similar protein scaffolds and had their stability and molecular size & function similar to the query sequence of NS1. Thus, it was reasonable to expect that those identified SBPs could be used as references to design novel SBPs of enhanced binding affinity and specificity (22). For the identified SBPs showing medium sequence similarity (e.g. centyrin anti-ERBB1 83v2 variant), their scaffolds were different from that of the query sequence, which might

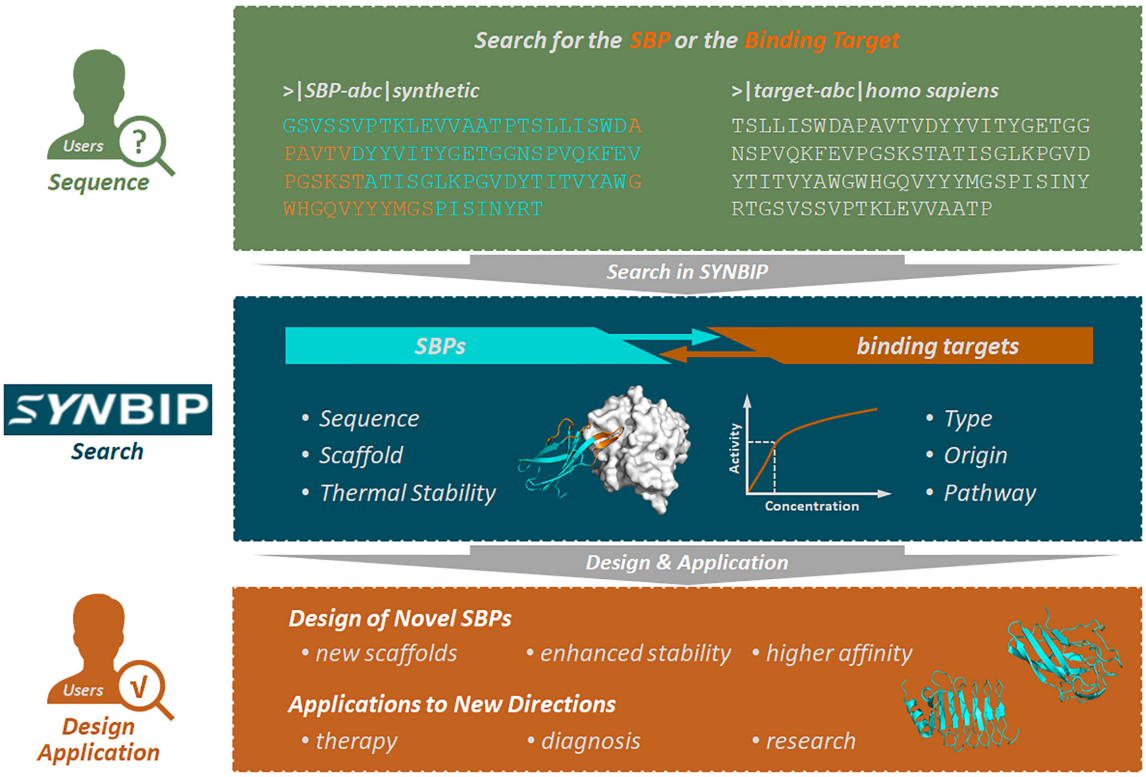


Figure 5. The function of sequence-based similarity search realized in SYNBIP. Sequence-based similarity search was carried out by BLASTing (107) against SBPs or their binding targets, which could facilitate the design of novel SBPs and application to the new research directions.

be adopted as template to design new SBP scaffold with enhanced stability (108). For the identified SBPs showing relatively low sequence similarity (e.g. nanobody anti-GlyT1 clone 5), their fold type might be the same as or similar to that of the query sequence, which could also provide useful information for the de novo design of SBPs (38).

Besides, the structure-based similarity search could also facilitate protein engineering and novel protein design. So far, some tools for structure alignment & comparison had been available, such as TM-align (109), Fr-TM-align (110) and MM-align (111). Although these available tools were powerful in structure-based alignment and similarity comparison, their applications were limited by their excessive time cost spent on comparing two structures, which made it impossible to scan all SBP structures by online calculation. To partially enable the structure-based similarity search function, SYNBIP allowed free download of all SBP structures (each was indicated by its unique SBP ID), and the users can use the local version of TM-align that is downloadable from the tool's official website (https://zhanggroup.org/TM-align/) to scan their query structures against all SBP structures in SYNBIP based on their local computing resources.

All in all, the similarity search functions based on either sequence or structure were expected to be useful for current protein engineering and design. These search functions provided in the latest version SYNBIP could therefore be essential for the related research communities. Furthermore, a user manual that provided a step-by-step instruction on the usage of SYNBIP was shown in the 'Help' page, whose web link could be readily found on the home page of SYNBIP.

Data access, retrieval and standardization

In SYNBIP, a user-friendly way to identify the SBPs of users' interest was designed and provided. Taking the searching of 'monobody BMS-962476' as an example, its corresponding entry could be identified by simply typing this keyword and searching against all SYNBIP data. As provided in Figure 3, BMS-962476 was explicitly described as the SBP with molecular weight of 11.3kDa, thermal denaturation temperature of 81°C, and highest clinical trial status in Phase I. Meanwhile, its sequence and structure were fully provided and could be directly download from its own page. Furthermore, the additional data of SBP scaffold (e.g. monobody), binding target (e.g. proprotein convertase 9), clinical development status (Phase I for treating hypercholesterolemia), and so on, were also explicitly described (see the webpage screenshots in Figures 3 and 4).

To make the access and analysis of SYNBIP data convenient for all users, the collected raw data were carefully cleaned up and then systematically standardized. These standardizations included: (i) all SYNBIP diseases were standardized using the latest version of International Classification of Disease (ICD-11, officially released by World Health Organization (112), which was expected to serve comprehensive health managements (113); (ii) all SBP binding targets were standardized by and crosslinked to available databases, and the extended data of each SBP could be

accessed by hyperlinks to UniProt (86), ClinicalTrials.gov (114), VARIDT (115), ChiCTR (116), EUCTR (117), TTD (118), PDB (85), INTEDE (119), etc.; (iii) a sequence-based similarity search against SBPs in SYNBIP and their binding targets was enabled to facilitate the design of SBPs and their application to new research fields (described in Figure 5). All SBP data can be viewed, assessed, and downloaded from the SYNBIP website, which is freely assessable without login requirement by all users at its official (https://idrblab.org/synbip/) and mirror (http://synbip.idrblab.net/) sites.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Entrepreneurship and Innovation Support Plan for Chinese Overseas Students of Chongqing [cx2020127]; Natural Science Foundation of Zhejiang Province [LR21H300001]; National Natural Science Foundation of China [81872798, U1909208]; Leading Talent of the 'Ten Thousand Plan'—National High-Level Talents Special Support Plan of China; 'Double Top-Class' University Project [181201*194232101]; Fundamental Research Fund for Central Universities [2018QNA7023, 2019CDYGYB005]; Key R&D Program of Zhejiang Province [2020C03010]; Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare; Alibaba Cloud; Information Technology Center of Zhejiang University. Funding for open access charge: Entrepreneurship and Innovation Support Plan for Chinese Overseas Students of Chongqing [cx2020127].

Conflict of interest statement. None declared.

REFERENCES

1. Alley,E.C., Khimulya,G., Biswas,S., AlQuraishi,M. and Church,G.M. (2019) Unified rational protein engineering with sequence-based deep representation learning. *Nat. Methods*, **16**, 1315–1322.
2. Yang,K.K., Wu,Z. and Arnold,F.H. (2019) Machine-learning-guided directed evolution for protein engineering. *Nat. Methods*, **16**, 687–694.
3. Huang,P.S., Boyken,S.E. and Baker,D. (2016) The coming of age of de novo protein design. *Nature*, **537**, 320–327.
4. Chevalier,A., Silva,D.A., Rocklin,G.J., Hicks,D.R., Vergara,R., Murapa,P., Bernard,S.M., Zhang,L., Lam,K.H., Yao,G. *et al.* (2017) Massively parallel de novo protein design for targeted therapeutics. *Nature*, **550**, 74–79.
5. Sun,M.G., Seo,M.H., Nim,S., Corbi-Verge,C. and Kim,P.M. (2016) Protein engineering by highly parallel screening of computationally designed variants. *Sci. Adv.*, **2**, e1600692.
6. Arnold,F.H. (2019) Innovation by evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed. Engl.*, **58**, 14420–14426.
7. Li,R., Wijma,H.J., Song,L., Cui,Y., Otzen,M., Tian,Y., Du,J., Li,T., Niu,D., Chen,Y. *et al.* (2018) Computational redesign of enzymes for regio- and enantio-selective hydroamination. *Nat. Chem. Biol.*, **14**, 664–670.
8. Quijano-Rubio,A., Yeh,H.W., Park,J., Lee,H., Langan,R.A., Boyken,S.E., Lajoie,M.J., Cao,L., Chow,C.M., Miranda,M.C. *et al.* (2021) De novo design of modular and tunable protein biosensors. *Nature*, **591**, 482–487.
9. Kang,S., Davidsen,K., Gomez-Castillo,L., Jiang,H., Fu,X., Li,Z., Liang,Y., Jahn,M., Moussa,M., DiMaio,F. *et al.* (2019) COMBINES-CID: an efficient method for de novo engineering of highly specific chemically induced protein dimerization systems. *J. Am. Chem. Soc.*, **141**, 10948–10952.
10. Pan,X. and Kortemme,T. (2021) Recent advances in de novo protein design: principles, methods, and applications. *J. Biol. Chem.*, **296**, 100558.
11. Binz,H.K. and Pluckthun,A. (2005) Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.*, **16**, 459–469.
12. Doerr,A. (2015) Protein binder woes. *Nat. Methods*, **12**, 373.
13. West,N.R., Hegazy,A.N., Owens,B.M.J., Bullers,S.J., Linggi,B., Buonocore,S., Coccia,M., Gortz,D., This,S., Stockenhuber,K. *et al.* (2017) Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. *Nat. Med.*, **23**, 579–589.
14. Muyldermans,S. (2021) Applications of nanobodies. *Annu. Rev. Anim. Biosci.*, **9**, 401–421.
15. Huang,Z., Li,Z., Zhang,X., Kang,S., Dong,R., Sun,L., Fu,X., Vaisar,D., Watanabe,K. and Gu,L. (2020) Creating red light-switchable protein dimerization systems as genetically encoded actuators with high specificity. *ACS Synth. Biol.*, **9**, 3322–3333.
16. Gomez-Castillo,L., Watanabe,K., Jiang,H., Kang,S. and Gu,L. (2020) Creating highly specific chemically induced protein dimerization systems by stepwise phage selection of a combinatorial single-domain antibody library. *J. Vis. Exp.*, **1**, 60738.
17. Pluckthun,A. (2015) Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.*, **55**, 489–511.
18. Nygren,P.A. and Uhlen,M. (1997) Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.*, **7**, 463–469.
19. Skerra,A. (2000) Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.*, **13**, 167–187.
20. Liang,T., Chen,H., Yuan,J., Jiang,C., Hao,Y., Wang,Y., Feng,Z. and Xie,X.Q. (2021) IsAb: a computational protocol for antibody design. *Brief Bioinform*, **1**, bbab143.
21. Nuttall,S.D. and Walsh,R.B. (2008) Display scaffolds: protein engineering for novel therapeutics. *Curr. Opin. Pharmacol.*, **8**, 609–615.
22. Sha,F., Salzman,G., Gupta,A. and Koide,S. (2017) Monobodies and other synthetic binding proteins for expanding protein science. *Protein Sci.*, **26**, 910–924.
23. McMahon,C., Baier,A.S., Pascolutti,R., Wegrecki,M., Zheng,S., Ong,J.X., Erlandson,S.C., Hilger,D., Rasmussen,S.G.F., Ring,A.M. *et al.* (2018) Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol.*, **25**, 289–296.
24. Gebauer,M. and Skerra,A. (2020) Engineered protein scaffolds as next-generation therapeutics. *Annu. Rev. Pharmacol. Toxicol.*, **60**, 391–415.
25. Yang,Q., Li,B., Tang,J., Cui,X., Wang,Y., Li,X., Hu,J., Chen,Y., Xue,W., Lou,Y. *et al.* (2020) Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. *Brief Bioinform*, **21**, 1058–1068.
26. Ahmadi,M.K.B., Mohammadi,S.A., Makvandi,M., Mamouei,M., Rahmati,M., Dehghani,H. and Wood,D.W. (2020) Recent advances in the scaffold engineering of protein binders. *Curr. Pharm. Biotechnol.*, **22**, 878–891.
27. Zimmermann,I., Egloff,P., Hutter,C.A.J., Kuhn,B.T., Brauer,P., Newstead,S., Dawson,R.J.P., Geertsma,E.R. and Seeger,M.A. (2020) Generation of synthetic nanobodies against delicate proteins. *Nat. Protoc.*, **15**, 1707–1741.
28. Lv,Z., Cui,F., Zou,Q., Zhang,L. and Xu,L. (2021) Anticancer peptides prediction with deep representation learning features. *Brief Bioinform*, **1**, bbab008.
29. Gilbreth,R.N. and Koide,S. (2012) Structural insights for engineering binding proteins based on non-antibody scaffolds. *Curr. Opin. Struct. Biol.*, **22**, 413–420.
30. Owens,B. (2017) Faster, deeper, smaller-the rise of antibody-like scaffolds. *Nat. Biotechnol.*, **35**, 602–603.
31. Tian,P., Steward,A., Kudva,R., Su,T., Shilling,P.J., Nickson,A.A., Hollins,J.J., Beckmann,R., von Heijne,G., Clarke,J. *et al.* (2018) Folding pathway of an Ig domain is conserved on and off the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E11284–E11293.

32. Schoof,M., Faust,B., Saunders,R.A., Sangwan,S., Rezelj,V., Hoppe,N., Boone,M., Billesbolle,C.B., Puchades,C., Azumaya,C.M. *et al.* (2020) An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive spike. *Science*, **370**, 1473–1479.
33. Koenig,P.A., Das,H., Liu,H., Kummerer,B.M., Gohr,F.N., Jenster,L.M., Schiffelers,L.D.J., Tesfamariam,Y.M., Uchima,M., Wuerth,J.D. *et al.* (2021) Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape. *Science*, **371**, eabe6230.
34. Wang,Y., Zhang,S., Li,F., Zhou,Y., Zhang,Y., Wang,Z., Zhang,R., Zhu,J., Ren,Y., Tan,Y. *et al.* (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.*, **48**, D1031–D1041.
35. Cao,L., Goreshnik,I., Coventry,B., Case,J.B., Miller,L., Kozodoy,L., Chen,R.E., Carter,L., Walls,A.C., Park,Y.J. *et al.* (2020) De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science*, **370**, 426–431.
36. Feng,Z., Chen,M., Xue,Y., Liang,T., Chen,H., Zhou,Y., Nolin,T.D., Smith,R.B. and Xie,X.Q. (2021) MCCS: a novel recognition pattern-based method for fast track discovery of anti-SARS-CoV-2 drugs. *Brief. Bioinform.*, **22**, 946–962.
37. Yang,J., Zhang,Z., Yang,F., Zhang,H., Wu,H., Zhu,F. and Xue,W. (2021) Computational design and modeling of nanobodies toward SARS-CoV-2 receptor binding domain. *Chem. Biol. Drug Des.*, **98**, 1–18.
38. Silva,D.A., Yu,S., Ulge,U.Y., Spangler,J.B., Jude,K.M., Labao-Almeida,C., Ali,L.R., Quijano-Rubio,A., Ruterbusch,M., Leung,I. *et al.* (2019) De novo design of potent and selective mimics of IL-2 and IL-15. *Nature*, **565**, 186–191.
39. Spencer-Smith,R., Koide,A., Zhou,Y., Eguchi,R.R., Sha,F., Gajwani,P., Santana,D., Gupta,A., Jacobs,M., Herrero-Garcia,E. *et al.* (2017) Inhibition of RAS function through targeting an allosteric regulatory site. *Nat. Chem. Biol.*, **13**, 62–68.
40. Cao,L., Chang,H., Shi,X., Peng,C. and He,Y. (2016) Keratin mediates the recognition of apoptotic and necrotic cells through dendritic cell receptor DEC205/CD205. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 13438–13443.
41. Cao,L., Shi,X., Chang,H., Zhang,Q. and He,Y. (2015) pH-dependent recognition of apoptotic and necrotic cells by the human dendritic cell receptor DEC205. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 7237–7242.
42. Shahsavar,A., Stohler,P., Bourenkov,G., Zimmermann,I., Siegrist,M., Guba,W., Pinard,E., Sinning,S., Seeger,M.A., Schneider,T.R. *et al.* (2021) Structural insights into the inhibition of glycine reuptake. *Nature*, **591**, 677–681.
43. Feng,Z., Liang,T., Wang,S., Chen,M., Hou,T., Zhao,J., Chen,H., Zhou,Y. and Xie,X.Q. (2020) Binding characterization of GPCRs-modulator by molecular complex characterizing system (MCCS). *ACS Chem. Neurosci.*, **11**, 3333–3345.
44. Zuraw,B., Yasothan,U. and Kirkpatrick,P. (2010) Ecallantide. *Nat. Rev. Drug Discov.*, **9**, 189–190.
45. Morrison,C. (2019) Nanobody approval gives domain antibodies a boost. *Nat. Rev. Drug Discov.*, **18**, 485–487.
46. Sheridan,C. (2019) Llama-inspired antibody fragment approved for rare blood disorder. *Nat. Biotechnol.*, **37**, 333–334.
47. Sevy,A.M., Chen,M.T., Castor,M., Sylvia,T., Krishnamurthy,H., Ishchenko,A. and Hsieh,C.M. (2020) Structure- and sequence-based design of synthetic single-domain antibody libraries. *Protein Eng. Des. Sel.*, **33**, gzaa028.
48. Biswas,S., Khimulya,G., Alley,E.C., Esvelt,K.M. and Church,G.M. (2021) Low-N protein engineering with data-efficient deep learning. *Nat. Methods*, **18**, 389–396.
49. Wittmann,B.J., Johnston,K.E., Wu,Z. and Arnold,F.H. (2021) Advances in machine learning for directed evolution. *Curr. Opin. Struct. Biol.*, **69**, 11–18.
50. Wu,Z., Kan,S.B.J., Lewis,R.D., Wittmann,B.J. and Arnold,F.H. (2019) Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8852–8858.
51. Tang,J., Fu,J., Wang,Y., Li,B., Li,Y., Yang,Q., Cui,X., Hong,J., Li,X., Chen,Y. *et al.* (2020) ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief Bioinform*, **21**, 621–636.
52. Frappier,V. and Keating,A.E. (2021) Data-driven computational protein design. *Curr. Opin. Struct. Biol.*, **69**, 63–69.
53. Burnside,D., Schoenrock,A., Moteshareie,H., Hooshyar,M., Basra,P., Hajikarimlou,M., Dick,K., Barnes,B., Kazmirchuk,T., Jessulat,M. *et al.* (2019) In silico engineering of synthetic binding proteins from random amino acid sequences. *iScience*, **11**, 375–387.
54. Shin,J.E., Riesselman,A.J., Kollasch,A.W., McMahon,C., Simon,E., Sander,C., Manglik,A., Kruse,A.C. and Marks,D.S. (2021) Protein design and variant prediction using autoregressive generative models. *Nat. Commun.*, **12**, 2403.
55. Wang,X.F., Gao,P., Liu,Y.F., Li,H.F. and Lu,F. (2020) Predicting thermophilic proteins by machine learning. *Curr. Bioinform*, **15**, 493–502.
56. Rothbauer,U. (2018) Speed up to find the right ones: rapid discovery of functional nanobodies. *Nat. Struct. Mol. Biol.*, **25**, 199–201.
57. Lima,W.C., Gasteiger,E., Marcatili,P., Duek,P., Bairoch,A. and Cosson,P. (2020) The ABCD database: a repository for chemically defined antibodies. *Nucleic Acids Res.*, **48**, D261–D264.
58. Carvalho,M.B., Molina,F. and Felicori,L.F. (2019) Yvis: antibody high-density alignment visualization and analysis platform with an integrated database. *Nucleic Acids Res.*, **47**, W490–W495.
59. Raybould,M.I.J., Marks,C., Lewis,A.P., Shi,J., Bujotzek,A., Taddese,B. and Deane,C.M. (2020) Thera-SAbDab: the therapeutic structural antibody database. *Nucleic Acids Res.*, **48**, D383–D388.
60. Wilton,E.E., Opyr,M.P., Kailasam,S., Kothe,R.F. and Wieden,H.J. (2018) sdAb-DB: the single domain antibody database. *ACS Synth. Biol.*, **7**, 2480–2484.
61. Adolf-Bryfogle,J., Xu,Q., North,B., Lehmann,A. and Dunbrack,R.L. (2015) PyIgClassify: a database of antibody CDR structural classifications. *Nucleic Acids Res.*, **43**, D432–D438.
62. Szklarczyk,D., Gable,A.L., Nastou,K.C., Lyon,D., Kirsch,R., Pyysalo,S., Doncheva,N.T., Legeay,M., Fang,T., Bork,P. *et al.* (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.*, **49**, D605–D612.
63. Oughtred,R., Stark,C., Breitkreutz,B.J., Rust,J., Boucher,L., Chang,C., Kolas,N., O'Donnell,L., Leung,G., McAdam,R. *et al.* (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, D529–D541.
64. Basha,O., Shpringer,R., Argov,C.M. and Yeger-Lotem,E. (2018) The DifferentialNet database of differential protein-protein interactions in human tissues. *Nucleic Acids Res.*, **46**, D522–D526.
65. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database: 2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
66. Orchard,S., Ammari,M., Aranda,B., Breuza,L., Briganti,L., Broackes-Carter,F., Campbell,N.H., Chavali,G., Chen,C., del-Toro,N. *et al.* (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
67. Chen,Z., Zhao,P., Li,C., Li,F., Xiang,D., Chen,Y.Z., Akutsu,T., Daly,R.J., Webb,G.I., Zhao,Q. *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.
68. Li,F., Chen,J., Leier,A., Marquez-Lago,T., Liu,Q., Wang,Y., Revote,J., Smith,A.I., Akutsu,T., Webb,G.I. *et al.* (2020) DeepCleave: a deep learning predictor for caspase and matrix metalloprotease substrates and cleavage sites. *Bioinformatics*, **36**, 1057–1065.
69. Mei,S., Li,F., Leier,A., Marquez-Lago,T.T., Giam,K., Croft,N.P., Akutsu,T., Smith,A.I., Li,J., Rossjohn,J. *et al.* (2020) A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief Bioinform*, **21**, 1119–1135.
70. Feldwisch,J. and Tolmachev,V. (2012) Engineering of affibody molecules for therapy and diagnostics. *Methods Mol. Biol.*, **899**, 103–126.
71. Rothe,C. and Skerra,A. (2018) Anticalin proteins as therapeutic agents in human diseases. *BioDrugs*, **32**, 233–243.
72. Griffiths,K., Dolezal,O., Cao,B., Nilsson,S.K., See,H.B., Pfleger,K.D.G., Roche,M., Gorry,P.R., Pow,A., Viduka,K. *et al.*

(2016) I-bodies, human single domain antibodies that antagonize chemokine receptor CXCR4. *J. Biol. Chem.*, **291**, 12641–12657.
73. Hantschel,O., Biancalana,M. and Koide,S. (2020) Monobodies as enabling tools for structural and mechanistic biology. *Curr. Opin. Struct. Biol.*, **60**, 167–174.
74. Lee,S.C., Park,K., Han,J., Lee,J.J., Kim,H.J., Hong,S., Heu,W., Kim,Y.J., Ha,J.S., Lee,S.G. *et al.* (2012) Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 3299–3304.
75. Hanna,R., Cardarelli,L., Patel,N., Blazer,L.L., Adams,J.J. and Sidhu,S.S. (2020) A phage-displayed single-chain fab library optimized for rapid production of single-chain IgGs. *Protein Sci*, **29**, 2075–2084.
76. Mak,L., Marcus,D., Howlett,A., Yarova,G., Duchateau,G., Klaffke,W., Bender,A. and Glen,R.C. (2015) Metrabase: a cheminformatics and bioinformatics database for small molecule transporter data analysis and (Q)SAR modeling. *J. Cheminform*, **7**, 31.
77. Kovaleva,M., Ferguson,L., Steven,J., Porter,A. and Barelle,C. (2014) Shark variable new antigen receptor biologics: a novel technology platform for therapeutic drug development. *Expert. Opin. Biol. Ther.*, **14**, 1527–1539.
78. Wuo,M.G. and Arora,P.S. (2018) Engineered protein scaffolds as leads for synthetic inhibitors of protein-protein interactions. *Curr Opin Chem Biol*, **44**, 16–22.
79. Lv,Z., Wang,P., Zou,Q. and Jiang,Q. (2020) Identification of sub-golgi protein localization by use of deep representation learning features. *Bioinformatics*, **36**, 5600–5609.
80. Tian,P. and Best,R.B. (2020) Exploring the sequence fitness landscape of a bridge between protein folds. *PLoS Comput. Biol.*, **16**, e1008285.
81. Tian,P., Louis,J.M., Baber,J.L., Aniana,A. and Best,R.B. (2018) Co-Evolutionary Fitness Landscapes for Sequence Design. *Angew. Chem. Int. Ed. Engl.*, **57**, 5674–5678.
82. Jakhar,R., Dangi,M., Khichi,A. and Chhillar,A.K. (2020) Relevance of molecular docking studies in drug designing. *Curr Bioinform*, **15**, 270–278.
83. Sayers,E.W., Beck,J., Brister,J.R., Bolton,E.E., Canese,K., Comeau,D.C., Funk,K., Ketter,A., Kim,S., Kimchi,A. *et al.* (2020) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **48**, D9–D16.
84. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
85. Burley,S.K., Bhikadiya,C., Bi,C., Bittrich,S., Chen,L., Crichlow,G.V., Christie,C.H., Dalenberg,K., Di Costanzo,L., Duarte,J.M. *et al.* (2021) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.*, **49**, D437–D451.
86. UniProt,C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
87. Binz,H.K., Amstutz,P. and Pluckthun,A. (2005) Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.*, **23**, 1257–1268.
88. Lofblom,J., Frejd,F.Y. and Stahl,S. (2011) Non-immunoglobulin based protein scaffolds. *Curr. Opin. Biotechnol.*, **22**, 843–848.
89. Gebauer,M. and Skerra,A. (2019) Engineering of binding functions into proteins. *Curr. Opin. Biotechnol.*, **60**, 230–241.
90. Scott,C.P., Abel-Santos,E., Wall,M., Wahnon,D.C. and Benkovic,S.J. (1999) Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 13638–13643.
91. Russo,A., Aiello,C., Grieco,P. and Marasco,D. (2016) Targeting "undruggable" proteins: design of synthetic cyclopeptides. *Curr. Med. Chem.*, **23**, 748–762.
92. Dunbar,J., Krawczyk,K., Leem,J., Baker,T., Fuchs,A., Georges,G., Shi,J. and Deane,C.M. (2014) SAbDab: the structural antibody database. *Nucleic Acids Res.*, **42**, D1140–D1146.
93. Ma,Y., Ding,Y., Song,X., Ma,X., Li,X., Zhang,N., Song,Y., Sun,Y., Shen,Y., Zhong,W. *et al.* (2020) Structure-guided discovery of a single-domain antibody agonist against human apelin receptor. *Sci. Adv.*, **6**, eaax7379.
94. McMahon,C., Staus,D.P., Wingler,L.M., Wang,J., Skiba,M.A., Elgeti,M., Hubbell,W.L., Rockman,H.A., Kruse,A.C. and Lefkowitz,R.J. (2020) Synthetic nanobodies as angiotensin receptor blockers. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 20284–20291.
95. Smolarczyk,T., Roterman-Konieczna,I. and Stapor,K. (2020) Protein secondary structure prediction: a review of progress and directions. *Curr Bioinform*, **15**, 90–107.
96. Dhiraviam,K.N., Balasubramanian,S. and Jayavel,S. (2018) Indole alkaloids as new leads for the design and development of novel DPP-IV inhibitors for the treatment of diabetes. *Curr Bioinform*, **13**, 157–169.
97. Ovchinnikov,S., Park,H., Varghese,N., Huang,P.S., Pavlopoulos,G.A., Kim,D.E., Kamisetty,H., Kyrpides,N.C. and Baker,D. (2017) Protein structure determination using metagenome sequence data. *Science*, **355**, 294–298.
98. Kuhlman,B. and Bradley,P. (2019) Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.*, **20**, 681–697.
99. Koga,N., Tatsumi-Koga,R., Liu,G., Xiao,R., Acton,T.B., Montelione,G.T. and Baker,D. (2012) Principles for designing ideal protein structures. *Nature*, **491**, 222–227.
100. Abbass,J. and Nebel,J.C. (2020) Rosetta and the journey to predict proteins structures, 20 years on. *Curr. Bioinform.*, **15**, 611–628.
101. Yang,J., Anishchenko,I., Park,H., Peng,Z., Ovchinnikov,S. and Baker,D. (2020) Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 1496–1503.
102. Abriata,L.A. and Dal Peraro,M. (2021) State-of-the-art web services for de novo protein structure prediction. *Brief Bioinform*, **22**, bbaa139.
103. Gross,G.G., Junge,J.A., Mora,R.J., Kwon,H.B., Olson,C.A., Takahashi,T.T., Liman,E.R., Ellis-Davies,G.C., McGee,A.W., Sabatini,B.L. *et al.* (2013) Recombinant probes for visualizing endogenous synaptic proteins in living neurons. *Neuron*, **78**, 971–985.
104. Ring,A.M., Manglik,A., Kruse,A.C., Enos,M.D., Weis,W.I., Garcia,K.C. and Kobilka,B.K. (2013) Adrenaline-activated structure of beta2-adrenoceptor stabilized by an engineered nanobody. *Nature*, **502**, 575–579.
105. Wallberg,H., Orlova,A., Altai,M., Hosseinimehr,S.J., Widstrom,C., Malmberg,J., Stahl,S. and Tolmachev,V. (2011) Molecular design and optimization of 99mTc-labeled recombinant affibody molecules improves their biodistribution and imaging properties. *J Nucl Med*, **52**, 461–469.
106. Theurillat,J.P., Dreier,B., Nagy-Davidescu,G., Seifert,B., Behnke,S., Zurrer-Hardi,U., Ingold,F., Pluckthun,A. and Moch,H. (2010) Designed ankyrin repeat proteins: a novel tool for testing epidermal growth factor receptor 2 expression in breast cancer. *Mod Pathol*, **23**, 1289–1297.
107. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403–410.
108. Bond,C.J., Marsters,J.C. and Sidhu,S.S. (2003) Contributions of CDR3 to V H H domain stability and the design of monobody scaffolds for naive antibody libraries. *J Mol Biol*, **332**, 643–655.
109. Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.*, **33**, 2302–2309.
110. Pandit,S.B. and Skolnick,J. (2008) Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. *BMC Bioinformatics*, **9**, 531.
111. Mukherjee,S. and Zhang,Y. (2009) MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic Acids Res.*, **37**, e83.
112. Lancet,T. (2019) ICD-11. *Lancet*, **393**, 2275.
113. Lancet,T. (2018) ICD-11: in praise of good data. *Lancet Infect Dis*, **18**, 813.
114. Zarin,D.A., Fain,K.M., Dobbins,H.D., Tse,T. and Williams,R.J. (2019) 10-year update on study results submitted to ClinicalTrials.gov. *N Engl J Med*, **381**, 1966–1974.
115. Yin,J., Sun,W., Li,F., Hong,J., Li,X., Zhou,Y., Lu,Y., Liu,M., Zhang,X., Chen,N. *et al.* (2020) VARIDT 1.0: variability of drug transporter database. *Nucleic Acids Res.*, **48**, D1042–D1050.
116. Song,M., Guo,H., Chen,H. and Hu,H. (2016) Characteristics of anticancer drug studies registered on the Chinese clinical trial registry (ChiCTR) from 2007 to 2015. *J Evid Based Med*, **9**, 59–68.

117. Goldacre,B., DeVito,N.J., Heneghan,C., Irving,F., Bacon,S., Fleminger,J. and Curtis,H. (2018) Compliance with requirement to report results on the EU clinical trials register: cohort study and web resource. *BMJ*, **362**, k3218.
118. Li,Y.H., Yu,C.Y., Li,X.X., Zhang,P., Tang,J., Yang,Q., Fu,T., Zhang,X., Cui,X., Tu,G. *et al.* (2018) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.*, **46**, D1121–D1127.
119. Yin,J., Li,F., Zhou,Y., Mou,M., Lu,Y., Chen,K., Xue,J., Luo,Y., Fu,J., He,X. *et al.* (2021) INTEDE: interactome of drug-metabolizing enzymes. *Nucleic Acids Res.*, **49**, D1233–D1243.