



Review

Florian J. Gisdon, Josef P. Kynast, Merve Ayyildiz, Anna V. Hine, Andreas Plückthun and Birte Höcker*

Modular peptide binders – development of a predictive technology as alternative for reagent antibodies

<https://doi.org/10.1515/hsz-2021-0384>

Received September 30, 2021; accepted January 11, 2022;

published online January 28, 2022

Abstract: Current biomedical research and diagnostics critically depend on detection agents for specific recognition and quantification of protein molecules. Monoclonal antibodies have been used for this purpose over decades and facilitated numerous biological and biomedical investigations. Recently, however, it has become apparent that many commercial reagent antibodies lack specificity or do not recognize their target at all. Thus, synthetic alternatives are needed whose complex designs are facilitated by multidisciplinary approaches incorporating experimental protein engineering with computational modeling. Here, we review the status of such an engineering endeavor based on the modular armadillo repeat protein scaffold and discuss challenges in its implementation.

Keywords: affinity reagent; armadillo repeat proteins; computational design; directed evolution; library generation; protein-peptide interface.

Florian J. Gisdon, Josef P. Kynast and Merve Ayyildiz contributed equally to this work.

*Corresponding author: **Birte Höcker**, Department of Biochemistry, University of Bayreuth, Universitätsstr. 30, D-95447 Bayreuth, Germany, E-mail: birte.hoecker@uni-bayreuth.de. <https://orcid.org/0000-0002-8250-9462>

Florian J. Gisdon, Josef P. Kynast and Merve Ayyildiz, Department of Biochemistry, University of Bayreuth, D-95447 Bayreuth, Germany, E-mail: florian.gisdon@uni-bayreuth.de (F.J. Gisdon), josef.kynast@uni-bayreuth.de (J.P. Kynast), Merve.Ayyildiz@uni-bayreuth.de (M. Ayyildiz)

Anna V. Hine, College of Health and Life Sciences, Aston University, Birmingham B4 7ET, UK, E-mail: a.v.hine@aston.ac.uk

Andreas Plückthun, Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland, E-mail: plueckthun@bioc.uzh.ch

Introduction

Current biomedical research relies on the use of reagent antibodies to detect biomolecules in medical diagnostics and basic life science research. The development of a chimeric antibody in 1984 (Morrison et al. 1984; Neuberger et al. 1984) as a first recombinant antibody opened new possibilities for the development of therapeutics and applications as affinity reagents. Recombinant production allows one to define the sequence, which is important as it ensures the reproducibility of experiments and reliability of results. In contrast, the sequence of monoclonal antibodies is not directly known, but can be obtained via protein sequencing, though it is time-consuming and costly. The efficient production and sophisticated technology of monoclonal antibodies that are derived by immunization is certainly a reason for their prevalence as specific binders in biological sciences. But the use of animal-derived antibodies has been more and more brought into question. On the one hand information about monoclonal antibodies, which are derived from hybridoma cell lines, can get lost due to cell line death or gene loss (Bradbury and Plückthun 2015). But apart from that, increasing awareness arose from the observation that animal-derived antibodies are varying between separate batches and often lack distinct specificity, which affects experimental reproducibility (Baker 2015). To address this issue, DNA sequencing can be applied and in fact, a recombinant production should then be possible. While this is technologically feasible, it is unfortunately not routinely done for commercially available reagent antibodies, which is likely due to commercial reasons (Bradbury and Plückthun 2015).

Consequently, in an interdisciplinary meeting in 2019, 35 years after the first recombinant antibody had been engineered, the development and use of animal-free recombinant antibodies were discussed with the objective to foster their increased use in basic research (Groff et al. 2020). Still, conventional antibodies are widely used in research applications, but antibodies with poor specificities or the

lack of reproducibility led to the development of alternative affinity reagents, which can be produced recombinantly and hence ensure their reliability (Groff et al. 2015). Recombinant production requires one to know the sequence of the reagent, and thus makes experimental results transparent and reproducible. Furthermore, recombinantly produced affinity reagents are truly monoclonal but can also be made polyclonal by using exactly defined pools. This procedure even provides knowledge about the full composition of the reagent mixture.

The first recombinant affinity reagents have been immunoglobulin derivatives. Immunoglobulins consist of a tail region, the Fc fragment, which interacts with cellular receptors, and a Fab fragment, which binds to antigens. In 1989 the first Fc-fusion protein was described, a fusion of the Fc fragment with the cell-surface glycoprotein CD4 (Capon et al. 1989). In the meantime, Fc-fusion proteins have been used as reagents for immunotherapy (to harvest their long half-lives) and laboratory research (to exploit detection reagents against the Fc part) (Duivelshof et al. 2021; Flanagan et al. 2007; Liu and Yu 2016), and the Fab regions have been utilized as recombinant affinity reagents (Conroy et al. 2017; Shih et al. 2012). However, the structure of antibodies entails technical challenges such as production in eukaryotic cells to obtain the required disulfide pattern and/or glycosylation (Gebauer and Skerra 2020). Such considerations have supported the development of alternative binding reagents which are not based on the immunoglobulin fold. First alternatives were e.g., based on natural folds such as fibronectin (Koide et al. 1998) or lipocalin (Beste et al. 1999), leading to monobodies or anticalins as designed affinity reagents. For both affinity reagents loop regions can be randomized to generate different variants, which can be selected for specific targets. However, a change of the scaffold has just been the first step. The mentioned affinity reagents are restricted by their size and variability in their binding mode. For a better adjustability and for control of the binding properties designed repeat proteins have been considered as scaffolds. Designed ankyrin repeat proteins (DARPs), for example, have the advantage to be fully characterized, and their size can be adapted by addition of further repeats (Binz et al. 2004; Forrer et al. 2003; Plückthun 2015). The repeats can also be easily randomized, which allows for a great variability that can be screened (see below) to find good binding reagents. Further, the binding site is slightly concave, which is favorable for binding large epitopes. However, DARPs have to be developed anew for every target, nevertheless, they are used as innovative affinity reagents (Schilling et al. 2021).

A fundamentally different concept, which is also based on a repeat protein scaffold, is currently investigated

within the collaborative ‘Predictive Reagent Antibody Replacement Technology’ (for PRE-ART) project. Here, the alternative affinity reagent is a designed armadillo repeat protein (dArmRP, see Figure 1), which can be varied in length of the concave binding surface, analogous to DARPs. However, the modularity of dArMRPs gives them an additional unique feature, as each internal repeat harbors binding capabilities for exactly two adjacent amino acid residues of a target. Furthermore, the distance between the repeats is optimized to match the periodicity of a peptide chain, so that the dArmRP can be applied to bind linear epitopes (Reichen et al. 2014). By designing different repeat modules, with specificities for all individual amino acids, a universal toolkit will be created from which desired binders can easily be assembled. This idea radically rethinks the established concept of affinity reagents and will affect a broad user base, as the recognition of linear epitopes is fundamental in many research applications, for instance protein purification with affinity tags or the recognition of unstructured regions such as found on western blots or intrinsically disordered proteins. Further, such unstructured regions are often targets for post-translational modifications such as phosphorylation and play an important role in the function of proteins (Dyson and Wright 2005; Liu et al. 2020).

This growing number of applications strengthens the need for robust and well-defined affinity reagents, which are less cost- and time-consuming in their production compared to commonly used reagent antibodies. This is especially important since many commercial reagent antibodies lack specificity or do not recognize their target at all. The modular dArMRPs define an innovative technology that fully reexamines the concept of existing affinity reagents and promises to revolutionize their applications.

Armadillo repeat proteins are modular scaffolds for peptide recognition

The natural armadillo repeat protein (ArmRP) scaffold harbors unique and useful features necessary for its development into recombinant affinity reagents. It is comprised of homologous structural units that stack to form an elongated, rigid structure. Crystal structures show that natural ArmRPs bind stretched peptides of up to six amino acids (Conti and Kuriyan 2000; Conti et al. 1998; Graham et al. 2000). This binding of peptides in extended conformation reveals a conserved modular recognition mechanism which is a key feature of the ArmRP scaffold. Every second main

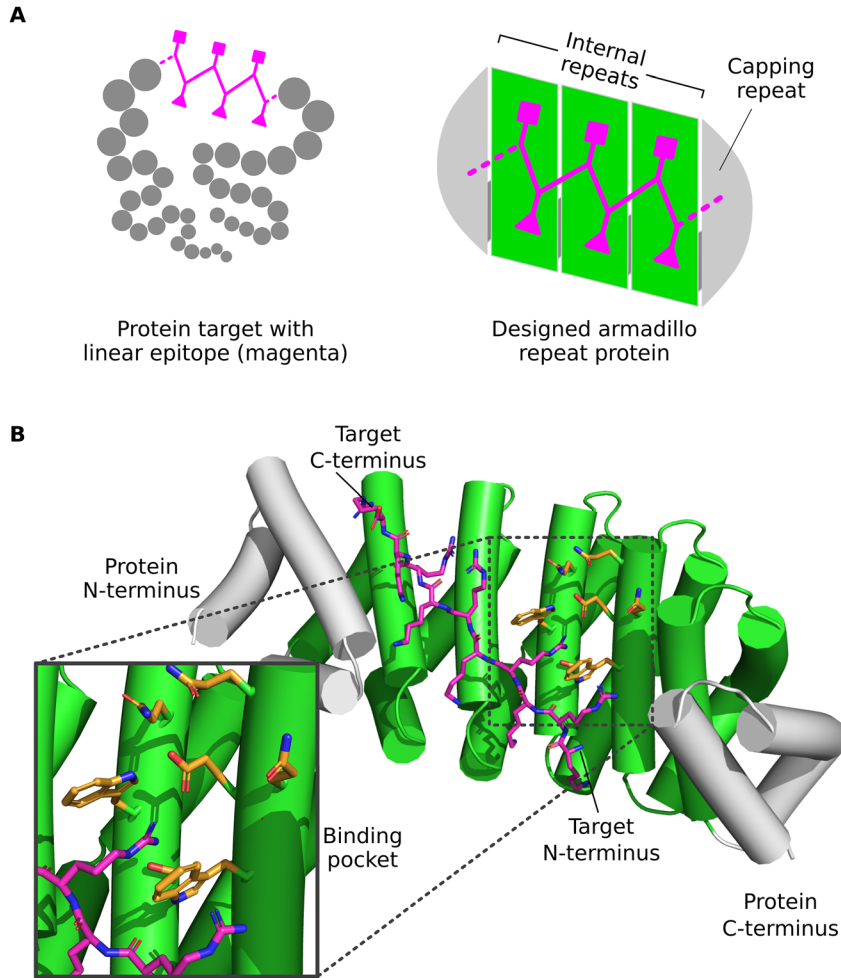


Figure 1: The modular nature of designed armadillo repeat proteins (dArmRPs). (A) Dipeptide units of a linear peptide stretch are bound in a modular fashion by dArmRPs. (B) The crystal structure of a dArmRP in complex with its target (PDB-ID: 5AEI) shows the modularity of binding of the extended peptide (magenta, as sticks) to the repeat modules of the protein (green, as cartoon). The residues of one arginine binding pocket are highlighted (orange, as sticks).

chain peptide bond of the target is held in place by a conserved asparagine residue on every ArmRP repeat. These interactions provide a general affinity and secure the regularity of the binding interactions. Each ArmRP repeat unit further binds two adjacent amino acid side chains in the target sequence in a specific manner (Figure 1).

These features were enhanced and regularized in iterative rounds of engineering. Using a consensus approach followed by computational and structural engineering for stability yielded a highly stable dArmRP, which consists of perfectly stackable repeats and optimized cap structures (Alfarano et al. 2012; Madhurantakam et al. 2012; Parmegiani et al. 2008). Each repeat is 42 amino acids long and forms three alpha-helices. The assembled repeats again form an extended superhelical structure. Reichen et al. (2016) analyzed the variation in curvature of natural ArmRPs and identified a repeat pair in yeast importin-alpha with the ideal curvature geometry for optimal binding of an extended peptide. Based on binding pockets from importin-alpha, a dArmRP could be built that has picomolar affinity to its target peptide of alternating lysine and arginine residues

(Hansen et al. 2016). The crystal structure of this protein, built from five identical repeats and N- and C-terminal caps in complex with a $(KR)_5$ peptide, confirmed the regular binding mode (Figure 1B). It lays the groundwork for the design of tailored binders with specific affinities by the assembly of dipeptide-specific dArmRP modules.

For the development of a diverse set of binding modules for different amino acids, a consistent design and testing approach is crucial for success as we discuss below. Furthermore, it is important that other binding modes are eliminated as it had been observed that repetitive sequences lead to register shifts and flipping of peptides during selections from libraries, which affects the investigation of binding specificities (Ernst et al. 2020). To prevent the peptide from binding in undesired orientations, a lock was incorporated into the dArmRP by grafting a hydrophobic binding site observed in beta-catenin onto the dArmRP, thereby locking the peptide with the complementary sequence in place. The interaction of the lock was improved by mutual optimization of the pocket and the bound peptide, which were then confirmed by X-ray

crystallography. The lock could further be moved from the N-terminus of the dArmRP to its middle nicely highlighting the modularity of the system (Ernst et al. 2020).

With stability and modular binding of dArmRPs established and with an efficient locking system in place, the main goal is now to develop modules that can bind any other amino acid including negatively charged or even phosphorylated ones. Clearly, further adjustments of the dArmRP scaffold will also be necessary as neighboring binding pockets and combination of modules might have effects on the overall binder. However, the current challenge is to identify sequences that form binding pockets for other amino acids and thereby design new binding modules. Here, a consistent strategy to reduce the number of theoretical binding pocket sequences to an experimentally testable level is the key to success.

Experimental strategies in the design of specific dArmRP modules

The repeat units of dArmRPs bind two adjacent amino acids in an alternating orientation (Figure 1). Originating from the importin-alpha framework one binding pocket is specific for arginine and the other one is specific for lysine (Hansen et al. 2016). The specificity of each pocket has to be adjusted to recognize other amino acids by mutating binding pocket residues. For an efficient search of specific binding pockets, DNA library selection technologies play a major role. These techniques allow to rapidly screen large numbers of DNA sequences encoding for the target protein that are randomized in regions responsible for the desired interaction. A complete randomization of a dArmRP module, however, is not useful. First, only a small fraction of residues is in direct contact with the ligand side chain of the target peptide. Second, uncontrolled randomization will incorporate unwanted termination codons. And third, due to the assignment of 64 codons to 20 canonical amino acids and termination codons, the distribution of amino acids will heavily differ at each position of randomization. Hence, the probability for certain amino acids to occur will be drastically reduced and create a bias. Additionally, the total number of sequences necessary to exhaustively screen a library will exponentially increase per randomized amino acid position.

A solution to these difficulties and to reduce the number of DNA sequences necessary for exhaustive screening is the use of MAX randomization as a non-degenerate saturation mutagenesis technology (Hughes et al. 2003). This technology allows one to build libraries with exactly 20 codons (one for each amino acid) or a desired subset of those for the randomized position. As a related technique

ProxiMAX even allows to saturate multiple contiguous codons in a non-degenerate manner (Ashraf et al. 2013). Both methods require no specialized chemistry, reagents, or equipment. Ultimately, the use of the MAX techniques allows to generate DNA libraries without amino acid bias, termination codons, and degeneracy. Limiting both library size and degeneracy is critical to maximizing the output from the applied screening technology.

Three main selection technologies exist that could be used for the selection of dArmRP libraries: phage display, ribosome display, and yeast display. Because of the starting consensus scaffold being dominated by importin-alpha, the libraries are heavily biased to bind positively charged peptides, which creates difficulties during panning. As ribosome display uses highly negatively charged mRNA molecules and filamentous phages are equally negatively charged, it is not possible to select specific binding to the positively charged peptides. In contrast, selections by yeast display can be successfully performed, as the yeast surface is apparently not as negatively charged.

During selections of pockets for individual amino acids it is key that the peptides bind specifically and efficiently to the dArmRPs. Due to the repetitive nature of the dArmRP binding pockets the target peptide can bind in different registers. To avoid flipping or sliding of the peptide it is important to provide a binding pocket that locks the peptide into place. This was achieved by grafting a binding site from β -catenin into the dArmRP as described above (Ernst et al. 2020). The lock allows that selections can now be focused onto the binding pocket residues to the new target side chain to which specificity should be achieved.

Selection by yeast display is a very powerful technique and many different variants can be sorted in a high-throughput manner. Nonetheless, even with this technique only a library of a certain size can be screened. While library design by MAX randomization is a huge advantage as it allows particular residue types in predefined positions, screening of these libraries is still time-consuming. Therefore, it is useful to focus the libraries further to the most likely variants. Here, computational techniques can help to predict precise mixtures of amino acids for each position of randomization.

Computational strategies in the design of specific dArmRP modules

The modularity of the dArmRP scaffold allows for the individual design of a single pocket at a time. However, there is still an enormous number of residue combinations and degrees of freedom that need to be sampled. Therefore, a

computational pre-selection of possible binding modes is useful and necessary to enable efficient experimental screening as described above.

The computational sampling of a very large number of combinations and degrees of freedom is challenging as well, though the past decade has seen significant improvements in the development and application of computational methods for protein design (Lechner et al. 2018). With algorithmic improvements and technological progress in computer hardware, new protein design approaches yielded increased accuracy and efficiency by allowing more flexibility and by applying simultaneous sampling of multiple sequences (Friedland et al. 2008; Murphy et al. 2012; Saunders and Baker 2005; Yin et al. 2007). In addition to the applied flexibility, different design objectives like creating single-state, multi-state, or ensemble-based designs influence the quality of the computational predictions.

One powerful method for sampling flexibility in computational protein design is Molecular Dynamics (MD) that has proven to provide valuable insights on protein stability, dynamics, and macromolecular interactions (Simonson et al. 2020). Technological advances such as parallelization on graphics processing units (GPU) have significantly accelerated MD calculations. Although the high computational cost is still a limiting factor, MD simulations of microseconds on a single GPU for protein systems such as dArmRPs are achievable within several days (Lazim et al. 2020). However, the design of binding pockets requires to sample many different combinations of amino acids, which spans an enormous combinatorial search space. For the evaluation of such a large amount of different variants, provable algorithms, including Branch and Bound, Dead-End Elimination, and Dynamic Programming, that have been successfully applied to protein design problems with backbone flexibility are promising developments for efficient calculations (Desmet et al. 1992; Gordon and Mayo 1999; Jou et al. 2016; Leaver-Fay et al. 2005; Ojewole et al. 2018). Also, deep learning techniques have experienced a large gain in interest in protein redesign since novel deep learning architectures achieve extraordinary prediction results in various fields due to clever model design and effective pattern recognition (Jumper et al. 2021; Krizhevsky et al. 2017). Thus, prediction of protein design features might be applicable on multiple sequences in a drastically smaller timescale. However, many state-of-the-art machine learning models, especially deep learning models, have not been extensively explored for protein design applications so far (Gao et al. 2020; Wang et al. 2018; Xu et al. 2020).

Nonetheless, computational strategies are often used as a complementary approach to experimental methods

since experimental work is time-consuming and expensive (Chen and Keating 2012; Ernst et al. 2020; Liang et al. 2021). Within the multidisciplinary approach of PRE-ART, computational tools with diverse features help to characterize existing and to design new binding modules. For the characterization of new or existing binding pockets, different computational options are available. Tools can be used to screen the possible sequence space with methods such as the non-exhaustive screening and scoring protocols FastDesign (Loshbaugh and Kortemme 2020; Maguire et al. 2021) and coupled moves (Ollikainen et al. 2015) included in the software suite Rosetta. FastDesign performs iterations of side chain repacking and global minimization to find energy minima while exchanging predefined residues within the sequence. Coupled moves, however, alters backbone and sidechain conformations as well as the sequence at a time, to allow for more effective sampling. Further, several well-established computational methods, including *flex ddG* and *Branch and Bound Over K** (*BBK**) algorithms implemented in the Rosetta and Osprey protein design suites, respectively, allow to specifically target single binder sequences with exchanges in one residue position (Barlow et al. 2018; Ojewole et al. 2018). The *flex ddG* protocol incorporates backrub motion to accurately calculate binding affinity changes upon mutation. The *BBK** algorithm efficiently evaluates the partition function to calculate the binding affinity, while additionally allowing for continuous flexibility. Complementing these algorithms, MD simulations can support the analysis of the influence of mutations on the dynamics and the protein-ligand interactions of the system.

To predict promising mutations in a binding pocket in the first place that potentially develops a specific binding ability for the desired peptide, the software suite ATLIGATOR has been developed (Kynast et al. 2022). It is based on a knowledge-based approach that extracts pairwise interactions from existing structures to be used in the design of new binding pockets. Furthermore, it incorporates the detection of frequent interaction groups for specific amino acid side chains. Subsequent evaluation of the suggested binding pockets from ATLIGATOR can be performed by algorithms such as *flex ddG* or *BBK**, which can be complemented with MD simulations. The combination of the described methods results in a detailed understanding of the new binding pocket candidates. Hence, even if the computational prediction of exact binder sequences is not entirely possible, the multidisciplinary PRE-ART approach established a feedback loop to use the findings from computational modeling for the design of focused libraries for experimental screening (Figure 2).

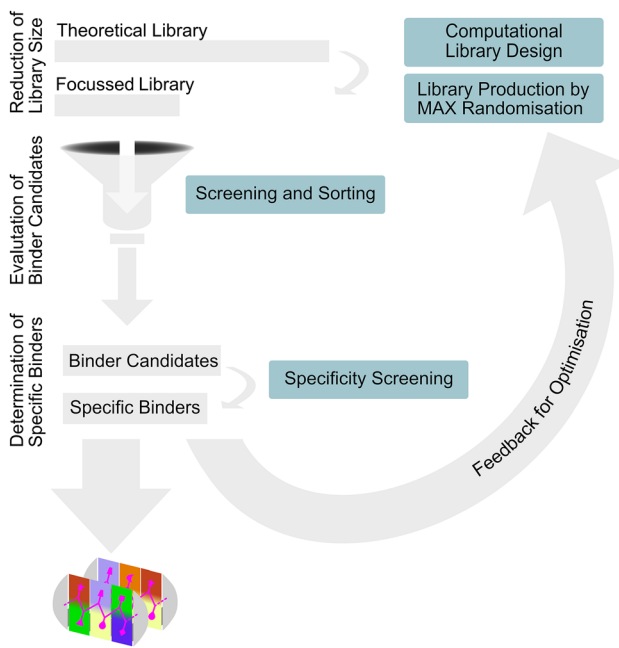


Figure 2: Workflow in the engineering of binding modules. Libraries are designed, synthesized, screened, and evaluated, providing feedback to the input techniques. The overall loop creates an ensemble of binding modules that can later be assembled to recognize predefined target peptides.

Complementarity of experimental and computational design

The design of specific protein-protein or protein-peptide interactions with experimental screening and selection methods as well as computational modeling and prediction tools has progressed significantly. Experimental screening of DNA libraries with molecular display technologies (Levin and Weiss 2006) allows one to sample millions of sequences at once. When combining fluorescence activated cell sorting (FACS) with bacterial or yeast display approaches, cells can be sorted according to desired features. However, experimental screening methods suffer from exponentially growing complexity, the more residue positions are randomized. The use of techniques such as MAX randomization optimizes the codon selection to the minimum required and allows to define limited sets of amino acids for randomized positions (Hughes et al. 2003). Thus, the total number of sequences to screen for a complete coverage of desired amino acid sequences is minimized and the effective screening capacity is drastically increased.

Still, a theoretical library for complete randomization of a binding pocket quickly exceeds screening capacities. Thus, sequence space has to be reduced to a relevant set of

sequences in the randomization. To screen only the relevant sequence space computational modeling can be used to exclude noninvolved positions and unfavorable mutations. An early attempt was by Voigt et al. (2001) who computationally focused a library and successfully selected proteins with increased stability. Also, in protein-protein interaction engineering, several groups have used computational design to focus libraries to select sequences compatible with the target fold that were screened for function later-on (Guntas et al. 2010; Hayes et al. 2002; Treynor et al. 2007). With increasing computational power and new protein modeling and design algorithms in the fields of deterministic (reviewed in Gainza et al. [2016]) or heuristic solving (reviewed for the Rosetta Suite in Kuhlman [2019]) as well as machine learning (reviewed in AlQuraishi [2021]) the potential to computationally focus libraries increased heavily.

The prediction of protein structures and stability are used successfully as a less cost- and labor-intense alternative to experimental methods. Even though the prediction of protein complex structures and their binding free energy is still not feasible for “bigger systems” in many cases, current software protocols can give crucial insights into those events (Barlow et al. 2018; Ojewole et al. 2018). Thus, functionally important positions can be identified or amino acid properties with potentially positive effects can be defined to reduce the size of the relevant search space. An incorporation of this knowledge into a designed library for experimental screening allows one to screen a bigger part of potentially advantageous sequences and to sort out disfavored sequences with a higher probability. Hence, the interplay of computational and experimental techniques leads to a higher likelihood to find variants with an improvement of the desired functionality. In the case of the PRE-ART project individual binding pockets are designed in dArmRPs, which detect and discriminate single amino acid side chains with high specificity. Randomization of all possible interacting positions would lead to a search space that largely exceeds screening capacities, which is why complementation with computational methods to design focused libraries is highly beneficial.

The precise objective of such a library design process for subsequent experimental screening is not immediately obvious. Possible priorities in creating such a library can be the inclusion of the best predicted sequences, the most frequently predicted sequences or a preferably high sequence diversity (Chen and Keating 2012), as well as sequences with highest affinity versus specificity. A reasonable choice would be to focus on *affinity* with computational selection and on *specificity* with subsequent experimental screening. Additionally, a library can be designed by scanning and scoring relevant shares of

the sequence space (Barlow et al. 2018; Gainza et al. 2013) or by considering interaction motifs found in natural proteins (Kynast et al. 2022). The results of screening such a focused library will potentially detect more desired binder sequences.

These variants of the binder sequences can be further characterized for their binding specificity as well as the structure of the protein-peptide complex. Such specific binding affinity information is crucial for the establishment and improvement of computational prediction tools (as seen in Barlow et al. [2018], Kadukova et al. [2021], and Spiliotopoulos et al. [2016]) to enable effective evaluation of methodological parameters. Furthermore, computational approaches can complement or explain experimental findings by simulations of the dynamic behavior of the dArmRP-peptide complex. Additionally, the selection rounds during experimental screening can be sequenced with next-generation sequencing techniques. By that strategy, a gigantic amount of sequencing data is generated whose analysis can lead to an even more sophisticated design of focused libraries or selection methods.

Conclusions

Most affinity reagents for scientific research applications are still monoclonal antibodies derived from immunization, which either already exist and thus can be ordered from a supplier (catalog antibody), or they do not exist and have to be produced by immunization of animals (custom antibody). In fact, for most targets, epitopes and applications no suitable catalog antibody exists. And even if they exist, catalog antibodies frequently do not perform for reasons of cross-reactivity or low affinity, and the production of custom antibodies is costly in terms of time and money.

A major issue for common catalog or custom antibodies is that their genetic information is not available unless the antibody is sequenced in a labor-intensive step. However, applications with fusion proteins or the expression on cell or virus surfaces require the knowledge of the protein sequence to produce the binder recombinantly. Therefore, many catalog or custom antibodies are not suitable for such applications (Bradbury and Plückthun 2015). Additionally, common recombinant antibodies also have to be created anew for every new target sequence.

The collaborative PRE-ART project addresses these issues. A modular affinity reagent has been built based on the Armadillo repeat scaffold, where the modularity of the binder matches the target peptide architecture. Now, individual binding pockets are being designed to be specific

for individual amino acids on the target that can later be combined. Thus, with an existing set of binding pockets in place it will be possible to assemble an affinity reagent for a specific target sequence in a very short time. Apart from slight adaptations at the pocket interfaces no further experimental selections and computational optimizations will be necessary during the assembly of new sequence-specific binding proteins.

This fundamentally new concept allows one to bind linear target sequences in an unfolded state. Such stretches are often available at the termini of proteins or in linker regions, or they can be obtained by denaturation of the target protein as in SDS-PAGE or western blots. Unstructured targets of great interest are also the tails of receptors or regions of signal transduction molecules which are phosphorylated, or intrinsically disordered. Since unstructured regions are often post-translationally modified, these modular affinity reagents could be used to specifically target and investigate post-translational modifications. It would also be highly interesting to build pairs of binders for phosphorylated and unphosphorylated targets to visualize effects of candidate drugs on signaling pathways. Such an approach could accelerate mass spectrometry detection by orders of magnitude, circumvent labeling and thus permit to incorporate such a workflow into drug discovery. Because of the modular nature, “calibration” binders could be added that detect constant parts of the proteins in question, which would further add to the robustness of the concept.

Overall, the application of modular affinity reagents that can be assembled from predefined binding pockets has enormous potential for a wide range of applications. Because of the sequence-specific binding nature, these applications are completely out of reach of monoclonal antibodies or other conventional affinity reagent scaffolds.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work was supported by H2020-FETopen-RIA grant agreement 764434 (‘PRE-ART’)

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

- Alfarano, P., Varadamsetty, G., Ewald, C., Parmeggiani, F., Pellarin, R., Zerbe, O., Plückthun, A., and Caffisch, A. (2012). Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. *Protein Sci.* 21: 1298–1314.

- AlQuraishi, M. (2021). Machine learning in protein structure prediction. *Curr. Opin. Chem. Biol.* 65: 1–8.
- Ashraf, M., Frigotto, L., Smith, M.E., Patel, S., Hughes, M.D., Poole, A.J., Hebaishi, H.R.M., Ullman, C.G., and Hine, A.V. (2013). ProxiMAX randomization: a new technology for non-degenerate saturation mutagenesis of contiguous codons. *Biochem. Soc. Trans.* 41: 1189–1194.
- Baker, M. (2015). Blame it on the antibodies. *Nature* 521: 274–276.
- Barlow, K.A., Ó Conchúir, S., Thompson, S., Suresh, P., Lucas, J.E., Heinonen, M., and Kortemme, T. (2018). Flex ddG: Rosetta ensemble-based estimation of changes in protein-protein binding affinity upon mutation. *J. Phys. Chem. B* 122: 5389–5399.
- Beste, G., Schmidt, F.S., Stibora, T., and Skerra, A. (1999). Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc. Natl. Acad. Sci. U.S.A.* 96: 1898–1903.
- Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grütter, M.G., and Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* 22: 575–582.
- Bradbury, A. and Plückthun, A. (2015). Reproducibility: standardize antibodies used in research. *Nature* 518: 27–29.
- Capon, D.J., Chamow, S.M., Mordenti, J., Marsters, S.A., Gregory, T., Mitsuya, H., Byrn, R.A., Lucas, C., Wurm, F.M., Groopman, J.E., et al. (1989). Designing CD4 immunoadhesins for AIDS therapy. *Nature* 337: 525–531.
- Chen, T.S. and Keating, A.E. (2012). Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods. *Protein Sci.* 21: 949–963.
- Conroy, P.J., Law, R.H.P., Caradoc-Davies, T.T., and Whisstock, J.C. (2017). Antibodies: from novel repertoires to defining and refining the structure of biologically important targets. *Methods* 116: 12–22.
- Conti, E. and Kuriyan, J. (2000). Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin α . *Structure* 8: 329–338.
- Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α . *Cell* 94: 193–204.
- Desmet, J., De Maeyer, M., Hazes, B., and Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356: 539–542.
- Duivelshof, B.L., Murisier, A., Camperi, J., Fekete, S., Beck, A., Guillaume, D., and D’Atri, V. (2021). Therapeutic Fc-fusion proteins: current analytical strategies. *J. Separ. Sci.* 44: 35–62.
- Dyson, H.J. and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6: 197–208.
- Ernst, P., Zosel, F., Reichen, C., Nettels, D., Schuler, B., and Plückthun, A. (2020). Structure-guided design of a peptide lock for modular peptide binders. *ACS Chem. Biol.* 15: 457–468.
- Flanagan, M.L., Arias, R.S., Hu, P., Khawli, L.A., and Epstein, A.L. (2007). Soluble Fc fusion proteins for biomedical research. In: Albitar, M. (Ed.), *Monoclonal antibodies: methods and protocols*. Humana Press, Totowa, pp. 33–52.
- Forrer, P., Stumpp, M.T., Binz, H.K., and Plückthun, A. (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* 539: 2–6.
- Friedland, G.D., Linares, A.J., Smith, C.A., and Kortemme, T. (2008). A simple model of backbone flexibility improves modeling of side-chain conformational variability. *J. Mol. Biol.* 380: 757–774.
- Gainza, P., Nisonoff, H.M., and Donald, B.R. (2016). Algorithms for protein design. *Curr. Opin. Struct. Biol.* 39: 16–26.
- Gainza, P., Roberts, K.E., Georgiev, I., Lilien, R.H., Keedy, D.A., Chen, C.-Y., Reza, F., Anderson, A.C., Richardson, D.C., Richardson, J.S., et al. (2013). OSPREY: protein design with ensembles, flexibility, and provable algorithms. *Methods Enzymol.* 523: 87–107.
- Gao, W., Mahajan, S.P., Sulam, J., and Gray, J.J. (2020). Deep Learning in protein structural modeling and design. *Patterns* 1: 100142.
- Gebauer, M. and Skerra, A. (2020). Engineered protein scaffolds as next-generation therapeutics. *Annu. Rev. Pharmacol. Toxicol.* 60: 391–415.
- Gordon, D.B. and Mayo, S.L. (1999). Branch-and-terminate: a combinatorial optimization algorithm for protein design. *Structure* 7: 1089–1098.
- Graham, T.A., Weaver, C., Mao, F., Kimelman, D., and Xu, W. (2000). Crystal structure of a β -catenin/Tcf complex. *Cell* 103: 885–896.
- Groff, K., Allen, D., Casey, W., and Clippinger, A.J. (2020). Increasing the use of animal-free recombinant antibodies. *ALTEx*, <https://doi.org/10.14573/altex.2001071> (Epub ahead of print).
- Groff, K., Brown, J., and Clippinger, A.J. (2015). Modern affinity reagents: recombinant antibodies and aptamers. *Biotechnol. Adv.* 33: 1787–1798.
- Guntas, G., Purbeck, C., and Kuhlman, B. (2010). Engineering a protein-protein interface using a computationally designed library. *Proc. Natl. Acad. Sci. U.S.A.* 107: 19296–19301.
- Hansen, S., Tremmel, D., Madhurantakam, C., Reichen, C., Mittl, P.R.E., and Plückthun, A. (2016). Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity. *J. Am. Chem. Soc.* 138: 3526–3532.
- Hayes, R.J., Bentzien, J., Ary, M.L., Hwang, M.Y., Jacinto, J.M., Vielmetter, J., Kundu, A., and Dahiyat, B.I. (2002). Combining computational and experimental screening for rapid optimization of protein properties. *Proc. Natl. Acad. Sci. U.S.A.* 99: 15926–15931.
- Hughes, M.D., Nagel, D.A., Santos, A.F., Sutherland, A.J., and Hine, A.V. (2003). Removing the redundancy from randomised gene libraries. *J. Mol. Biol.* 331: 973–979.
- Jou, J.D., Jain, S., Georgiev, I.S., and Donald, B.R. (2016). BWM*: a novel, provable, ensemble-based dynamic programming algorithm for sparse approximations of computational protein design. *J. Comput. Biol.* 23: 413–424.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583–589.
- Kadukova, M., Machado, K.D.S., Chacón, P., and Grudinin, S. (2021). KORP-PL: a coarse-grained knowledge-based scoring function for protein-ligand interactions. *Bioinformatics* 37: 943–950.
- Koide, A., Bailey, C.W., Huang, X., and Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.* 284: 1141–1151.
- Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60: 84–90.
- Kuhlman, B. (2019). Designing protein structures and complexes with the molecular modeling program Rosetta. *J. Biol. Chem.* 294: 19436–19443.

- Kynast, J.P., Schwägerl, F., and Höcker, B. (2022). ATLIGATOR: editing protein interactions with an atlas-based approach. *bioRxiv*, <https://doi.org/10.1101/2022.01.19.476980>.
- Lazim, R., Suh, D., and Choi, S. (2020). Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems. *Int. J. Mol. Sci.* 21: 1–20.
- Leaver-Fay, A., Kuhlman, B., and Snoeyink, J. (2005). An adaptive dynamic programming algorithm for the side chain placement problem. *Proc. Pacific Symp. Biocomput.* 27: 16–27.
- Lechner, H., Ferruz, N., and Höcker, B. (2018). Strategies for designing non-natural enzymes and binders. *Curr. Opin. Chem. Biol.* 47: 67–76.
- Levin, A.M. and Weiss, G.A. (2006). Optimizing the affinity and specificity of proteins with molecular display. *Mol. Biosyst.* 2: 49–57.
- 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. *Briefings Bioinf.* 22: 1–14.
- Liu, N., Guo, Y., Ning, S., and Duan, M. (2020). Phosphorylation regulates the binding of intrinsically disordered proteins via a flexible conformation selection mechanism. *Commun. Chem.* 3: 1–9.
- Liu, Y. and Yu, J. (2016). Oriented immobilization of proteins on solid supports for use in biosensors and biochips: a review. *Microchim. Acta* 183: 1–19.
- Loshbaugh, A.L. and Kortemme, T. (2020). Comparison of Rosetta flexible-backbone computational protein design methods on binding interactions. *Proteins Struct. Funct. Bioinf.* 88: 206–226.
- Madhurantakam, C., Varadamsetty, G., Grütter, M.G., Plückthun, A., and Mittl, P.R.E. (2012). Structure-based optimization of designed Armadillo-repeat proteins. *Protein Sci.* 21: 1015–1028.
- Maguire, J.B., Haddox, H.K., Strickland, D., Halabiya, S.F., Coventry, B., Griffin, J.R., Pulavarti, S.V.S.R.K., Cummins, M., Thieker, D.F., Klavins, E., et al. (2021). Perturbing the energy landscape for improved packing during computational protein design. *Proteins Struct. Funct. Bioinf.* 89: 436–449.
- Morrison, S.L., Johnson, M.J., Herzenberg, L.A., and Oi, V.T. (1984). Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proc. Natl. Acad. Sci. U.S.A.* 81: 6851–6855.
- Murphy, G.S., Mills, J.L., Miley, M.J., Machius, M., Szyperski, T., and Kuhlman, B. (2012). Increasing sequence diversity with flexible backbone protein design: the complete redesign of a protein hydrophobic core. *Structure* 20: 1086–1096.
- Neuberger, M.S., Williams, G.T., and Fox, R.O. (1984). Recombinant antibodies possessing novel effector functions. *Nature* 312: 604–608.
- Ojewole, A.A., Jou, J.D., Fowler, V.G., and Donald, B.R. (2018). BBK* (Branch and Bound over K*): a provable and efficient ensemble-based protein design algorithm to optimize stability and binding affinity over large sequence spaces. *J. Comput. Biol.* 25: 726–739.
- Ollikainen, N., de Jong, R.M., and Kortemme, T. (2015). Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity. *PLoS Comput. Biol.* 11: 1–22.
- Parmeggiani, F., Pellarin, R., Larsen, A.P., Varadamsetty, G., Stumpp, M.T., Zerbe, O., Cafilisch, A., and Plückthun, A. (2008). Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.* 376: 1282–1304.
- Plückthun, A. (2015). Designed ankyrin repeat proteins (DARPin): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* 55: 489–511.
- Reichen, C., Hansen, S., Forzani, C., Honegger, A., Fleishman, S.J., Zhou, T., Parmeggiani, F., Ernst, P., Madhurantakam, C., Ewald, C., et al. (2016). Computationally designed armadillo repeat proteins for modular peptide recognition. *J. Mol. Biol.* 428: 4467–4489.
- Reichen, C., Hansen, S., and Plückthun, A. (2014). Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. *J. Struct. Biol.* 185: 147–162.
- Saunders, C.T. and Baker, D. (2005). Recapitulation of protein family divergence using flexible backbone protein design. *J. Mol. Biol.* 346: 631–644.
- Schilling, J., Jost, C., Mariuca Ilie, I., Schnabl, J., Buechi, O., Eapen, R.S., Truffer, R., Cafilisch, A., and Forrer, P. (2021). Thermostable designed ankyrin repeat proteins (DARPin) as building blocks for innovative drugs. *J. Biol. Chem.* 298: 1–12.
- Shih, H.H., Tu, C., Cao, W., Klein, A., Ramsey, R., Fennell, B.J., Lambert, M., Ní Shúilleabháin, D., Autin, B., Kouranova, E., et al. (2012). An ultra-specific avian antibody to phosphorylated tau protein reveals a unique mechanism for phosphoepitope recognition. *J. Biol. Chem.* 287: s44425–s44434.
- Simonson, T., Mignon, D., Druart, K., Michael, E., Opuu, V., Polydorides, S., Villa, F., Gaillard, T., Panel, N., and Archontis, G. (2020). Physics-based computational protein design: an update. *J. Phys. Chem.* 124: 10637–10648.
- Spiliotopoulos, D., Kastiris, P.L., Melquiond, A.S.J., Bonvin, A.M.J.J., Musco, G., Rocchia, W., and Spitaleri, A. (2016). dMM-PBSA: a new HADDOCK scoring function for protein-peptide docking. *Front. Mol. Biosci.* 3: 46.
- Treynor, T.P., Vizcarra, C.L., Nedelcu, D., and Mayo, S.L. (2007). Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc. Natl. Acad. Sci. U.S.A.* 104: 48–53.
- Voigt, C.A., Mayo, S.L., Arnold, F.H., and Wang, Z.G. (2001). Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. U.S.A.* 98: 3778–3783.
- Wang, J., Cao, H., Zhang, J.Z.H., and Qi, Y. (2018). Computational protein design with deep learning neural networks. *Sci. Rep.* 8: 1–9.
- Xu, C., Lu, P., Gamal El-Din, T.M., Pei, X.Y., Johnson, M.C., Uyeda, A., Bick, M.J., Xu, Q., Jiang, D., Bai, H., et al. (2020). Computational design of transmembrane pores. *Nature* 585: 129–134.
- Yin, S., Ding, F., and Dokholyan, N.V. (2007). Modeling backbone flexibility improves protein stability estimation. *Structure* 15: 1567–1576.