PNAS RESEARCH ARTICLE | APPLIED BIOLOGICAL SCIENCES

Modular binder technology by NGS-aided, high-resolution selection in yeast of designed armadillo modules

Yvonne Stark[a,1], Faye Menard[a,2], Jeliazko R. Jeliazkov[a,3], Patrick Ernst[a,4], Anupama Chembath[b], Mohammed Ashraf[b], Anna V. Hine[b], and Andreas Plückthun[a,5]

Edited by Andrei N. Lupas, Max-Planck-Institut fur Entwicklungsbiologie, Tuebingen, Germany; received November 3, 2023; accepted May 7, 2024 by Editorial Board Member William F. DeGrado

Establishing modular binders as diagnostic detection agents represents a cost- and time-efficient alternative to the commonly used binders that are generated one molecule at a time. In contrast to these conventional approaches, a modular binder can be designed in silico from individual modules to, in principle, recognize any desired linear epitope without going through a selection and hit-validation process, given a set of preexisting, amino acid–specific modules. Designed armadillo repeat proteins (dArmRP) have been developed as modular binder scaffolds, and we report here the generation of highly specific dArmRP modules by yeast surface display selection, performed on a rationally designed dArmRP library. A selection strategy was developed to distinguish the binding difference resulting from a single amino acid mutation in the target peptide. Our reverse-competitor strategy introduced here employs the designated target as a competitor to increase the sensitivity when separating specific from cross-reactive binders that show similar affinities for the target peptide. With this switch in selection focus from affinity to specificity, we found that the enrichment during this specificity sort is indicative of the desired phenotype, regardless of the binder abundance. Hence, deep sequencing of the selection pools allows retrieval of phenotypic hits with only 0.1% abundance in the selectivity sort pool from the next-generation sequencing data alone. In a proof-of-principle study, a binder was created by replacing all corresponding wild-type modules with a newly selected module, yielding a binder with very high affinity for the designated target that has been successfully validated as a detection agent in western blot analysis.

designed armadillo repeat proteins | yeast surface display | modular binder technology | specificity selection | next-generation sequencing (NGS)

Significance

The aim of the modular binder technology is to provide on-demand binders for user-specified epitopes. This could significantly enhance the scientific output in biology and medicine-related fields. The development of the herein presented selection strategies used to generate specific modules represents a milestone to advance this technology. At the same time, it demonstrates that the resolution limit of selections in combination with FACS can be pushed to distinguish individual amino acid mutations in an epitope, which in itself represents a technological achievement. The proof-of-principle study of the binder assembled from a newly selected module validates the concept of the modular binders and demonstrates its efficacy in an application example.

Specific binding proteins are a central pillar of biological and biomedical research due to their application as detection agents in a wide variety of different assays, ranging from western blots, ELISA, and immunohistochemistry assays to affinity chromatography and many more (1). Historically, antibodies are the most commonly used agents, and the majority of them outside of therapy are still created by immunization. Polyclonal antibodies suffer from batch-to-batch variability, but even monoclonal antibodies show cross-reactivities, clones can be lost, and they are usually not molecularly defined by the sequence (2). Besides recombinant antibodies that are selected from libraries, over the last decade, many specially engineered scaffolds have been brought forward that by design do not contain disulfides or glycosylation sites and can thus be expressed in *Escherichia coli* inexpensively and are by definition sequence-defined and monoclonal, thereby greatly enhancing binder production and quality (3, 4). However, even with these new scaffolds, binders have to be selected de novo individually for each new target, significantly limiting the throughput and speed at which those binders can be generated for new targets. To overcome this remaining issue, we propose the use of modular binders.

Modular binders are built from custom modules that recognize subsections of an epitope and are assembled to generate a binder specific for the full epitope as shown in Fig. 1. Given a preexisting set of modules, binders can be rapidly assembled according to the desired target sequence. The concept of modular binding is especially well suited to generate binders against linear epitopes since the prediction of secondary or tertiary structure elements at an atomic level is still difficult and the diversity of structured epitopes renders a modular approach impractical. Nevertheless, in many of the desired applications, the target proteins will by default be detected in a denatured state (e.g., western blots, proteomics) and, considering that also correctly folded proteins often contain unstructured parts in loops or termini (5), the range of possible applications for linear peptide binders is very broad. It was also found over the last decade that up to 15 to 40% of all cellular interactions are actually protein–peptide interactions (6), being involved in essential tasks

Author affiliations: [a]Department of Biochemistry, University of Zürich, Zürich CH-8057, Switzerland; and [b]College of Health and Life Sciences, School of Biosciences, Aston University, Aston Triangle, Birmingham B4 7ET, United Kingdom

[1]Present address: Mabylon, Schlieren 8952, Switzerland.

[2]Present address: Neurimmune, Schlieren 8952, Switzerland.

[3]Present address: Protein Design and Informatics, Data and Predictive Science, Glaxo Smith Kline R&D, Collegeville, PA 19426.

[4]Present address: University of Zurich, Dean's office of the Medical Faculty.

[5]To whom correspondence may be addressed. Email: plueckthun@bioc.uzh.ch.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2318198121/-/DCSupplemental.

Published June 25, 2024.

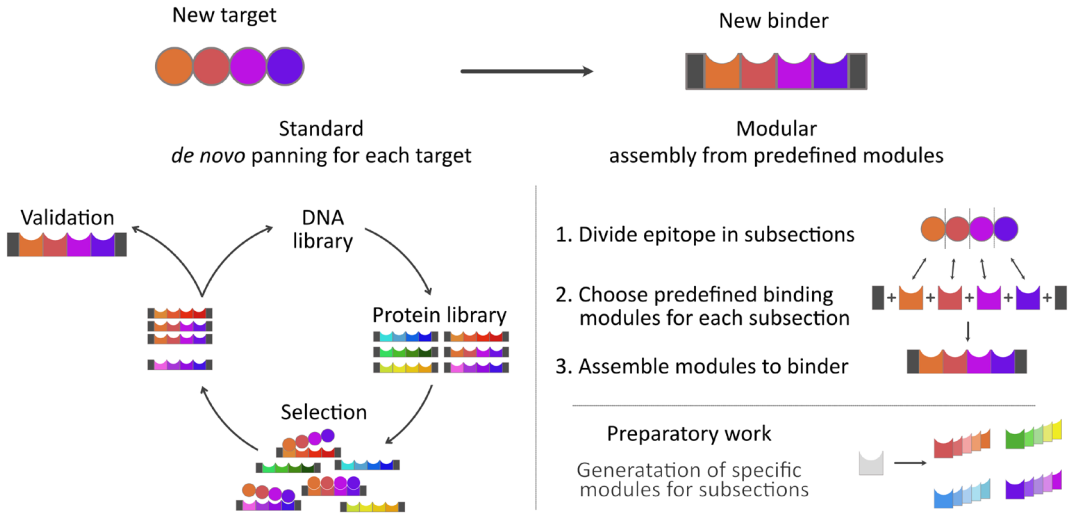

Fig. 1. Generation of modular binders. In contrast to the standard approaches (*Left*) which require a de novo selection for each new target, the modular binder technology (*Right*) is based on a preexisting set of compatible modules with individual specificities such that a binder can be assembled to match the desired target sequence. In principle, a binder can be generated for any linear target sequence or length, after the modules have been obtained. The generation of the individual, specific modules is achieved in a preparatory stage and is required only once since they all have the attribute for generic binder assembly (while being specific for a particular interaction with the target).

such as signaling, regulation of cell adhesion, or protein trafficking (7, 8). For simplicity, we will refer to linear epitopes as peptides, even if they are part of a folded protein. It should be noted that this is a very different problem from the generation of binders against tags forming poly-proline helices, which has recently been reported in elegant design studies (9) since the current approach allows binding of any unfolded sequence.

As discussed in ref. 10, repeat proteins represent an attractive starting point for the design of modular peptide binders. To initiate this work, we derived a designed ArmRP (dArmRP) (10) by consensus design of natural ArmRPs (11), mostly influenced by importin-α, which naturally targets highly positive charged peptides. As a result, the initial consensus dArmRPs binds $(KR)_n$ peptides.

The binding interface of designed armadillo repeat proteins (dArmRPs) is formed by the juxtaposition of its repetitive subunits (modules), naturally resulting in a modular interaction with its extended peptide ligand. Iterative cycles of engineering and design resulted in generic internal modules retaining all the peptide-binding features while being "stackable" to form binders of varying lengths (12, 13). Furthermore, the consensus dArmRP scaffold has been shown to bind its designated $(KR)_n$ target peptide in a highly modular fashion (14). Moreover, the overall affinity can be adjusted over many orders of magnitude by altering the length of the protein and the peptide (14).

Our previous work has focused on improving the scaffold over several generations, and we demonstrated that individual binding sites can be exchanged by grafting naturally occurring binding sites (15). Here, we describe the generation of dArmRPs with novel recognition sequences. We report the development of a powerful selection technology that favors selection for specificity, and not just affinity, which can be adjusted independently, and the characterization of binders, including a high-resolution crystal structure. Furthermore, we report the analysis of the selection progress by next-generation sequencing that helps to rationalize the success of different selection strategies.

Results

Yeast Surface Display. Yeast surface display (YSD) was chosen as a selection technology to generate specific modules for different amino acids, based on the dArmRP consensus scaffold. To ensure the correct binding register of the repetitive peptide target, a hydrophobic interaction was grafted onto the consensus scaffold to specifically recognize a Leu-Ser-Phe (LSF) motif in the target peptide, thereby anchoring it in the desired position (14, 15) as shown in Fig. 2.

A structure-based, rationally designed library with seven randomized positions was designed and integrated within a dArmRP with six internal repeats as shown in Fig. 2. The N-terminal part of the dArmRP was adapted to accommodate the LSF lock motif (15) to ensure the correct binding register of the target peptide. The peptides used for the YSD selection contain the variable target amino acid X positioned to engage with the randomized dArmRP module. A schematic drawing of the selection setup is shown in Fig. 2. The residues spatially flanking the target amino acid X in the peptide were changed to Ala to minimize selection bias toward these flanking sites. The introduction of Ala additionally decreased the general affinity of the target peptide from the picomolar to the low nanomolar range, which had been found previously to be ideal to distinguish the binding energy contributed by individual target amino acid side chains. Nonetheless, for correct target binding, the complete peptide must be engaged by the dArmRP. Hence, during selection, the additional energetic gain resulting from favorable interactions with the target amino acid X has to be distinguished from the basal peptide binding energy.

Selection with Single Side-Chain Resolution. The main challenge of the dArmRP module selection is the need to select modules that have a high selectivity for one specific amino acid side chain in a peptide, but not necessarily the highest affinity. To this end, differently labeled off-target peptides were used as competitors during the selection process. Such off-target peptides are identical except for variable position X where they contain any residue other than the target amino acid. Since the target and off-target peptides differ in only this one amino acid, a large part of the respective affinities results from the constant part, which is identical in both the target and the off-target peptides. We will refer to this affinity, attributed to the constant part of the peptide, as the basal (peptide) affinity. Hence, the selection process requires the discrimination of the affinity given by the sum of the basal peptide affinity and the beneficial interactions of the target amino acid side chain, versus the basal peptide affinity

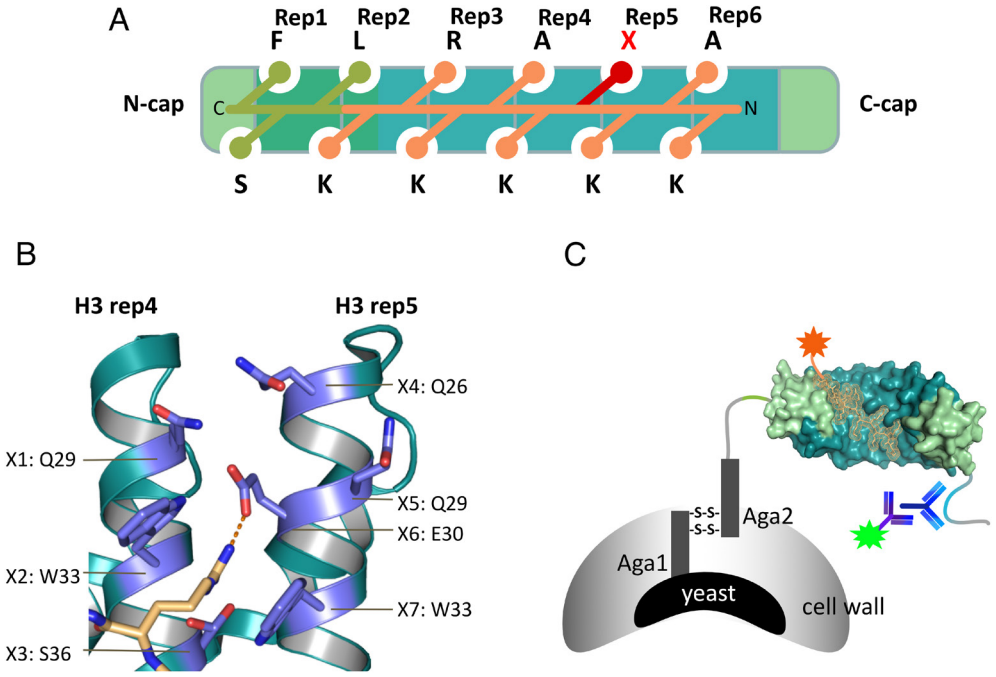

Fig. 2. dArmRP library and selection setup. (*A*) Schematic drawing of a dArmRP as used for YSD. The N-cap and first two repeats harbor the lock mutation, recognizing the peptide's LSF motif at the C-terminal end, shown in olive. Repeats 4 and 5 contain the randomized library residues engaging target amino acid X shown in red. Ala residues are introduced to spatially flank the amino acid X, preventing a potential selection bias and reducing the basal peptide affinity to a range suitable to discriminate the impact of binding the target amino acid X. (*B*) Structure of a consensus dArmRP with the seven library positions highlighted in blue as used in the selection adapted from PDB 5MFC. In orange is depicted the Arg side chain engaging the pocket. (*C*) YSD setup with the dArmRP linked to the cell wall. A C-terminal myc tag is detected by standard antibodies as expression control (shown with a green label) in parallel to the target binding (shown as a peptide with an orange label).

combined with the effect of off-target side chains. We therefore require single side-chain resolution.

The selection complexity is further exacerbated by the relatively flat topography of the dArmRP module binding surface. Large amino acids cannot be excluded from binding by generating a small cavity since they can potentially avoid repulsions or clashes by adapting a different rotamer conformation. Hence, the depletion of cross-reactive binders, which show a high affinity for the target, but also for other amino acids in this position, is the most critical step for a successful selection.

In order to achieve this depletion of cross-reactive binders, an excess of competitor peptides is typically added during the sorting experiments, decreasing the signal intensity in the target signal channel for those yeast cells that simultaneously bind to target and competitor and hence display a cross-reactive dArmRP. In theory, variants that then still maintain a high target signal should be specific for the target and non-cross-reactive for all competitors used during this selection. This approach has been successfully used as, e.g., described in ref. 16. However, for us, this assumption turned out to be oversimplified when distinguishing very similar epitopes, as the discrimination of high-affinity but cross-reactive binders from high-affinity, specific binders has proven to be very challenging. We attribute this to the fact that the highest affinity binders are not necessarily the most specific ones, which renders it very difficult to set the right target-to-competitor ratio: Too high a competitor concentration will eventually also affect target-specific binders (due to the same basal affinity of the constant parts of the target peptide), whereas too low a concentration will not lead to a notable shift in target signal intensity for cross-reactive binders. Also, the definition of a too-high or too-low competitor concentration is highly dependent on the precise affinities for the target and off-target peptides and obviously differs for the individual binders contained in the selection pools. Using an excess of competitor peptides during the selection is therefore always a compromise between keeping lower affinity but higher specificity binders versus depleting high-affinity but cross-reactive variants.

The above-described "standard" competition strategy was successfully applied to deplete binders with very low general affinity or those with high specificity for off-targets, but high amounts of cross-reactive binders were still retained. Thus, this standard competitor selection proved to be very useful, but only to enrich a pool of binders all with a high affinity for the designated target but with variable specificity. We therefore developed a two-step competition strategy: first, to remove those binders with higher affinity to off-targets than for the desired one using standard competition (as described above), and second, an additional competition step to remove those binders with high affinity to the desired target but still significant cross-reactivity to some other target(s) (as described in the next sections).

To address the issue of specificity, we therefore developed an additional competition setup which weighs the specificity of a binder above its affinity. This strategy was termed "reverse competitor" selection because the target peptide is employed as a competitor to decrease the binding signal of the desired clones. It follows the rationale that, compared to cross-reactive binders (with higher affinity to off-targets than to the desired target), specific binders (with higher affinity to the cognate target, but nonzero cross-reactivity) are more susceptible to competition when using their designated target as a competitor.

Schematic drawings of a standard and reverse competition selection setup and a comparison between an in-parallel performed standard and reverse competition selection are shown in Fig. 3. For the reverse competition, a mix of target peptide and off-target peptides are labeled with the same dye [e.g., mNeonGreen (17)] and are incubated with an excess of only the target peptide as a

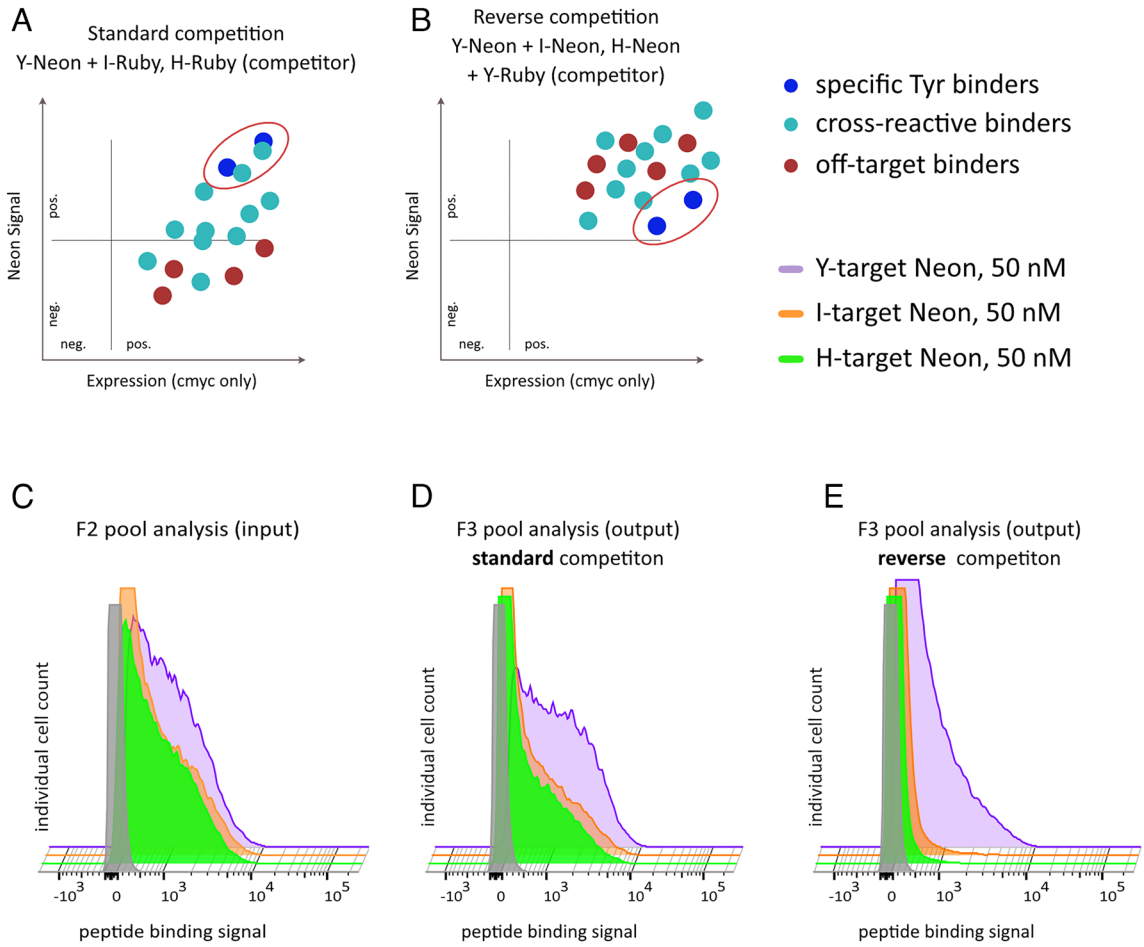

Fig. 3. Competition strategies. (*A*) Schematic drawing of the flow cytometric dot plot for the respective library members in the standard competition approach. Here, selection of a Tyr binder is depicted. Off-target binders (i.e., mostly binding to targets other than Tyr) and cross-reactive binders with a strong preference for the used competitors (binding Tyr, but also H- and I- Ruby in this example) are efficiently depleted. The sorting gate is shown as a red oval. (*B*) Schematic drawing of the flow cytometric dot plot for the respective library members with the reverse competition strategy. The whole population is shifted into the double-positive quadrant due to the incubation with Y-, I,- and H-Neon. Only the specific binders experience a downshift for the Neon signal (*y* axis) upon addition of the Y-Ruby competitor. The sorting gate is shown as a red circle. (*C*) Actual pool analysis after the second selection round (F2), carried out as a competitor sort, to enrich high-affinity binders for the target. This pool was the input for the third selection round (F3), for which the competitor and reverse competitor sort were performed in parallel. (*D*) Pool analysis after the third sort round, performed as a competitor sort. An enrichment for the designated targets is observed compared to the input pool (F2). However, the specificity did not markedly improve since the noncognate I and H- targets are still binding to a similar extent as seen in the F2 pool. (*E*) Pool analysis after the third sort round, performed as reverse-competitor sort. Compared to the input pool, a clear specificity improvement can be observed since only the designated target shows a significant binding signal.

competitor with a different label [e.g., labeled with mRuby3 (17)]. As the competitor peptide labeled with mRuby3 target is identical to the target peptide labeled with mNeonGreen (except for the dye), it has the same affinity to the presented binders and will participate in the same equilibrium. Due to the excess mRuby3 peptide, a significant fraction of the bound target peptide will be labeled with mRuby3. Hence, the competition effect (fewer bound mNeonGreen-peptides, which results in a lower mNeonGreen signal) is most prominent for target-specific binders because these binders more exclusively interact with the target peptide alone such that a significant part of the bound peptides carries an mRuby3 label.

The cross-reactive binders in contrast engage the mNeonGreen target and off-targets peptides, and therefore, the fraction of bound mNeonGreen peptide that can be effectively competed with target-competitor is lower. This effect is even potentiated for binders with slightly higher affinities for off-targets compared to the designated target since the competition effect is obviously weaker when competing a higher-affinity (off-target) peptide with a lower-affinity target peptide.

In other words. we exploit the fact that the truly specific binder can only bind the cognate peptide out of the mixed "green" peptides, while less specific binders will bind both the green cognate peptide and the green off-target peptides. The excess of "red" cognate peptide therefore significantly displaces the green cognate peptide from the specific binder, but much less so from the less specific binders.

Hence, in a reverse competitor setup, specific binders are singled out because they preferentially engage the designated target (which is used in the function of a competitor) (Fig. 3). In contrast, in a normal competition setup, one of the criteria for a variant to be selected is that it should not bind to the employed competitors (which correspond to the off-targets in a normal competition setup). However, this does not automatically mean that they are specific for the designated target. Additionally, by using the target peptide in the function of a competitor, the selection stringency increases with lower competitor concentration since the competitor is the target itself. This significantly decreases the risk of false positives due to a high competitor concentration. As for any competition experiment, the degree of competition depends on the binder's affinities for target and off-targets, respectively, and the relative target/competitor concentrations.

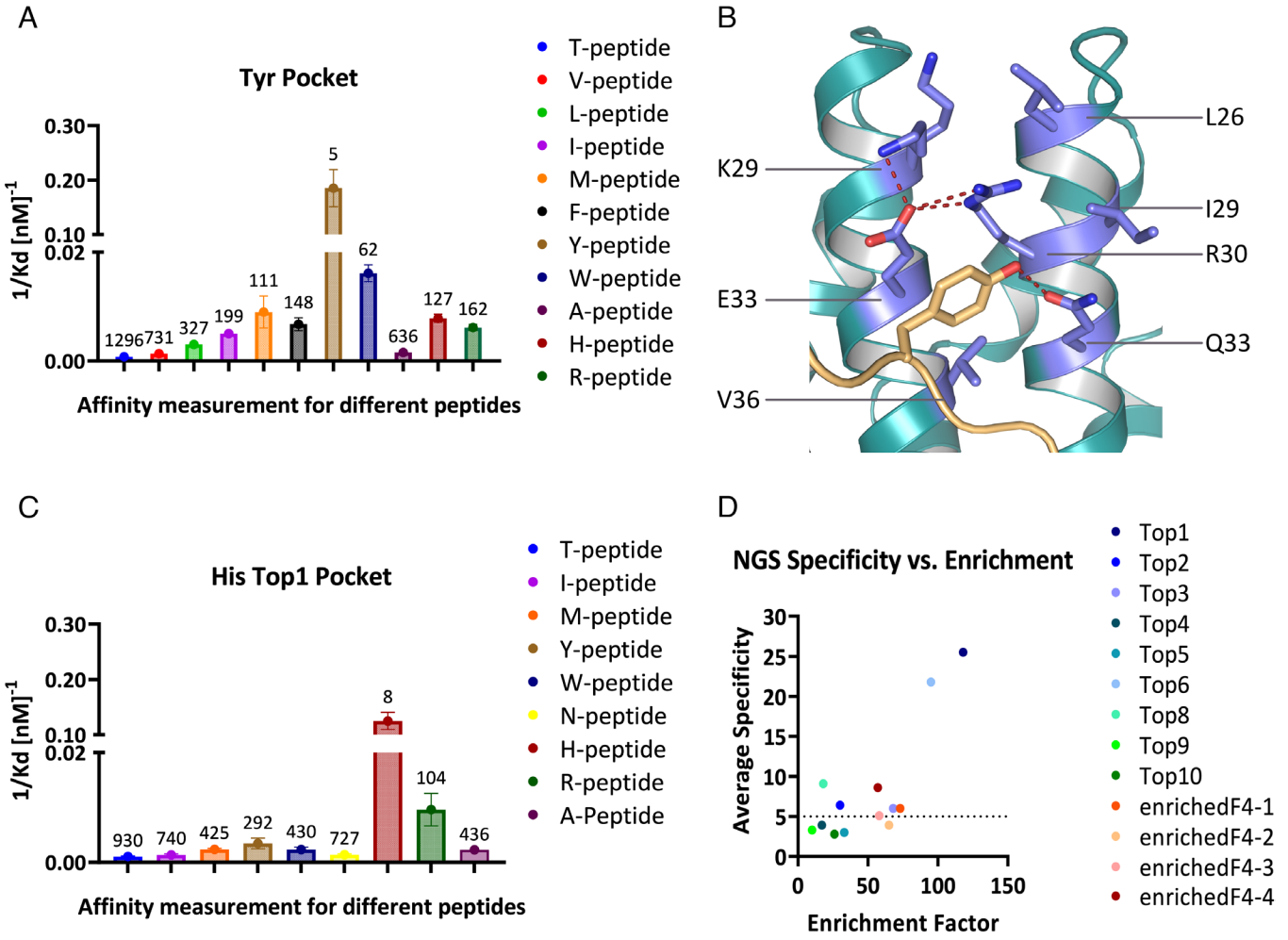

Fig. 4. Characterization of hits. (*A*) K_d data measured by fluorescence anisotropy for the selected Tyr pocket with peptides of the sequence KAKXKAKRKLSF, differing in one amino acid X. The one-letter code abbreviation of the legend on the *x* axis represents amino acid X. The *y* axis is $1/K_d$, while the numbers above the bar represent the measured K_d in nM. Showing the affinities as $1/K_d$ allows to depict higher affinities as higher bars. (*B*) 1.4 Å resolution structure of the selected Tyr pocket interacting with its target amino acid (PDB ID: 8QZN). (*C*) K_d data measured by fluorescence anisotropy for the selected Top1 His pocket, shown as a bar graph, the figure is equivalent to (*A*). (*D*) Representation of the NGS binder specificity as a function of their enrichment in the last selection round. The specificity factor is defined as the affinity to the designated target divided by the affinity to the next best noncognate amino acid. The average specificity shown here is the mean of the specificity factor for the two best off-targets. The enrichment factor is the ratio of the abundance fraction of a binder in the F4 pool, divided by its abundance fraction in pool F3. Binders *enrichedF4-1* to *enrichedF4-4* were selected from the top200 most abundant binders in the F4 pool, solely based on their enrichment factor. They would not have been analyzed based on their abundance, as it was not high. However, their specificities are at least as good as those from the 10 most abundant binders (Top1 to Top10). This finding demonstrates the correlation between the enrichment factor and the binder specificity. Error bars show SD of duplicate measurements.

The comparison of a standard and reverse competition sort (Fig. 3) clearly demonstrates the ability of the reverse competition setup to deplete cross-reactive variants. As will be discussed below, the deep sequencing of selection pools at different stages of selection allowed us to visualize the effect of the reverse competition strategy, again indicating its efficacy to separate cross-reactive from specific binders.

In our understanding, the success of the reverse competitor over the standard competitor selection depends on the constitution of the epitopes that have to be distinguished and the conformational space the binder can sample to engage the target epitope. We expect that the reverse competitor strategy outperforms a standard competition approach whenever the discrimination of highly similar epitopes is required and/or if the conformational space that the binder can sample is limited. As an additional advantage, the reverse competition setup does not require the designated target to be labeled as long as the off-targets can be detected (although it provides an additional layer of control and is therefore recommended if possible). This feature renders the reverse competitor strategy amenable for targets that are difficult to produce.

Selection of Tyr and His Pockets. Based on the described strategies, selections against His and Tyr were performed as the amino acid X, in the context of an otherwise identical peptide. To generate this "single-pocket library," all amino acids were exchanged compared to the original Arg pocket (Fig. 4), thereby changing the specificity of one single binding pocket within the scaffold. Hits were identified from the last selection pool by a single-clone screen of yeast cells using flow cytometry. The sequence-verified hits were then expressed in *E. coli* and characterized by measuring the K_d for different peptides by fluorescence anisotropy to determine the affinity and specificity of the selected modules. The selectivity and affinity trends observed in the yeast single-clone assay were confirmed. The obtained affinities for the best Tyr and His module are shown in Fig. 4 and their sequences in *SI Appendix*, Table S1.

With a K_d of 5 nM in the context of the chosen peptide, the pocket selected for Tyr shows a high affinity for its target and at the same

time discriminates against the next best target (Trp) by a factor of more than 10. All other amino acids are lower in affinity by at least a factor 20, demonstrating that a complete change in amino acid specificity (here from Arg to Tyr) can be achieved. A high-resolution crystal structure (Fig. 4) was obtained, elucidating the interaction of the Tyr side chain and the selected Tyr pocket. The pocket is formed by residues from helices of two subsequent structural repeats as shown in the crystal structure, and the residues are numbered by their respective internal residue numbers. An H-bond network between the selected residues K29, E33, and R30 orients the E33 in an ideal conformation to form a C-H–π stacking interaction with its Cβ hydrogen to the aromatic moiety of the Tyr side chain. Additionally, the Tyr hydroxyl head group of the incoming Tyr side chain is engaged in an H-bond to residue Q33. Upon grafting a Trp as the target amino acid into the peptide structure (Pymol), a clash with R30 is observed, whereas the spatially fitting Phe lacks the hydroxyl group to undergo the H-bond with Q33. Hence, the observed binding mode supports the experimental K_d data.

The characterization data for the top1 His pocket [named after the next-generation sequencing (NGS) ranking discussed below] reveal a similar performance in terms of specificity and affinity as the Tyr pocket, as shown in Fig. 4. The Top1 His pocket was measured to have an affinity of 8 nM for its designated target in this peptide context, discriminating against Arg as the next best binding amino acid by a factor of 13 and against all other measured amino acids by more than a factor of 30. Hence, also the Top1 His pocket fulfills the criteria with regard to specificity and selectivity required for a modular binder assembly approach and validates the above-described selection strategy.

Next-Generation Sequencing. The selection rounds for the His modules were analyzed by Illumina MiSeq NGS. The seven library positions were extracted from the processed data for each single round. A comparison of the amino acid distribution over different selection rounds, as well as the visualization of the selection progress by projecting the sequence differences as distances using UMAP (18), impressively visualizes the effect of the different selection strategies (Fig. 5). The data confirm the desired distribution of amino acids for the randomized library positions in the native pool, which is very close to the designed distribution. The first round (FACS1) was sorted with low stringency, aiming mainly to separate binding vs. nonbinding variants and to decrease the library size. Hence, the FACS1 pool still maintains a high diversity. Enrichment of individual amino acids in different library positions is observed after FACS2 and remains similar for FACS3, both of which employ a standard competitor selection strategy to deplete binders with predominant off-target specificity and enrich binders with high affinity for the target. The most significant change in amino acid distribution is observed for the reverse competition sort in FACS4, which aims to separate the high-affinity but cross-reactive binders from high-affinity specific binders, as described above.

Interestingly, during that final step, in some positions, the amino acids that were enriched in FACS2 and FACS3 are significantly reduced again in FACS 4 (e.g., position X4, Fig. 5). The shift in sequence space upon reverse competitor selection is also clearly observed in the UMAP plots. The sequence space explored by the FACS4 pool comprises only a small portion of that of the FACS3 pool, underlining the importance of this selection strategy to retrieve modules with the desired selectivity.

The 10 most frequent clones as identified by NGS were additionally characterized as shown in Table 1. All of them show high affinity for the His target but differ in their specificity. Comparing the relative abundance of the top10 binders in pools FACS3 and FACS4, it can be observed that overproportionally enriched binders in the step from FACS3 to FACS4 tend to have better specificity. Hence, the selectivity of individual binders can be estimated based on their enrichment factor that is calculated as a ratio of the respective sequence abundance fraction in FACS3 and FACS4. The correlation of this enrichment with measured selectivity mirrors the success of the reverse competition. It can also be seen that the selection for affinity (FACS3, standard competition), in comparison, has not enriched clones that are truly specific, while the direct selection for selectivity (FACS4, reverse competition) did achieve the intended purpose.

We next wanted to test the hypothesis that a notable enrichment from FACS3 to FACS4 would conversely help to identify highly selective clones. Therefore, binders with a high enrichment factor were picked out of the top200 NGS hits and characterized for their selectivity. Three out of four tested binders indeed show a specificity factor average above 5, which makes them more specific than 50% of the top 10 binders (Fig. 4). Hence, the enrichment factor allows one to harvest highly sequence-diverse binders from the NGS data alone, which would be very unlikely to have been identified otherwise.

Binder Generation Based on Assembly of Selected Modules. Based on the selected Tyr pocket, a binder with two adjacent Tyr pockets was designed. The assembly of two adjacent pockets presents some challenges because the central H3 helix is contributing to both pockets (Fig. 6). Screening was performed investigating four variants that are obtained from rationally combining the selected amino acids in the critical positions as shown in Fig. 6, as either of the shared positions could prefer the identity of the residue from the "left" or the "right" pocket. Retaining a Glu in the shared library position X7 proved to be important to obtain the best affinity toward the designated target.

The most promising combination for the double Tyr motif was then used to create a binder for $(KY)_4$ and $(KY)_6$, respectively, by propagating these double pockets. The binders and their corresponding targets were expressed in *E. coli* and characterized by FA measurements. The TyrM4 and TyrM6 dArmRPs show a very high affinity to their respective targets. For comparison, the consensus M4 dArmRP binds with 250 nM to its cognate $(KR)_4$ target peptide (14), whereas the TyrM4 binds its cognate target $(KY)_4$ with 3 nM (cf. Fig. 7). The TyrM6 affinity to its cognate target was estimated to be below picomolar, as the K_d could not be exactly determined, because in a FRET assay to measure the off-rate (15), no disassociation of the $(KY)_6$ target was detected over the maximal measuring time of 16 h.

However, assuming that a 10% dissociation of the complex over the 16 h measuring time would have been observed, the dissociation rate can be estimated to be smaller than 1.7×10^{-6} s^{-1}. Together with the association rate that was determined to be 1.3×10^{7} M^{-1} s^{-1} the equilibrium dissociation constant for the TyrM6 and its cognate target $(KY)_6$ can be determined to be below 1.5×10^{-13} M.

The impact of individual amino acids of the $(KY)_n$ peptide to the binding affinity was also investigated. Similar to the original $(KR)_n$ binder (14), a very regular contribution to the free energy of binding was observed for the $(KY)_n$ binder TyrM4, depending on the nature of the respective amino acid, but regardless of its position in the target peptide. As shown in Fig. 7, the substitution of a Tyr by an Ala reduces the affinity by a factor of 10, whereas the exchange from a Lys to an Ala results in a 2.5 times diminished affinity. This validates the modularity of the TyroMX binders and the success of introducing the new Tyr pockets.

The TyrMX binders thus serve as proof of principle that binders can be assembled based on modules selected by YSD. The characteristics of a modular binder are maintained, for instance, the

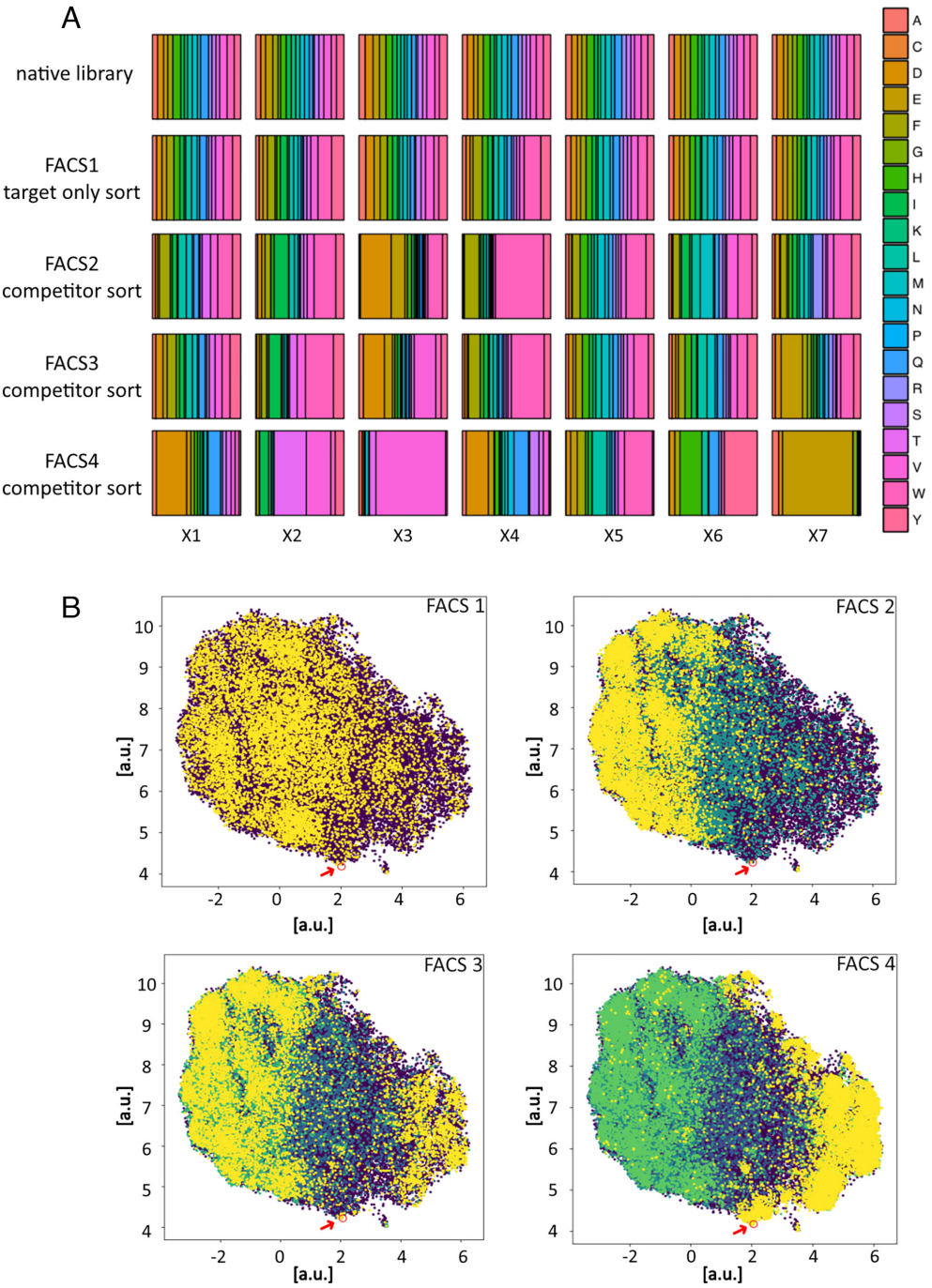

Fig. 5. NGS analysis of the His selection. (*A*) Amino acid distribution in the seven randomized library positions over the course of the selection. (*B*) UMAP representation of the library sequence space (seven randomized positions only) over the course of the selection. The sequences for the stated selection pool are always shown in yellow on top of the previous pools. In the FACS1 scheme the native pool is shown in purple and the FACS1 pool in yellow. In the FACS2 scheme the native pool is shown in purple, the FACS1 pool in cyan and the FASC2 pool in yellow. In the FACS3 scheme the native pool is shown in purple, the FACS1 pool in cyan, the FASC2 pool in green and the FACS3 pool in yellow. In the FACS4 scheme the native pool is shown in purple, the FACS1 pool in cyan, the FASC2 pool in green, the FACS3 pool in light green and the FACS4 pool in yellow. The red circle (arrow) denotes the top1 binder sequence.

ability to tune the affinity by adjusting the length of the binder or the regular binding energy contribution of individual amino acids in the target peptide, as shown in Fig. 7.

A rather obvious application for modular peptide binders is their use as primary detection agents in assays such as western blots (WB). To investigate the performance of the dArmRP as staining reagents, an HA-tagged TyrM6 was used to detect blotted $(KY)_6$-GFP in a WB assay. The TyrM6 itself was stained via a standard anti-HA antibody sandwich. As a control, the GFP was also detected via an anti-GFP antibody sandwich. As shown in Fig. 7, the dArmRP specifically detects the expressed $(KY)_6$-GFP in the *E. coli* cell lysate without background. This result clearly demonstrates the potential for the use of custom-made dArmRPs as sequence-specific protein detection agents.

Discussion

The selection of amino acid–specific dArmRP modules has proven to be a challenging quest since the discrimination of a single side chain is required. Nonetheless, a series of interwoven technology

Table 1. K_d measurements of NGS hits

	K_d (nM)			Specificity factor			NGS	
Variant	H-peptide	R-peptide	Y-peptide	R/H	Y/H	Average	Enrichment factor	Rank FACS4
top1	7.8 ± 1.3	104.0 ± 22	292.0 ± 76	13.4	37.6	25.5	118	1
top2	14.9 ± 1.9	109.7 ± 47	81.7 ± 18	7.4	5.5	6.4	30	2
top3	11.6 ± 3.4	72.7 ± 23	65.0 ± 20	6.3	5.6	6.0	68	3
top4	22.0 ± 3.2	137.3 ± 43	33.3 ± 9	6.2	1.5	3.9	17	4
top5	12.6 ± 3.3	32.2 ± 7	44.3 ± 10	2.6	3.5	3.0	33	5
top6	10.8 ± 3.4	176.0 ± 29	296.3 ± 86	16.2	27.4	21.8	95	6
top8	10.0 ± 2.6	31.0 ±10	150.7 ± 26	3.1	15.1	9.1	18	8
top9	16.5 ± 5.3	61.7 ± 19	47.3 ± 14	3.7	2.9	3.3	10	9
top10	6.7 ± 2.6	21.3 ± 4	16.3 ± 4	3.2	2.4	2.8	26	10
enrichedF4-1	4.0 ± 1.3	39.0 ± 6	9.0 ± 2	9.8	2.3	6.0	73	196
enrichedF4-2	16.0 ± 3.7	49.0 ± 9	76.0 ± 9	3.1	4.8	3.9	65	170
enrichedF4-3	10.0 ± 2.1	47.0 ± 11	55.0 ± 11	4.7	5.5	5.1	58	99
enrichedF4-4	3.5 ± 1.7	43.0 ± 13	17.0 ± 3	12.3	4.9	8.6	57	92

K_d data ± SD of duplicate measurements of the ten most abundant hits (*top1* to *top10*) of the His selection present in the FACS4 pool and the variants selected based on their enrichment factor (*enrichedF4-1* to *enrichedF4-4*). The K_d was determined by a fluorescence anisotropy assay. For the calculation of the specificity factor, the K_d of the off-target R- and Y-peptides is divided by the K_d of the designated H-peptide target. The enrichment factor is the ratio of the abundance fraction of the respective clone in FACS3 and FACS4 (numbers not shown). The rank in FACS4 is based on the absolute abundance ranking of the variants in FACS4.

and protein engineering steps have allowed us now to reach this goal. After recognizing that the use of multivalent peptide constructs tends to obliterate the binding difference by the variable target amino acid X, a selection strategy based on strictly monovalent targets was developed to achieve the specificity required for the individual modules. Using this approach, the key component for success was the development of a powerful counterselection strategy. The classical competition setup was found to be

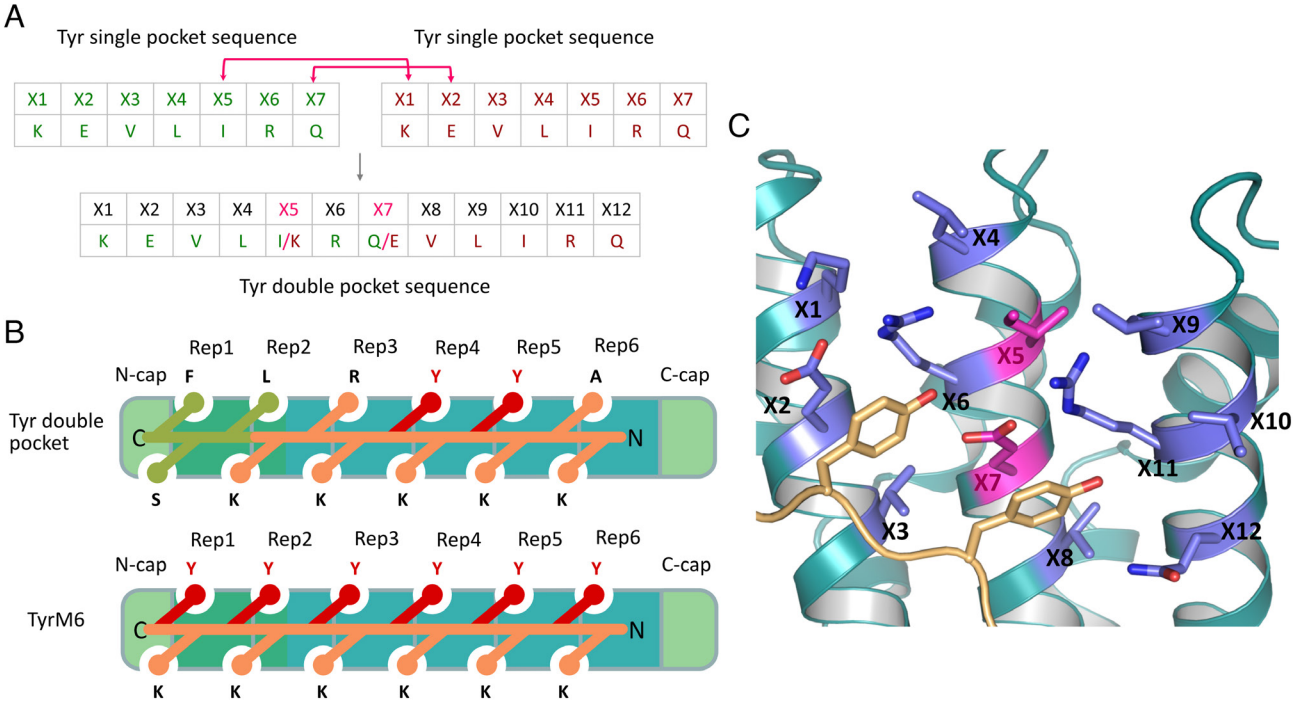

Fig. 6. Generation of double pockets. (*A*) Schematic overview for combining two single pockets into a double-pocket sequence using the Tyr single pocket as an example. Only the variable residues are shown; X1 to X7 represent the library residues from the single-pocket selection. Due to the same basic structure of each repeat, the residue X5 of the first single-pocket sequence is the same as residue X1 of the adjacent pocket, likewise X7 of the first pocket and X2 of the adjacent pocket as indicated with the pink arrows. Therefore, a double-pocket graft involves only twelve point mutations, but in positions X5 and X7, two different amino acids have to be considered. (*B*) *Upper*: Schematic drawing of a dArmRP with a grafted double Tyr pocket. The N-cap and the first two repeats still harbor the mutations required to bind the "LSF"-lock motif on the C-terminal end of the peptide (olive). The mutations introduced by the double-pocket graft are located in repeats 3, 4, and 5 indicated by the dark blue color. Repeat 4 harbors the two residues that are shared between the two Tyr pockets (pink). The peptide has been adapted to contain two spatially neighboring Tyr as target amino acids. *Lower*: Schematic drawing of the TyroM6. To generate a full $(KY)_x$ binder, repeat4, as in the double-pocket construct, replaces the internal repeats rep2-5. The rep1 and rep6 in the full construct correspond to rep3 and rep5 in the double-pocket construct. (*C*) Model of the Tyr double pocket adapted from PDB 8QNZ by changing the corresponding residues in the single-pocket structure to obtain a double-pocket model in Pymol. The two shared residues X5 and X7 are shown in pink, all others X1 to X12 are shown in blue. Shown here is the most successful variant with an Ile in X5 and an Glu in X7.

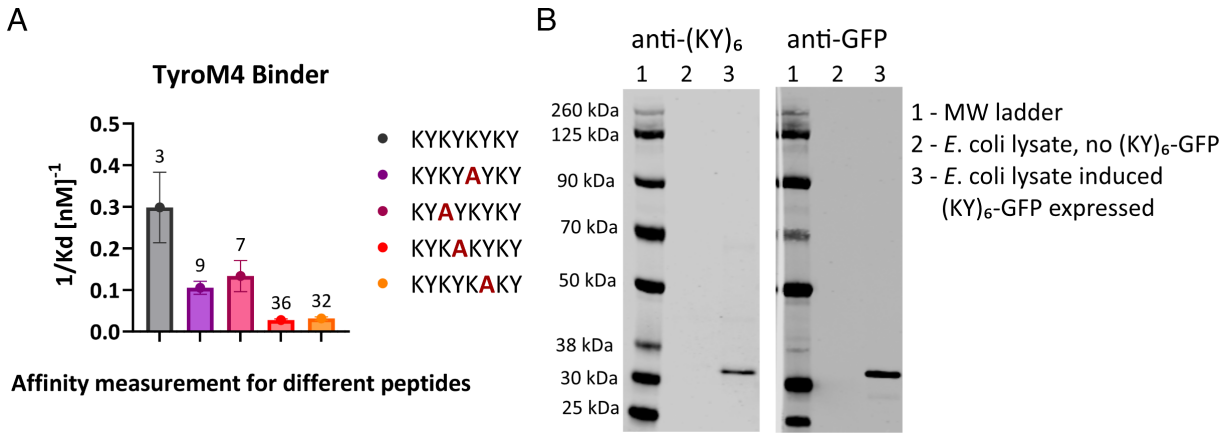

Fig. 7. TyroTag binder analysis. (*A*) K_d data measured by fluorescence anisotropy for the TyrM4 binder shown as a bar graph. The *y* axis is $1/K_d$, while the numbers above the bar represent the measured K_d in nM. The Ala substitution of a Tyr or Lys in the target peptide reduces the affinity by a factor of 3.5 for Lys and a factor of 10 for the Tyr substitutions, regardless of the position. (*B*) Detection of $(KY)_6$-GFP fusion in an *E. coli* lysate in a WB assay. *Left*: HA-tagged TyroM6 dArmRP was used as a primary agent to detect the $(KY)_6$ peptide. The dArmRP was stained by its fused HA-tag, using a primary anti-HA antibody and secondary antibody combination. *Right*: Detection of GFP with standard antibodies. Error bars show SD of duplicate measurements.

necessary but not sufficient, as it does enrich for the target binding and counterselects binders with affinity for the fluorescent label, but it does not select for the absence of cross-reactivity to undesired side chains at the position in question. This is achieved by subsequently applying the target itself as a competitor in a reverse competition setup to remove cross-reactive variants. It was clearly demonstrated that specific binders were more effectively separated by this strategy, compared to the standard competition approach.

In a reverse-competition setup, specific binders are directly identified because they react more exclusively to the designated target as a competitor. Hence, we can directly select binders that do specifically engage the desired target (reverse competition) instead of selecting those that do not bind to the off-targets (standard competition), resulting in better discrimination between the target and off-target binders. Furthermore, we believe that the ability to analyze the binding phenotype for each binder during the FACS experiments is a requirement to achieve the resolution to separate high-affinity but cross-reactive binders from high-affinity selective binders.

The selectivity of the obtained modules has shown to be very high, increasing over the starting consensus derived from natural ArmRP, while discriminating even closely related amino acids by a factor of ten or more. The obtained crystal structure confirms the specific interaction of the target amino acid and its selected module as measured for the K_d characterization.

The selection rounds for the peptide containing His have been analyzed by NGS, and the selection progress was visualized by plotting the unique sequences with UMAP. The observed shift in sequence space, as a result of the reverse competition strategy, emphasizes again the positive effect of this selection setup. Comparing the binder pool before and after the reverse competition selectivity sort has furthermore allowed us to identify low-abundant variants with the desired phenotype from the NGS enrichment progress data alone, almost regardless of their abundance in the output pool. This is possible because an enrichment during the reverse competitor sort can be taken as a measure for the desired specificity phenotype.

As a proof of principle that selected modules were successfully used to generate a modular peptide binder, two dArmRPs targeting a $(KY)_n$ peptide were designed based on the selected Tyr module. The affinity characterization of those binders confirmed that the assembled binder recognizes its designated target peptide and that the affinities are adjustable according to the binder length. The high affinity of the TyrM6 binder exemplifies the affinity range that can be reached by this modular peptide binder approach. Furthermore, we confirmed the potential of the dArmRPs as WB detection agents as a first application since the TyrM6 binder has been shown to specifically recognize its $(KY)_6$-GFP target in an *E. coli* lysate with at least similar sensitivity as the conventional anti-GFP mAb, and without any off-target binding detectable. Current efforts are directed to direct labeling of the dArmRPs, obviating the need for any antibody detection.

These results suggest that binders to arbitrary sequences may now be generated very rapidly and at high throughput, based on a future repertoire of selected binding pockets, and following the strategy outlined herein. Importantly, after a portfolio of pockets has been identified, the selection work described here will be no longer needed, and modular binders can be assembled from the known parts, enormously simplifying the creation of future binders to arbitrary sequences.

Materials and Methods

Target Generation. Initially, chemically synthesized, biotinylated target peptides were preincubated with labeled streptavidin (SA) to detect target binding in yeast flow cytometry experiments. To avoid the avidity effect, the total concentration of peptide–SA was lowered, and more stringent gates were set. However, the effect between affinity and cytometry signal was obscured, suggesting that an avidity effect within the tetrameric peptide–SA complex is the cause (*SI Appendix*, Fig. S1).

To overcome this issue, monovalent targets were produced as peptide fusions to fluorescent proteins, restoring the correlation between peptide binding signal in flow cytometry experiments and fluorescence anisotropy measurements (*SI Appendix*, Fig. S1). To avoid selection of binders against the fluorescent proteins, different versions (mNeonGreen and mRuby) were used either in conjunction or in alternation.

Yeast Surface Display and FACS. The tryptophan-auxotrophic *S. cerevisiae* strain EY100 was used in all experiments. Yeast transformants were grown at 30 °C in SD-CAA media (6.9 g/L yeast nitrogen base without amino acids, 5 g/L casamino acids, 20 g/L *D*-glucose, 35.3 mM tribasic sodium citrate, and 35.3 mM citric acid). Expression of the display construct was induced at an OD_{600} of 1 and cultivated for 24 h at 20 °C in SG-CAA media (6.9 g/L yeast nitrogen base without amino acids, 5 g/L casamino acids, 2 g/L anhydrous *D*-glucose, 18 g/L anhydrous *D*-galactose, 38 mM dibasic sodium phosphate, and 55 mM monobasic sodium phosphate).

Yeast Library Preparation. The yeast display vector pCTCON2 (19) was modified to contain restriction sites directly at the N and C terminus of the dArmRP gene to facilitate the downstream processing of the selected variants. The initial randomized library inserts used for the His selections were ordered at Geneart Synthesis, using trinucleotide technology (20). In the randomized positions, all amino acids were allowed except Pro, Gly, and Cys, with the exception of positions X2 and X7 (Fig. 2), where also Gly was allowed since Gly was observed in some natural ArmRP at these positions.

For the Tyr selection, a second design was used, in which position X3 was limited to the amino acids Val, Thr, Ala, and Ser. This library was created by MAX randomization (21) (*SI Appendix*, Fig. S2) and amplified by PCR using HR primers (HR-fwd: AAT GAA CAA ATC TTG CAA GAG GCC TTG TGG GCC CTC, HR-rev: ATC GAT TAC TGC TTG GAT TTG CTC ATT ACC CCC GCT AG). The pCTCON2 library backbone was linearized by restriction digest with NheI (NEB) and BbvCI (NEB), located within the dArmRP gene, and purified via a PCR purification kit (Qiagen). The yeast libraries were created by homologous recombination of the library insert and the linearized library backbone, and they were used to cotransform EBY100 cells by electroporation. The cells were made competent using the lithium acetate method (22). For the transformation, the cell suspension was mixed with 3 μg linearized backbone and 6 μg library insert, transferred to a chilled cuvette (0.2 cm Gene Pulse, BioRad), and pulsed at 1.5 kV, 25 mF, and 200 V using Gene Pulser Xcell (Bio-Rad). The recovered cells were cultivated in SD-CAA until OD_{600} of 5 was reached. After transformation, dilution series were plated to estimate the actual library size.

Flow Cytometry. Flow cytometry experiments were performed in PBSAE buffer (137 mM sodium chloride, 3 mM potassium chloride, 8 mM dibasic sodium phosphate, 1.5 mM monobasic potassium phosphate, 1 g/L bovine serum albumin, and 2 mM EDTA). Cell staining was performed on ice using 5×10^7 cells/50 μL PBSAE buffer. The required expression culture volume was collected, and the pellet was resuspended in 50 μL mouse anti-c-myc antibody (Sigma, M4439), 1/1,000 diluted in PBSAE. After incubating for 30 min while rotating at 4 °C, the yeast cells were centrifuged (4,000 g, 5 min), and the pellet was resuspended in 50 μL PBSAE with goat anti-mouse-AF647 (Invitrogen, A21235, 1/1,000 diluted) and the desired concentration of peptide-mNeongreen or -mRuby3 fusions. The mixture was incubated for 60 min while rotating at 4 °C. Subsequently, the cells were centrifuged (4,000 g, 5 min), and the pellet was resuspended in 200 μL PBSAE directly before subjecting the cells to flow cytometry analysis or FACS.

Plasmid Construction for ArmRP Expression. For protein expression, a pQE30LIC_3C-based vector, which contains a 3C-protease-cleavable MRGSHis6-tag, was used (14). Selected dArmRPs variants were subcloned from the display vector pCTCON2 using restriction enzymes BamHI and HindIII. DNA fragments coding for rationally designed dArmRP pockets or assemblies were ordered from IDT or Twist Bioscience and subcloned using restriction enzymes NheI and BbvCI. The target-peptide fusions were obtained by ligation of the BamHI- and KasI-digested expression vector variant with the sticky-end dsDNA insert encoding the target peptide that was obtained by heat-induced annealing of the respective ssDNA sequences. To obtain mCherry-dArmRP fusions, the dArmRP were subcloned using restriction enzymes BamHI and HindIII in an expression vector coding for an MRGSHis6-tag-3C-mCherry fusion N-terminal to the BamHI site. All construct sequences were confirmed by DNA sequencing (Microsynth).

Expression and Purification. Protein expression was carried out in autoinduction medium for 15 h at 25 °C as described in ref. 23. The expression cultures were centrifuged (5,000 *g*, 5 min) and resuspended in resuspension buffer (RB) (50 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole). Protease inhibitors Petabloc SC (50 μg/mL), Leupeptin (5 μg/mL), and Pepstatin-A (1 μg/mL) were added. The cells were lysed by sonication, and the insoluble contents were removed by centrifugation (25,000 *g*, 20 min). The crude extracts were loaded onto Ni-NTA superflow resin columns (3 mL, Qiagen), and 30 column volumes (CV) of RB were applied for washing. Elution was achieved using 2.5 CV of EB (RB with 300 mM imidazole). Human rhinovirus 3C-protease (2% w/w) was mixed with eluted protein fractions, and the reaction mixtures were dialyzed against PBS (137 mM sodium chloride, 3 mM potassium chloride, 8 mM dibasic sodium phosphate, and 1.5 mM monobasic potassium phosphate) to cleave off the His-tags. Proteins destined for crystallization were dialyzed against crystallization buffer (CB) (50 mM Tris pH 7.4 and 300 mM NaCl). Protease and uncleaved proteins were removed by reverse IMAC. The monomeric protein fractions were isolated by SEC on an ÄKTA Explorer chromatography system using a HiLoad 16/60 Superdex 200 pg column and the same buffers as used for dialysis.

Affinity Characterization. For K_d determination, fluorescence anisotropy measurements were performed as described in ref. 14. As reporter, target peptide-sfGFP or -mNeongreen fusions were used. The assay was performed in black nonbinding 96-well plates (Greiner bio-One). All measurements were performed in PBS supplied with 0.05 % (v/v) Tween 20. Binding kinetics were determined using a FRET binding assay as described in refs. 14 and 15. The dArmRP variants were produced as mCherry fusions for the FRET assay.

Western Blot. The newly generated dArmRP binders were HA-tagged to be used as primary detection agents. The C-terminal HA tag was detected with commercial antibodies, primary mouse anti-HA (Sigma), and secondary goat anti-mouse Daylight790 conjugate (Biotium). As positive control, GFP was detected with primary mouse anti-GFP (Cell Signaling) and secondary goat anti-mouse Daylight790 conjugate (Biotium). The assay was performed in TBS (137 mM NaCl and 50 mM Tris). The samples were separated by SDS/PAGE Mini-PROTEAN TGX™ 4 to 20% gels and blotted onto a nitrocellulose membrane using a Trans-Blot® Turbo™ transfer system (BioRad) according to the manufacturer's protocol (mixed MW, 25 V, 7 min). The membranes were air dried at RT for 1 h and blocked for 20 min, RT, with TBSC [TBS containing 1× Casein Blocking buffer (Thermo Scientific)]. The primary incubation with the HA-tagged dArmRP (5 μg/mL) or anti-GFP mAb (1:1,000 diluted) was performed overnight at 4 °C in 5 mL TBSC. The membranes were washed with TBST (TBS supplied with 0.1% Tween 20) and incubated at RT for 1 h with the anti-HA antibody (1:2,000) in 5 mL TBST. The incubation step was also performed for the GFP membrane with 5 mL TBST. The membranes were washed with TBST and incubated at RT for 1 h with the secondary mAb in 5 mL TBST. After a final wash, the fluorescence intensity was detected on an Odyssey imaging system (LI-COR Biosciences).

Crystallization. The crystallization chaperone DARPin D12 (24) and a GGSGG-linked target peptide were N-terminally fused to the selected dArmRP variants. The proteins were concentrated to 40 to 70 mg/mL in 50 mM Tris/HCl at pH 8 and 300 mM NaCl using Amicon centrifugal concentrators (Amicon Ultra Centrifugal Filters, Merck Millipore). Sitting drop vapor diffusion crystallization was performed in Intelli-Plates R96-3 LV (Art Robbins, California) with a mother-liquor-to-protein ratio of 1:1, 2:1, and 3:1 per condition. The plates were set up and kept at 20 °C. Prior to data collection, the crystals were incubated for 10 s in the mother liquor containing 20% (v/v) ethylene glycol and flash-frozen in liquid nitrogen. Data were collected at the Paul Scherer Institute with the Synchrotron light source (PSI, Villigen, Switzerland). The data were processed using XDS, XSCALE, and XDSCONV (25) for analysis. Molecular replacement using CCP4 with Phaser (26) was employed to solve the structure. The refinement was done in CCP4 using refmac5 (27) and in Phenix with PhenixRefine (28). The Tyr module structure was obtained in 2 M $(NH_4)_2SO_4$ and 0.1 M Bis-Tris, pH 5.5. Data collection and refinement statistics are summarized in *SI Appendix*, Table S2.

NGS. Yeast cells were lysed by mixing an approximately 1 mm diameter globule of freshly grown cells with 10 μL 0.02 M NaOH and incubated for 10 min at 95 °C. Five microliters of lysate was used as PCR template to amplify the NGS amplicon using Taq polymerase and the provided Q-solution (HotStarTaq DNA Polymerase, Cat No./ID: 20320). The adaptor ligation, the Illumina Miseq run, and preprocessing of the data were performed by the Functional Genomics Center Zurich. From the obtained FASTQ files, approximately 99% of the reads were paired, and these were combined into single reads by FLASH (29) with a minimum/maximum overlap of 230/300 base pairs. Single reads were then mapped back to the wild-type dArmRP sequence using Bowtie2 (30) with triple penalties for gap opening and extension and the noncoding sequences excluded. The remaining sequences were clustered at 100% identity by a custom Python script. To visualize the highly dimensional sequence space, sequences were one-hot encoded and mapped to two dimensions by UMAP (18).

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank A. Henning for the initial library designs as well as her work on the identification of the most suitable display system and Y. Waltenspühl for the helpful discussions, especially with regard to the selection strategies, and J. Kapp for general advice and inputs concerning the western blot experiments. We also would like to acknowledge J. Wepfer for his contribution during the selection strategy development and V. Fabricius for his practical work on the initial TyrMX binder characterization. This work was supported by Schweizerische Nationalfonds BRIDGE Program grant 20B2-1_176535 (to A.P.), European Union Horizon 2020 FETOPEN program PRe-ART, grant agreement No. 764434, and European Union EIC Transition program PRe-ART-2T, grant agreement No. 101058027.

Author contributions: Y.S. and A.P. designed research; Y.S., F.M., P.E., A.C., and M.A. performed research; J.R.J. and A.V.H. contributed new reagents/analytic tools; Y.S., J.R.J., P.E., and A.V.H. analyzed data; and Y.S. and A.P. wrote the paper.

Competing interest statement: A patent has been filed on the obtained sequences from this work. It is owned by the University of Zurich.

This article is a PNAS Direct Submission. A.N.L. is a guest editor invited by the Editorial Board.

1. C. A. Borrebaeck, Antibodies in diagnostics–From immunoassays to protein chips. *Immunol. Today* **21**, 379–382 (2000).
2. A. Bradbury, A. Plückthun, Reproducibility: Standardize antibodies used in research. *Nature* **518**, 27–29 (2015).
3. H. K. Binz, P. Amstutz, A. Plückthun, Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23**, 1257–1268 (2005).
4. H. K. Binz, A. Plückthun, Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.* **16**, 459–469 (2005).
5. N. London, D. Movshovitz-Attias, O. Schueler-Furman, The structural basis of peptide-protein binding strategies. *Structure* **18**, 188–199 (2010).
6. E. Petsalaki, A. Stark, E. Garcia-Urdiales, R. B. Russell, Accurate prediction of peptide binding sites on protein surfaces. *PLoS Comput. Biol.* **5**, 1–10 (2009).
7. T. Pawson, P. Nash, Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452 (2003).
8. F. Diella *et al.*, Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* **13**, 6580–6603 (2008).
9. K. Wu *et al.*, De novo design of modular peptide-binding proteins by superhelical matching. *Nature* **616**, 581–589 (2023).
10. F. Parmeggiani *et al.*, Designed armadillo repeat proteins as general peptide-binding scaffolds: Consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.* **376**, 1282–1304 (2008).
11. M. Hatzfeld, The armadillo family of structural proteins. *Int. Rev. Cytol.* **186**, 179–224 (1999).
12. P. Alfarano *et al.*, Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. *Protein Sci.* **21**, 1298–1314 (2012).
13. S. Hansen *et al.*, Curvature of designed armadillo repeat proteins allows modular peptide binding. *J. Struct. Biol.* **201**, 108–117 (2018).
14. S. Hansen *et al.*, Structure and energetic contributions of a designed modular peptide-binding protein with picomolar affinity. *J. Am. Chem. Soc.* **138**, 3526–3532 (2016).
15. P. Ernst *et al.*, Structure-guided design of a peptide lock for modular peptide binders. *ACS Chem. Biol.* **15**, 457–468 (2020).
16. V. Puri, E. Streaker, P. Prabakaran, Z. Zhu, D. S. Dimitrov, Highly efficient selection of epitope specific antibody through competitive yeast display library sorting. *mAbs* **5**, 533–539 (2013).
17. B. T. Bajar *et al.*, Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep.* **6**, 20889 (2016).
18. L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv [Preprint]. https://arxiv.org/abs/1802.03426 (Accessed 1 November 2023).
19. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768 (2006).
20. B. Virnekäs *et al.*, Trinucleotide phosphoramidites: Ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res.* **22**, 5600–5607 (1995).
21. A. Chembath *et al.*, Nondegenerate saturation mutagenesis: library construction and analysis via MAX and ProxiMAX randomization. *Methods Mol. Biol.* **2461**, 19–41 (2022).
22. J. A. Van Deventer, K. D. Wittrup, Yeast surface display for antibody isolation: Library construction, library screening, and affinity maturation. *Methods Mol. Biol.* **1131**, 151–181 (2014).
23. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein. Expr. Purif.* **41**, 207–234 (2005).
24. Y. Wu *et al.*, Rigidly connected multispecific artificial binders with adjustable geometries. *Sci. Rep.* **7**, 11217 (2017).
25. W. Kabsch, XDS. *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
26. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
27. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta. Crystallogr. D. Biol. Crystallogr.* **67**, 355–367 (2011).
28. P. V. Afonine *et al.*, Joint X-ray and neutron refinement with phenix.refine. *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 1153–1163 (2010).
29. T. Magoč, S. L. Salzberg, FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
30. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).