Protein Engineering, Design and Selection, 2024, **37**, gzae007
https://doi.org/10.1093/protein/gzae007
Original Article

PEDS PROTEIN ENGINEERING, DESIGN & SELECTION
OXFORD

DexDesign: an OSPREY-based algorithm for designing *de novo* D-peptide inhibitors

Nathan Guerin[1], Henry Childs[2], Pei Zhou[3], Bruce R. Donald[1,2,3,4,*]

[1]Department of Computer Science, Duke University, 308 Research Drive, Durham, NC 27708, United States
[2]Department of Chemistry, Duke University, 124 Science Drive, Durham, NC 27708, United States
[3]Department of Biochemistry, Duke University School of Medicine, 307 Research Drive, Durham, NC 22710, United States
[4]Department of Mathematics, Duke University, 120 Science Drive, Durham, NC 27708, United States

*Corresponding author. Department of Computer Science, Duke University, Durham, NC 27708, United States. E-mail: brd+peds23@cs.duke.edu

Edited by: Dr Chica Roberto

Abstract

With over 270 unique occurrences in the human genome, peptide-recognizing PDZ domains play a central role in modulating polarization, signaling, and trafficking pathways. Mutations in PDZ domains lead to diseases such as cancer and cystic fibrosis, making PDZ domains attractive targets for therapeutic intervention. D-peptide inhibitors offer unique advantages as therapeutics, including increased metabolic stability and low immunogenicity. Here, we introduce DexDesign, a novel OSPREY-based algorithm for computationally designing *de novo* D-peptide inhibitors. DexDesign leverages three novel techniques that are broadly applicable to computational protein design: the Minimum Flexible Set, K*-based Mutational Scan, and Inverse Alanine Scan. We apply these techniques and DexDesign to generate novel D-peptide inhibitors of two biomedically important PDZ domain targets: CAL and MAST2. We introduce a framework for analyzing *de novo* peptides—evaluation along a replication/restitution axis—and apply it to the DexDesign-generated D-peptides. Notably, the peptides we generated are predicted to bind their targets tighter than their targets' endogenous ligands, validating the peptides' potential as lead inhibitors. We also provide an implementation of DexDesign in the free and open source computational protein design software OSPREY.

Key words: algorithms; D-peptide; OSPREY; peptide design; DexDesign

Introduction

Since the 1921 discovery (Banting 2024) of the peptide hormone insulin to treat diabetes, many peptides and peptide-derived therapeutics have come into clinical use, with more than 30 achieving final regulatory approval just since the year 2000 (Wang et al. 2022a). The use of peptides as therapeutics has a number of advantages, including standard protocols for synthesis, good efficacy, high potency, and selectivity (Craik et al. 2013, Fosgerau and Hoffmann 2015). On the other hand, peptide therapeutics have a number of drawbacks, including poor stability, oral bioavailability, membrane permeability, and retention (Craik et al. 2013). The substitution of D-amino acids (*Dexter*-amino acids that are mirror images of L-amino acids) for L-amino acids in peptides is one strategy medicinal chemists have used to address these shortcomings. The following sections describe the background, benefits of D-peptides, previous work in noncanonical design, and contributions of DexDesign (i.e. *Dexter*Design), an algorithm we developed and incorporated into the protein design software OSPREY to generate and analyze *de novo* D-peptide inhibitors of two biomedically important PDZ domains targets: CAL and MAST2.

Benefits of including D-amino acids in peptides

The inclusion of D-amino acids can increase peptide stability by decreasing the substrate recognition by proteolytic enzymes (Di 2014). For example, Chen *et al.* improved both stability and binding affinity of a bicyclic peptide inhibitor of the cancer-related protease urokinase-type plasminogen activator by substituting a single D-serine for a glycine (Angelini et al. 2012, Chen et al. 2013). Haugaard-Kedström et al. observed that the simple substitution of D-amino acids in two positions of their *de novo* PDZ domain inhibitor greatly improved metabolic stability by increasing their peptide's half-life 24-fold (Haugaard-Kedström et al. 2021). More ambitious uses of D-amino acids have also been performed. Liu et al. constructed an entirely D-peptide inhibitor of the MDM2 oncoprotein using mirror image phage display, an experimental technique used to discover D-peptide drug candidates, that inhibited growth of glioblastoma both in cell culture and nude mouse xenograph models (Liu et al. 2010). Nevertheless, the challenge of preparing an enantiomeric protein target for mirror image phage display remains a drug-discovery bottleneck (Lander et al. 2023).

PDZ domains

With over 270 unique occurrences in more than 150 human proteins, PDZ domains constitute the largest family of peptide-recognition domains in the human genome (Amacher et al. 2020). A typical PDZ domain has 80–100 amino acids folded into five core β-strands (β1-β5) and two α-helices (α1 and α2) (Lee and Zheng 2010, Ivarsson 2012, Amacher et al. 2020). They facilitate a variety of cellular functions, such as modulating polarization, signaling, and trafficking pathways, through interaction with short linear motifs (SLiMs) located at the C-terminus of their ligands (Harris and Lim 2001, Jemth and Gianni 2007, Christensen et al. 2019, Amacher et al.

Received: September 3, 2023. **Revised:** April 17, 2024. **Editorial decision:** April 22, 2024

2020). Usually SLiMs bind into a groove of the PDZ domain between $\alpha 2$ and $\beta 2$, extending the $\beta 2/\beta 3$ sheet (Jemth and Gianni 2007). Modulating the interaction between a SLiM and its PDZ binding partner is a strategy that both viruses and therapeutics aim to exploit (Davey et al. 2011, Amacher et al. 2020) and has been explored by previous computational design techniques (Smith and Kortemme 2010, Roberts et al. 2012, Smith et al. 2013, Melero et al. 2014, Nakariyakul et al. 2014, Zheng et al. 2015, Mignon et al. 2017, Opuu et al. 2020, Panel et al. 2021).

CFTR-associated ligand

Mutations in the cystic fibrosis transmembrane conductance regulator (CFTR), such as ΔF508, cause destabilized, misfolded, and less efficient CFTR (Roberts et al. 2012, Dougherty et al. 2020). The CFTR-associated ligand (CAL) binds CFTR *via* CAL's PDZ domain (CALP), which shepherds CFTR through rapid degradation *via* a lysosomal pathway (Roberts et al. 2012). Our lab used computational peptide design to develop a hexamer that bound 170-fold more tightly to CALP than CALP bound the CFTR C-terminus, rescuing CFTR activity in monolayers of polarized human upper airway epithelial cells that contain the ΔF508 deletion in CFTR—80% of cystic fibrosis patients are homozygous for this mutation (Roberts et al. 2012, Holt et al. 2019). Cyclic peptides (Dougherty et al. 2020) have also been developed.

MAST2

During viral infection, the rabies virus exploits SLiM/PDZ-domain interactions to further its propogation (Caillet-Saguy et al. 2015, Delhommel et al. 2015). The rabies virus glycoprotein's C-terminal residues interact with MAST2's PDZ domain, disrupting the ability of MAST2 and PTEN to form a complex and inhibits neurite outgrowth and apoptosis (Préhaud et al. 2010, Caillet-Saguy et al. 2015, Khan et al. 2019). Recognizing the therapeutic potential of promoting neurite outgrowth in the treatment of neurodegenerative disease, Khan et al. developed three peptides that mimic and improve upon the rabies virus glycoproteins' interaction with MAST2's PDZ domain, stimulating neurite outgrowth in proportion to the affinity the peptide bound MAST2 (Khan et al. 2019).

Computational tools and algorithms for designing D-peptides

OSPREY is a free and open-source software program containing a suite of computational protein design algorithms developed in our lab (Hallen et al. 2018). OSPREY has been used to, among many other things, design and structurally characterize peptide inhibitors of CALP for treating CFTR (Roberts et al. 2012, Holt et al. 2019), predict resistance mutations in bacteria (Frey et al. 2010, Reeve et al. 2015) and cancer (Guerin et al. 2022), and improve broadly neutralizing antibodies against HIV-1 (Rudicell et al. 2014, Holt et al. 2023). OSPREY has been used to computationally redesign proteins with canonical and non-canonical amino acids (Lilien et al. 2005, Roberts et al. 2012, Lowegard et al. 2020, Wang 2021), as well as optimize protein: small molecule interactions (Kaserer and Blagg 2018, Guerin et al. 2022, Guerin et al. 2023), but until now has not had the capability to design D-peptides. Given the promising biomedical potential of D-peptides (Liu et al. 2010, Wei et al. 2015, Wang et al. 2017, Miles et al. 2018, Zhou et al. 2020a, Doti et al. 2021), having the ability to apply OSPREY's ensemble-based, provable protein design algorithms in pursuit of D-peptide design could greatly decrease the required quantity of expensive, time-intensive experiments.

There are a few previous computational techniques for D-peptide design (Donald 2011). One of the earliest was by Elkin et al., which used the Multiple Copy Simultaneous Search (Miranker and Karplus 1991) method to predict candidate D-peptide inhibitors of hepatitis delta antigen dimerization (Elkin et al. 2000). Recent versions of Rosetta have included functionality to incorporate non-canonical and D-amino acids (Renfrew et al. 2012, Bhardwaj et al. 2016). Philip Kim's group has developed a computational D-peptide design technique based on creating a mirror image of the PDB, identifying hotspot interactions, and searching the D-PDB for similar configurations of hotspot residues (Garton et al. 2017, 2018). They applied this technique to develop two D-peptide inhibitors to the SARS-CoV-2 spike protein receptor binding domain and the human angiotensin-converting enzyme 2 (ACE2) that mimic the ACE2 $\alpha 1$-binding helix (Valiente et al. 2021, 2022). Overall, the number of algorithms the protein designer has available for D-peptide design is notably sparser than for L-design, and the development of additional computational protocols for this important task is warranted.

DexDesign

In this paper we present a computational protocol, DexDesign, for designing *de novo* D-peptides in OSPREY. DexDesign constructs D-peptide scaffolds by mirroring the structure of a L-protein: peptide complex into D-space (D-amino acids), then uses the geometric search algorithms in MASTER (Zhou and Grigoryan 2015) to search hundreds of thousands of L-protein structures for substructures with backbones similar to the D-peptide. It then uses the iMinDEE/K* algorithm (Georgiev et al. 2008, Gainza et al. 2012) in OSPREY to redesign a scaffold D-peptide's sidechains to optimize target binding (see Fig. 1). We use DexDesign to predict D-peptides inhibitors of CALP and MAST2 PDZ domains.

Methods

Algorithm and computational protocol

DexDesign generates *de novo* D-peptides by combining MASTER's molecular structure search (Zhou and Grigoryan 2015) with provable computational protein redesign algorithms in OSPREY (Donald 2011, Hallen et al. 2018), mediated *via* energy-equivariant geometric transformations (EEGT). EEGTs, such as translation, rotation, or reflection, are geometric transformations of a molecular structure that do not affect the energy of that structure. Each EEGT corresponds to a symmetry in the energy field (Noether 1983). For example, an energy function will compute the same energy of protein structure s and s reflected over the Cartesian x-y plane. The MASTER algorithm searches a database of protein structures for a user-specified query structure and is guaranteed to find all protein substructures in the database with a backbone RMSD below a cutoff threshold (Zhou and Grigoryan 2015). The K* algorithm in the OSPREY software suite (Hallen et al. 2018) searches for amino acid substitutions that maximize a design objective, such as binding affinity or specificity, while translating and rotating the ligand (Lilien

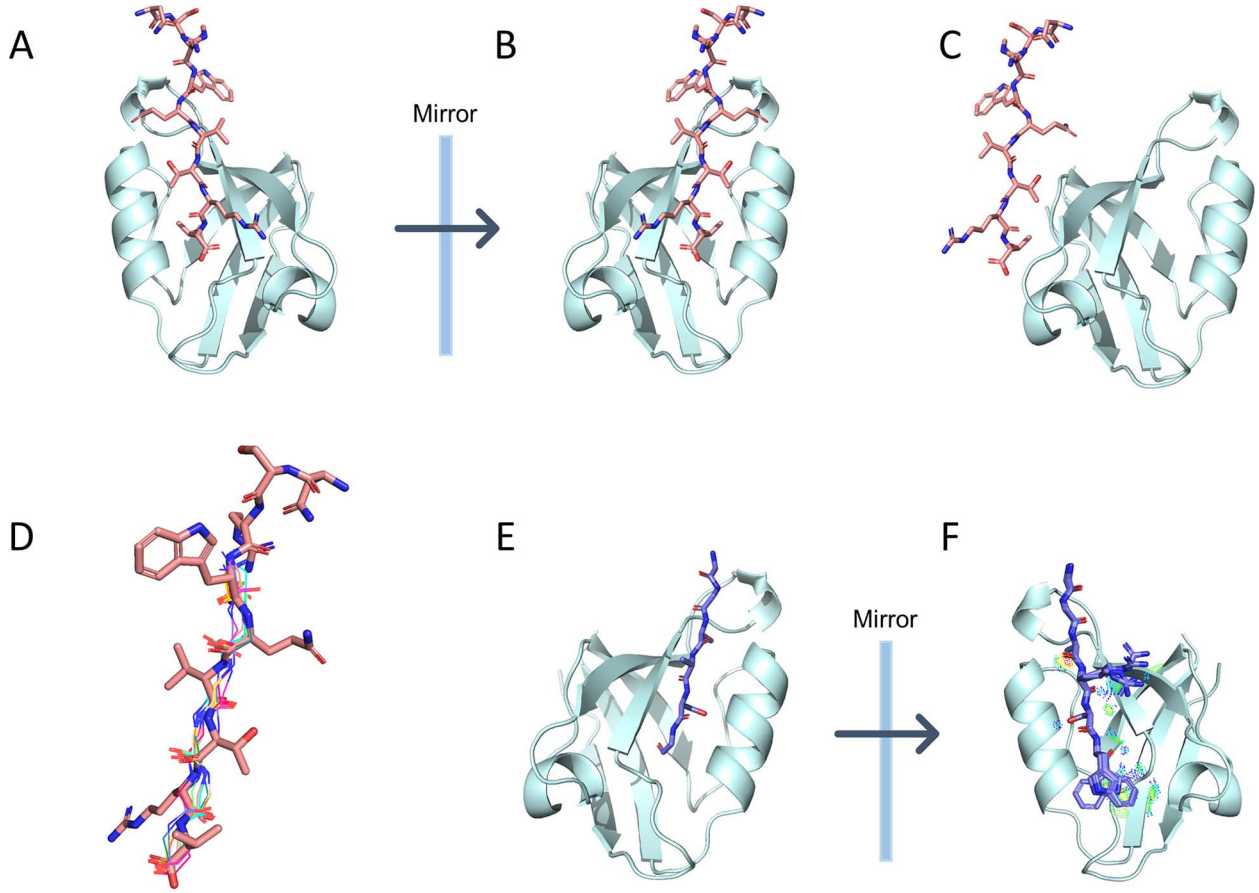

Figure 1. Example of the DexDesign protocol applied to CALP. (A) The potent inhibitor kCAL01 (in pink) bound to CALP (in cyan) (PDB ID 6ov7) (Holt et al. 2019) is used as a starting point for a DexDesign search for a D-peptide inhibitor of CALP. Both kCAL01 and CALP are composed of solely L-amino acids. (B) The input structure is reflected to produce a mirror-image of the kCAL01: CALP complex, which flips the chirality of all amino acids to D. (C) The complex is split into its constituent peptide and protein components. (D) Residues P^0 to P^{-5}, which are the residues located within CALP's binding pocket, are used as the query structure to conduct a MASTER (Zhou and Grigoryan 2015) search of a large database of L-protein structure to find substructures with similar backbones (as determined by backbone RMSD) to the D-version of kCAL01. 10 representative matches of L-peptide segments (in multicolored wire representation) are overlaid on the pink stick D-version of kCAL01. (E) Each L-peptide match is aligned to the D-version of kCAL01 in the D-kCAL01: CALP complex structure and D-kCAL01 is removed. Shown (in purple sticks) is an L-peptide substructure GGAASG (residues 168–173) that MASTER identified in *Mycobacterium tuberculosis* Rv0098 (PDB ID 2PFC) (Wang et al. 2007). This L-peptide forms the basis for OSPREY redesign. (F) The L-peptide: D-CALP complex is reflected once again to form an D-peptide: L-CALP complex. The K* algorithm (Lilien et al. 2005, Gainza et al. 2012) in OSPREY (Hallen et al. 2018) is then invoked to conduct a search over D-peptide sequences and continuous sidechain conformations to optimize the D-peptide for binding. K* identified two mutations at positions P^0 and P^{-2} predicted to improve binding of the peptide with a normalized $\Delta\Delta G$ of -1.4 kcal/mol, improving K_D by 9-fold (see Supplementary Information B for information on the normalization procedure). Position P^0 is mutated from Gly to Trp, and P^{-2} is mutated from Ala to Arg. Shown is an OSPREY-predicted low energy ensemble of the D-peptide GGARSW with Molprobity probe dots (Word et al. 1999, Williams et al. 2018, Jou et al. 2024) showing goodness-of-fit the OSPREY-predicted mutated D-sidechains make with CALP.

et al. 2005, Georgiev et al. 2008, Ojewole et al. 2018, Jou et al. 2020). It does this by exploiting molecular ensembles to compute a provably accurate ε-approximation to the binding constant, K_a (see Supplementary Information A: the K* algorithm) (Lilien et al. 2005, Georgiev et al. 2008). In essence, DexDesign invokes MASTER search as a subroutine to suggest D-peptide scaffolds with backbone conformations similar to their L-peptide counterpart, then invokes K* as a subroutine to optimize amino acid sequences and side chain conformations on those scaffolds.

After preparing a database (DB) of L-protein structures, a protein designer initiates DexDesign by identifying a protein target (t) of interest, for which there exists a structure of a protein or peptide (p) bound to t. In MASTER terminology, the structure of this bound complex will become our query, q_{tp}. Below, we define terms used in the DexDesign algorithm:

1. Let s_n be a protein structure with n residues. We define substructure $s_{i,j}$ of s, where $1 \leq i < j \leq n$, to be a structure of residues i through j of s.
2. Let $r(s, a)$ be a function that reflects all atoms in protein structure s across a plane a. Without loss of generality, we let a be the x-y plane and define $r(s) = r(s, a)$ henceforth. We note that r is an involution, i.e. $r(r(s)) = s$.
3. Let $M(DB, s, c)$ be the MASTER subroutine. M returns a set of substructures from DB with backbone RMSD, when optimally aligned with protein substructure s, less than c Å.
4. Let $\mathcal{O}(p, t)$ be the OSPREY K* subroutine. $\mathcal{O}$ redesigns peptide p towards increased binding affinity with protein target t by searching over mutated and continuously minimized amino acid sidechains, and returns a set of mutant sequences (and structural molecular ensembles) derived from p that have improved binding with t.

```
DexDesign

Inputs:
q_TP: The L Target:Peptide bound complex query
DB: A database of L-proteins
c: Maximum difference RMSD (in Å)
Output:
Set of L targets bound to K*-redesigned D-peptides

1. d_TP          ← r(q_TP)
2. (d_T, d_P)    ← Split(d_TP)
3. L_M           ← M(DB, d_P, c)
4. {K*, l_T, d_P} ← O(r(d_T), r(l_P)) for l_P in L_M
5. Return {K*, l_T, d_P}
```

Figure 2. The DexDesign Algorithm. It takes as input a query structure (q) of a bound target (t) and peptide (p), a database (DB) of L-protein structures, and a cutoff (c). Line 1 reflects the L-protein complex into D-space. Line 2 splits the target and peptide into two structures, the D-target (d_T) and the D-peptide (d_P). Line 3 calls MASTER (Zhou and Grigoryan 2015) (M) to search the L-protein database for all substructures with a backbone RMSD to d_P less than c. The set of results is saved in L_M. Line 4 reflects each MASTER peptide (l_P) and D-target (d_T), to make a D-version of l_P bound to the original L-target. OSPREY (Hallen et al. 2018) K* redesign (Lilien et al. 2005, Gainza et al. 2012) is then run on each target: peptide complex ($l_T + d_P$), resulting in a set of K* scores (K*), along with an OSPREY-predicted structural ensemble of the D-peptide (d_P) and L-target (l_T) complexes, with the sequence and continuously minimized sidechains of d_P optimized to bind l_T. The K* scores and computed structural ensembles are returned on Line 5.

The DexDesign algorithm is described in Figs 1 and 2.

Capabilities added to OSPREY

Customize existing or add new conformation libraries

In previous works (Kaserer and Blagg 2018, Reeve et al. 2019, Guerin et al. 2022, Wang et al. 2022b, Kugler et al. 2023), a typical OSPREY-based computational protein redesign entailed 1) selecting a starting molecular structure, 2) adding hydrogens, 3) specifying the design algorithm and its input parameters, 4) running the algorithm, and 5) analyzing the results. To enable Step 4, OSPREY included a default library of amino acid atom connectivity templates from Amber (Case et al. 2022) and rotamers from Lovell et al. (Lovell et al. 2000). These templates and rotamers then became starting points for continuous minimization within a voxel during the sequence and computational search (Georgiev et al. 2008, Hallen and Donald 2017). Embedding the templates within the algorithm provided protein designers with simple defaults for the majority of protein redesign problems. And while some works (Stevens et al. 2006, Wang 2021, Holt et al. 2023) have expanded these defaults in certain cases, such as in our use of non-canonical amino acids to design CALP inhibitors (Wang 2021), the necessity of providing a simple, general approach that enabled protein designers to experiment with diverse and novel biochemical building blocks remained. The implementation of a general, in contrast to application-specific, approach to modeling templates and flexibility enables designers to design proteins with chemistries that the creators of the protein design software did not even anticipate!

To meet this need, we have simplified the process of specifying rotamers, voxel-based continuous minimization, new molecular fragment templates, or even entire conformation libraries in OSPREY. OSPREY continues to provide intelligent defaults, but they are moved from deep within the software and are now exposed to the designer, allowing the designer to modify them as needed in a simple graphical user interface (see Fig. 3, right). This seemingly simple change has profound implications for OSPREY. When the complete conformation space specification (i.e. the design parameters such as the mutable residues, the flexible residues, etc. See Applying DexDesign to CALP and MAST2 for further definition) is a user-modifiable input to the algorithm, new classes of design capabilities, such as design with D-amino acids *via* DexDesign, are unlocked.

An algorithm for D-protein/peptide design

The molecular interaction forces between two molecules are invariant over a reflection of those two molecules. Put another way, if $K_D\,(x, y)$ is the dissociation constant for protein x binding protein y, then $K_D\,(x, y) = K_D\,(r(x), r(y))$. The OSPREY energy function, as described in detail in previous works (Georgiev and Donald 2009, Gainza et al. 2012, Hallen et al. 2018, Martin and Donald 2024), mimics this physics precisely by being (exactly) energy equivariant with respect to reflection, allowing us to add the ability to design D-proteins and peptides in OSPREY (see Supplementary Table 1 for examples of stereogenic atom configurations upon reflection, and Supplementary Fig. 1 for a visualization). We accomplished this by reflecting OSPREY's default L-conformation library into D-space. A protein designer can now use the functionality described in Customizing existing or add new conformation libraries to specify a D- or L- conformation library on a per-protein basis (see Fig. 3, left). DexDesign requires this capability because designing a D-peptide targeting an L-protein requires the use of both D- and L-conformation libraries.

Applying DexDesign to CALP and MAST2

To use DexDesign to predict *de novo* D-peptide inhibitors to CALP and MAST2, we started with structures of their bound complexes: kCAL01 bound to CALP (PDB ID 6ov7) (Holt et al. 2019) and PTEN bound to MAST2 (PDB ID 2kyl) (Terrien et al. 2010). We created a database of high-resolution L-protein structures by mining the RCSB PDB (Berman et al. 2000) for crystallographically determined structures with a resolution better than 2.5 Å, omitting DNA, RNA, and small molecules. This resulted in a database containing 119,160 structures (see Supplementary Fig. 2 for further description of the composition of the database). Using the DexDesign algorithm in Fig. 2, we first reflected each molecular structure to D-space and split the peptide and target PDZ domain into two separate structures, d_p and d_t, respectively. We then used the MASTER algorithm (Zhou and Grigoryan 2015) to query the database for L-protein substructures with backbones similar to d_p. MASTER returns a set of candidate L-peptides (L_M), each of which ($l_p \in L_M$) we superimposed over d_p in the d_p: d_t complex and subsequently removed d_p. We then again reflect each bound complex l_p:d_t, resulting in a *de novo* D-peptide candidate $r\,(l_p)$ bound to the original L-protein target $r\,(d_t)$ in a complex $r\,(l_p)$:$r\,(d_t)$.

Before executing Step 4 of the DexDesign algorithm (K* redesign; see Fig. 2) we further pruned the set of D-peptide candidates based on two additional criteria. First, we visualized candidate D-peptide: L-protein complex structures in PyMol (Schrödinger 2015) using Molprobity dots (Word et al. 1999, Williams et al. 2018) in our lab's Protein Design Plugin (Jou et al. 2024) to evaluate the number and severity of steric clashes, as steric clashes need to be resolved *via* sequence mutation and additional modeling of continuous side chain flexibility in the K* algorithm. Since clash resolution and

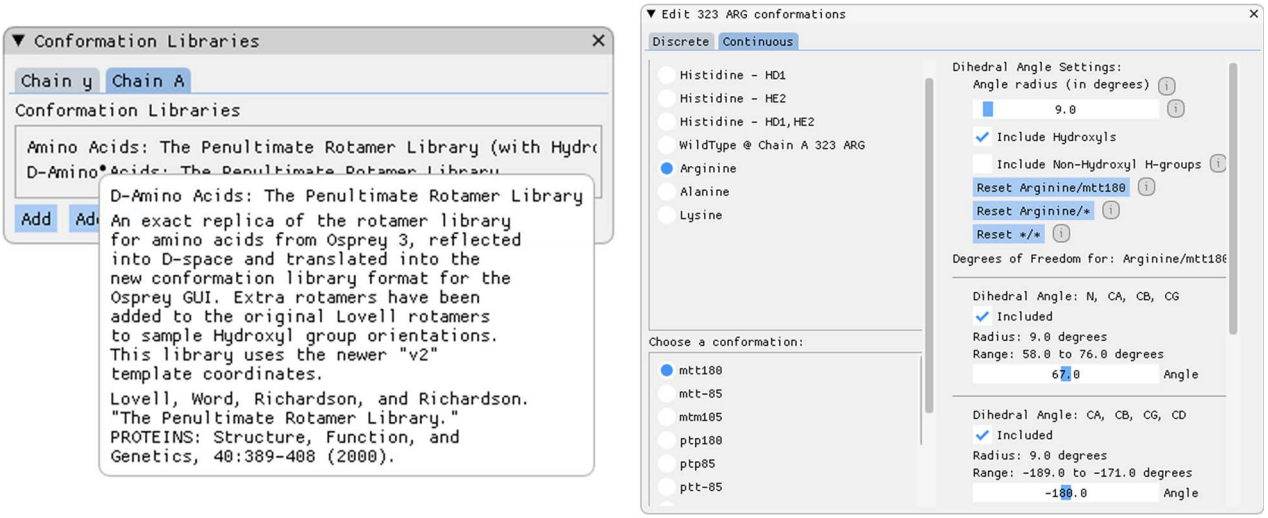

Figure 3. Screenshots of OSPREY protein design specification options. OSPREY (Hallen et al. 2018) now allows protein designers to add their own conformation libraries and easily control rotamer selections and allowed movements. Left: DexDesign includes, in addition to the standard L-conformation library, a D-conformation library that is the mirror image of the L-library. A conformation library describes the standard connectivity templates, rotamers, and allowed movements, all of which can be further customized by the protein designer. A protein designer can specify multiple, distinct conformation libraries per chain. Right: New detailed control over side chain conformational flexibility. Each of the conformation library rotamers (e.g. tptm, pttm, etc.) can be included or excluded. The angle of voxel in which OSPREY continuously minimizes a rotamer (Gainza et al. 2012) can now be set, and each dihedral angle can be included or excluded from the continuous minimization. The customizations available to the protein designer with DexDesign unlocks the ability to explore new types of conformation spaces, such as D-space amino acids.

peptide improvement *via* K* redesign utilize the same algorithmic technique and therefore draw from the same pool of limited computational resources, we chose to prioritize D-peptide candidates with fewer clashes so we could allocate more computational resources to improving a D-peptide candidate *via* K* sequence redesign. Second, we observed instances where MASTER found identical D-peptide candidate sequences with nearly identical structures in multiple distinct PDB files, which we resolved by removing the duplicate results.

Using the above criteria, we selected 8 promising D-peptide candidates to use as starting points for OSPREY K* redesign. We call these selected candidates *D-peptide redesign (DPR) scaffolds*. The complete set of DPR scaffolds is described in Supplementary Table 2. To evaluate and improve upon the DPR scaffolds, we developed three design techniques.

Design techniques: minimum flexible set, inverse alanine scanning, and mutational scanning

DexDesign's *de novo* peptide design has one important distinction from protein redesign: the model of the starting protein structure used as input for K* redesign is a theoretical model, rather than one determined by experiment. As described in Computational tools and algorithms for designing D-peptides, OSPREY's algorithms have been successfully applied to a large and diverse set of biomedical applications. Yet in the most common uses of the K*-family of algorithms, *viz.*, those provably approximating K_a (K* (Lilien et al. 2005, Georgiev et al. 2008), BBK* (Ojewole et al. 2018), MARK* (Jou et al. 2020), and EWAK* (Lowegard et al. 2020)), the protein designer starts a redesign to achieve a specific redesign goal (e.g. improving or ablating binding of a protein to a ligand) from an experimentally determined structure of the protein: ligand complex. The mere existence of this experimental structure provides a solid foundation upon which further redesign builds, i.e. the given protein and ligand bind *at least in vitro*, and *at least to some degree*. Unfortunately, we do not have this luxury when designing *de novo* peptides to bind a given protein target. For example, while DexDesign uses the backbones of known L-peptide binders as input to the algorithm, the resulting DPR scaffolds are sufficiently different in sequence, side chain conformation, and chirality that the protein designer should assume that their DPR scaffold will not bind its protein target *in vitro* without further optimization *via* K*. To address this challenge—which is inherent to *de novo* design—we have developed design techniques that systematically evaluate the quality of DPR scaffolds and rigorously suggest mutations that are predicted to improve the D-peptide's binding affinity to its target PDZ domain. These three design techniques assume that the protein designer has a fixed amount of time and computational resources at their disposal. To that end, they are formulated to allow designers to rapidly evaluate their DPR scaffolds by restraining the conformation space to the minimal size necessary to computationally evaluate a hypothesis. Conformation space size grows exponentially with the number of flexible residues. Here, restricting the size of a conformation space is an effective technique to obtaining computational predictions quickly.

Design technique 1: identifying a minimum flexible set.

The K* algorithm in OSPREY predicts a provable approximation to K_a by calculating provable bounds on the partition function values of three molecular ensembles: the protein: ligand complex, the apo protein, and the apo ligand (Lilien et al. 2005, Georgiev et al. 2008). To generate these low-energy ensembles, the K* algorithm enumerates a stream of conformations in order of increasing energy, stopping only when it reaches a point when its enumerated conformations are sufficient to calculate a provably good ε-approximation to the partition function value, and thus the K* score. The set of conformations that K* enumerates is determined entirely by its *conformation space*, or the combinatorial set of all

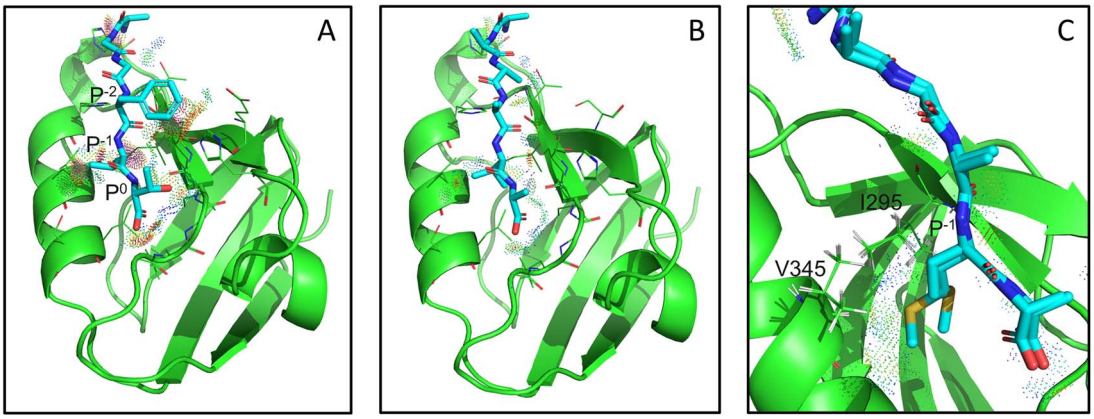

Figure 4. Illustration of the Minimum Flexible Set and Inverse Alanine Scanning design techniques applied to CALP-DPR5. (A) Choosing the Minimum Flexible Set. The CALP-DPR5 scaffold peptide (in cyan) is extracted from a crystal structure of the tobacco necrosis virus (PDB ID: 1c8n, residues C61-66, AGGFVT) (Oda et al. 2000). When aligned and superimposed over the query peptide kCAL01 and CALP (Holt et al. 2019) (in green) to create the DPR scaffold, sidechain and backbone clashes are present that the designer must address (red and pink MolProbity dots) (Word et al. 1999, Williams et al. 2018, Jou et al. 2024). The four peptide residues that clash with CALP are located at P^0 (Thr), P^{-1} (Val), P^{-2} (Phe), and P^{-5} (Ala). We specify the *Minimum Flexible Set,* or those residues that must be allowed to undergo continuous minimization (see Gainza et al. 2012) in all designs derived from CALP-DPR5, as the peptide residues located at P^0, P^{-1}, and P^{-2}. The peptide: CALP clash involving the alanine located at P^{-5} can be resolved by allowing OSPREY to translate and rotate the peptide during K* optimization, an option now available to the protein designer in the process of specifying a redesign's conformation space (related to Fig. 3). (B and C) The Inverse Alanine Scanning applied to CALP-DPR5. Complementing the Minimum Flexible Set technique, *Inverse Alanine Scanning* addresses peptide: target clashes by mutating all peptide residues modulo a single amino acid to alanine. In (B) and (C), we focus on position P^{-1} (Val) and use K* to mutate all other peptide residues to alanine, as well as to continuously minimize peptide: target sidechain conformations. (B) Inverse Alanine Scanning structural prediction with the source amino type, valine. As expected, the clashes present in CALP-DPR5 (A) vanish in the Inverse Alanine Scanning peptide (as indicated by the lack of red and pink Molprobity dots). Furthermore, K* has rotated P^{-1} (Val) to point towards the peptide's C-term, indicating that conformation is preferable. With this rotation, P^{-1} (Val) remains within CALP's hydrophobic pocket. (C) An OSPREY-predicted ensemble of the result of the Inverse Alanine Scan mutating position P^{-1} to methionine. CALP V345 and I295 form favorable van der Waals interactions (green and blue Molprobity dots) with multiple conformations of P^{-1} (Met), which is also reflected in the increase in K* score of P^{-1} (Met) compared to P^{-1} (Val). This result indicates that further K* binding affinity optimization can include methionine at P^{-1}, and that V345 and I295 should be allowed to flex continuously in CALP-DPR5 design candidates containing mutations at P^{-1}.

conformations that can be generated from given flexibility rules. Examples of such rules include the number of side chain rotamers an amino acid can explore, and the degree of continuous rotational flexibility permitted for a dihedral angle. The conformation space is in turn specified by the protein designer as an input to the K* algorithm (see Fig. 3). While specifying an appropriate conformation space has always been an important factor in K*'s ability to find mutations that accomplish a protein design's goal, specifying a sufficiently efficient, but still expressive, conformation space is an essential prerequisite of DPR scaffold redesign.

The MASTER search in Step 3 of the DexDesign algorithm (see Fig. 2) returns a set of L-peptides with low backbone RMSD to the D-peptide query. Due to the fact the MASTER search is by backbone-only RMSD, it is often the case that the L-peptide search results have side chains in sterically unfavorable positions that clash with the D-protein target. As discussed in Section 2.3, we prune DPR candidate peptides that cause many unfavorable clashes. On the other hand, we keep DPR candidate peptides with only a few small clashes because these clashes can typically be resolved with an appropriately specified conformation space. We call the set of residues that the protein designer must specify to be continuously flexible to resolve these clashes the *Minimum Flexible Set*.

The Minimum Flexible Set is different for each DPR scaffold, since the D-peptide in each DPR scaffold is unique. Given a fixed budget of computational resources, protein designers should prefer DPR scaffolds requiring smaller Minimum Flexible Sets, since such DPR scaffolds allow K* to expend more compute resources on searching for favorable mutations that increase binding to the target protein. See Fig. 4A for an example of specifying the Minimum Flexible Set for a CALP-DPR candidate.

Design technique 2: inverse alanine scanning.

Inverse Alanine Scanning is a technique complementary to the Minimum Flexible Set. Whereas the Minimum Flexible Set technique identifies a conformation space that resolves all clashes, Inverse Alanine Scanning allows the designer to investigate single residue mutations on the peptide that may increase binding to the target protein only when the target protein's nearby residues are provided sufficient flexibility in the conformation space. In contrast with Minimal Flexible Set, this technique resolves the problem of clashes between DPR scaffolds and the protein target by mutating all peptide residues, modulo the single residue under investigation, to alanine. We call this *Inverse* Alanine Scanning because this computational technique mutates all the peptide's residues *except* the residue of interest to alanine, the opposite of the canonical alanine scanning experiment. See Fig. 4B and C for a picture of this technique.

Notably, Inverse Alanine Scanning not only provides evidence as to which residues in the target protein must be flexible to accommodate certain favorable mutations, but it also provides evidence about which protein residues can safely remain rigid because they do not interact with peptide mutations in their vicinity. In the former case, a protein designer can visually inspect the OSPREY-generated structural ensemble of a favorable mutant to determine the set of residues which flexed and interacted with the mutated sidechains: these residues must remain flexible in the final conformation

space. Conversely, when Inverse Alanine Scanning identifies a favorable mutant that, upon visual inspection of the OSPREY-generated structural ensembles, has residues that *do not* flex, then these residues can safely be omitted from the final conformation space. This knowledge is valuable because it allows the designer to specify a smaller conformational space than they otherwise would in the next technique, K*-based Mutational Scanning.

Design technique 3: K-based mutational scanning.*

After learning which residues must flex from Minimum Flexible Set and obtaining hints as to which mutations might improve binding from Inverse Alanine Scanning, the K*-based Mutational Scanning technique (hereafter *Mutational Scan*) can be used. A *Mutational Scan* uses K* to systematically sample rotamers for all 20 amino acids at each DPR residue. We specified a K* design implementing a Mutational Scan for each residue in a DPR peptide and set the conformation space as the union of the set of flexible residues from the Minimum Flexible Set and the Inverse Alanine Scanning steps. We ran K* Mutational Scans on each of the DPR scaffolds. In many cases we observed mutations that notably improved the K* score. We then used the results of the Mutational Scans to inform the specification of additional K* designs that permitted multiple simultaneous peptide mutations in order to optimize peptide: target binding. We then further refined that set by removing sequences whose increase in K* score was driven primarily by peptide destabilization, as indicated by a large decrease in the unbound peptide's partition function value q_L (see Supplementary Information A: the K* algorithm). We then sorted the remaining favorable DPR sequences by their K* score. Finally, we analyzed the OSPREY-predicted structures of the top 3 sequences for each DPR scaffold. Our analysis is included below.

Results

DPR validation criteria

The aim of DexDesign is to predict novel D-peptides inhibitors. For this reason, we validated each of the DPR peptides across multiple criteria relevant to PDZ domain inhibitors. These criteria include:

1. The DPR's binding affinity. As an effective inhibitor must interact with its protein target in such a way that it disrupts the target protein's ability to bind its endogenous ligand, to validate the inhibitory potential of the D-peptides we compared their K* scores to the K* scores of each PDZ-domain's endogenous ligand, as well as some previous L-peptide inhibitors. We use the K* algorithm (Lilien et al. 2005, Georgiev et al. 2008) to optimize the D-peptide's sequence to increase binding affinity between the peptide and the endogenous ligand's binding site (the groove between $\alpha 2$ and $\beta 2$ in the PDZ domain). See Supplementary Information A for a definition of K* scores and how they are generated by the K* algorithm.
2. The DPR's ability to replicate biophysical facets common to PDZ domain binding. Due to their central role in regulating cellular trafficking and signaling pathways (Amacher et al. 2020), much research has been conducted to better understand and characterize PDZ domains (Harris and Lim 2001, Skelton et al. 2003, Jemth and Gianni 2007, Lee and Zheng 2010, Ivarsson 2012, Caillet-Saguy et al. 2015, Christensen et al. 2019, Amacher et al. 2020, Nardella et al. 2021, Tahti et al. 2023). This research has identified structural and biophysical elements that are commonly found facilitating canonical PDZ domain interactions (Christensen et al. 2019). One such element is the presence of a hydrogen bond network formed between the peptide's C-terminal carboxylate and the loop connecting $\beta 1$ and $\beta 2$, termed the *carboxylate binding loop (CBL)* (Doyle et al. 1996, Lee and Zheng 2010, Christensen et al. 2019). Another is the presence of β-strand-β-strand interactions between the peptide and $\beta 2$ (Harris and Lim 2001, Christensen et al. 2019). A third commonality is the presence of a hydrophobic pocket in the $\alpha 2$-$\beta 2$ groove, which canonically is filled by a hydrophobic residue at position P^0 in the peptide (Harris and Lim 2001), and in some peptides, by residue P^{-2} (Christensen et al. 2019).
3. We assessed the following biophysical facets in our validation of D-peptides: 1) the H-bond network formed by the D-peptide carboxylate and the CBL; 2) β-strand interactions between the D-peptide backbone and $\beta 2$, and; 3) the ability of the D-peptide to fill the hydrophobic pocket. These facets have proven sufficient for the systems presented in this manuscript, however, there may exist D-peptide inhibitors that initiate novel interactions not following this paradigm (see Discussion - Replication and Restitution for additional discussion on this point).
4. The presence of novel and favorable interactions in the DPR. Our validation includes a structural analysis of the OSPREY-predicted low-energy ensemble of the DPR bound to its target PDZ domain. Since DexDesign predicts *de novo* D-peptides, and since empirical structures of D-peptide inhibitors bound to PDZ domains are lacking, it is possible that a D-peptide could bind its target PDZ domain in a mode quite distinct from that of canonical L-peptides. For example, L-peptide residues at positions that point into the PDZ domain's binding groove may, in a D-peptide, point away (and *vice-versa*, see Supplementary Fig. 3 for an example), providing the possibility for some peptide residues to interact with parts of the PDZ domain in ways not formerly possible. To account for the possibility of novel modes of binding (and not, e.g. disregard D-peptides that fail to replicate all the criteria listed in 2), we analyzed the OSPREY-predicted low-energy molecular ensembles of the DPR bound to its target PDZ domain. In these analyses, we highlight the presence (or absence) of notable structural features capable of further validating the quality of the DPRs.

CALP

Using our crystal structure of kCAL01 bound to CALP (PDB ID 6ov7) (Holt et al. 2019), we used DexDesign to generate 5 DPR scaffolds: CALP-DPR[1–5] (see Supplementary Table 2 for further information about the DPR scaffolds). We then applied the design techniques and selection procedures from Applying DexDesign to CALP and MAST2 - Design Techniques to optimize the DPRs, thereby obtaining a final set of 15 D-peptide CALP inhibitors, CALP-PEP[1-15]. We assessed each of the CALP-PEPs using the quantitative and structural validation criteria described in Results - DPR validation criteria. We also compared the CALP-PEPs to CALP's endogenous

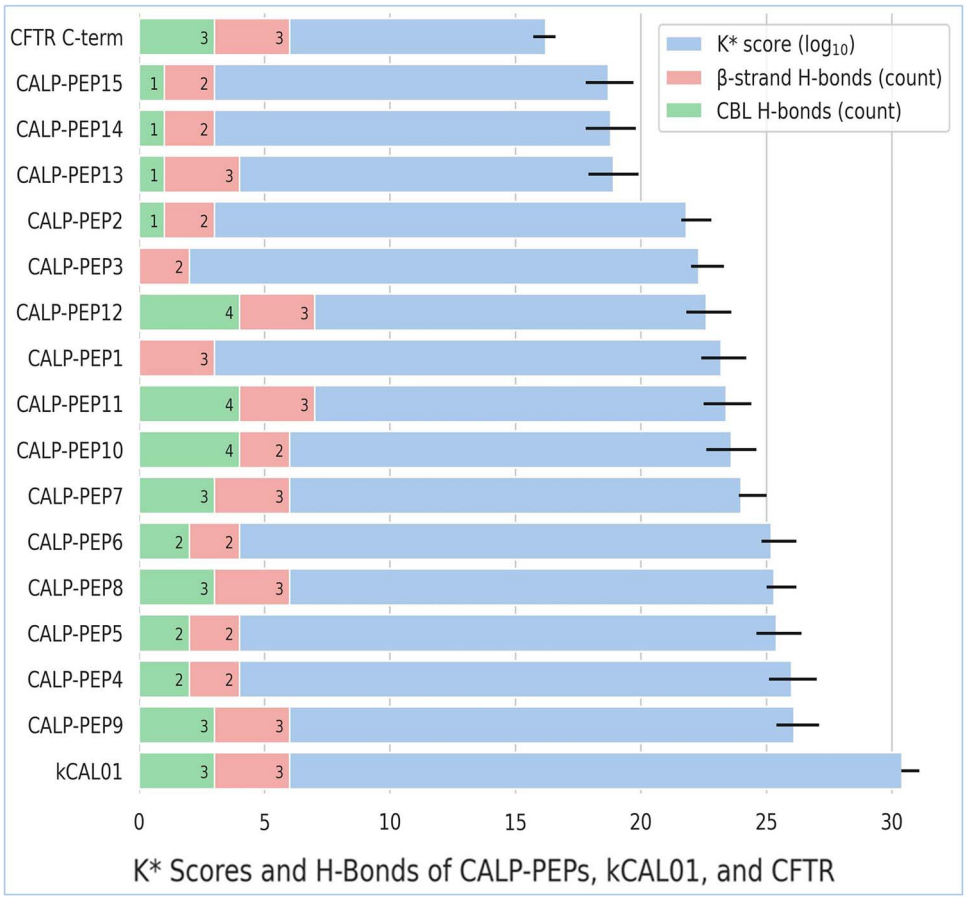

Figure 5. Quantitative and Structural Analysis of the CALP-PEPs, kCAL01, and the CFTR C-terminal SLiM. The OSPREY-predicted binding affinity ($\log_{10}$ K* score) of the tightest known peptide binder of CALP (kCAL01, PDB ID 6ov7) (Holt et al. 2019), the CALP-PEPs, and CALP's endogenous ligand (the CFTR C-terminal SLiM, PDB ID 2lob) (Piserchio et al. 2012) show (blue bars) that the CALP-PEPs are predicted to bind more tightly, with $\log_{10}$ K* scores ranging from 18.7 (CALP-PEP15) to 26.1 (CALP-PEP9), than the CFTR C-terminal SLiM ($\log_{10}$ K* score of 16.2). Conversely, the CALP-PEPs are predicted to bind CALP less tightly than the best known CALP peptide inhibitor, kCAL01 (Roberts et al. 2012, Holt et al. 2019), which OSPREY predicts to have a $\log_{10}$ K* score of 30.4. Since the primary objective of a competitive inhibitor is to outcompete an endogenous ligand in binding to the target protein, the K* scores, *viz.* provably accurate ε-approximations to K_a (see Supplementary Information A), of the 15 *de novo* D-peptide CALP-PEPs exceeding that of CFTR's C-terminal SLiM indicates that the CALP-PEPs meet their fundamental design objective. For example, CALP-PEP9 has a $\Delta\Delta G$ of -2.3 kcal/mol, improving K_D 46-fold, compared to the CFTR C-terminus (see Supplementary Information B). While not predicted to bind as tightly as kCAL01, D-peptides have therapeutic advantages over L-peptides, including improved metabolic stability (described in Background and introduction - Benefits of including D-amino acids in peptides), that can compensate for not reaching the binding affinity of the strongest CALP peptide inhibitor. The red bars show the number of β-strand H-bonds contributing to the common β2-sheet extension PDZ-binding motif. The green bars show the number of H-bonds between the peptide's C-terminal carboxylate and the CBL. The number of CBL and β-strand H-bonds varies across the CALP-PEPs, but the one predicted to bind tightest, CALP-PEP9, has 3 CBL and 3 β-strand H-bonds, the same number CFTR and kCAL01 have. The K* scores of the CALP-PEPs and empirical structures were determined using the K* algorithm (Lilien et al. 2005, Georgiev et al. 2008) in OSPREY. Supplementary Information A provides a definition of the K* algorithm and K* score. The error bars on the K* scores show the provable upper- and lower-bound of the K* approximation. The number and type of H-bonds between the peptides and CALP were determined using Pymol (Schrödinger 2015).

ligand (CFTR) and also to the most binding-efficient L-peptide CALP inhibitor, kCAL01, which our group reported in 2012 (Roberts et al. 2012) and solved a crystal structure of in 2019 (Holt et al. 2019). An overview of each of the CALP-PEP's K* scores, CBL H-bonds, and peptide: β2 backbone H-bonds is shown in Fig. 5. Notably, each of the CALP-PEPs is predicted to bind CALP tighter than the CFTR C-terminal SLiM, a critical prerequisite of an effective inhibitor. After normalization (see Supplementary Information B for information on the normalization procedure) and conversion to Gibbs free energy, the top 3 peptides, CALP-PEP9, CALP-PEP4, and CALP-PEP5, when compared to the CFTR C-terminus, have a ΔG of 2.3, 2.3, and 2.1 kcal/mol lower than the CFTR C-terminus (CALP-PEP9: $\Delta G = -6.9$ kcal/mol, CALP-PEP4: $\Delta G = -6.9$ kcal/mol, CALP-PEP5: $\Delta G = -6.7$ kcal/mol, CFTR C-terminus: $\Delta G = -4.6$ kcal/mol), improving K_D over the CFTR C-terminus by 46-, 44-, and 33-fold, respectively. Below, we provide the results and analyze the OSPREY-predicted structural ensembles.

K* redesign of the peptide sequence enabled each of the CALP-PEPs to achieve a tighter binding affinity to CALP when compared to the DPR scaffold from which it was generated. This is indicated by their positive $\log_{10}$ ΔK* score (see Supplementary Table 3). Notably, the CALP-PEPs ΔK* scores strongly correlate with their K* scores (Spearman correlation = 0.95). We postulate this strong correlation indicates that DexDesign's K* optimization is not merely alleviating clashes in the DPR scaffolds, but that it is also identifying peptide sequences forming novel side chain interactions that increase binding affinity. For example, CALP-PEP9's P^{-2}

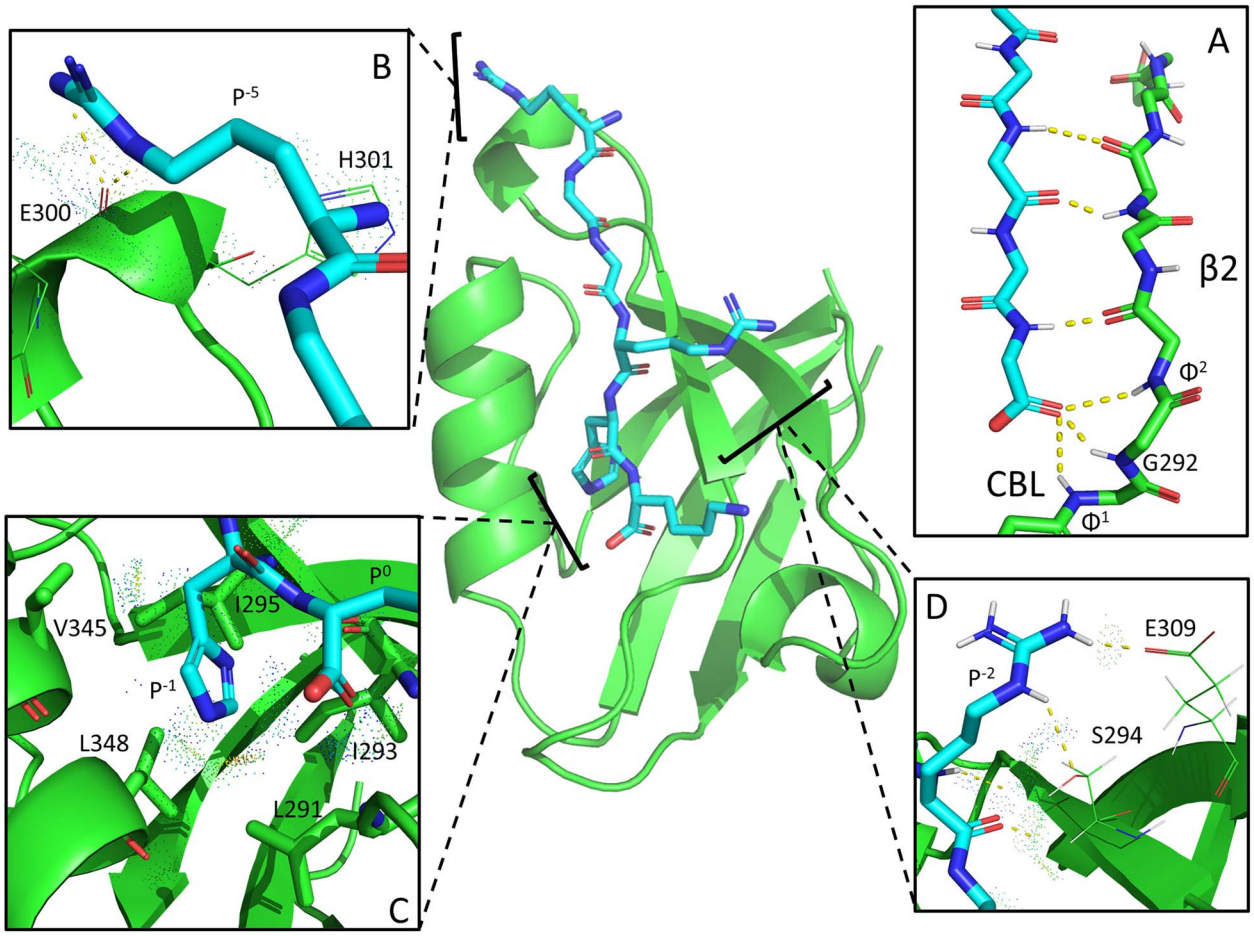

Figure 6. Structural analysis of OSPREY-generated ensemble of CALP-PEP9. Of the 15 CALP-PEPs, CALP-PEP9 (RGGRHK) is predicted to be the tightest binder to CALP with a $\log_{10}$ K* score of 26.1, which approaches the predicted affinity of the most binding-efficient L-peptide inhibitor kCAL01 (previously reported (Roberts et al. 2012) by our lab; with a $\log_{10}$ K* score of 30.4), and vastly exceeds the predicted binding affinity of the CFTR C-terminal SLiM ($\log_{10}$ K* score of 16.2). CALP-PEP9 is predicted to improve K_D by 46-fold over the CFTR C-term, with a normalized K_D of 8.9 μM versus 420 μM (Amacher et al. 2013) for the C-terminal SLiM (see Supplementary Information B). (A) CALP-PEP9's P^0 carboxylate forms favorable H-bonds with the carboxylate binding loop (CBL: G290-I293, $G\Phi^1G\Phi^2$) and strand β2, mimicking canonical PDZ binding interactions (Christensen et al. 2019) of L-peptides. (B) The amino acid at position P^{-5} in CALP-PEP9 is arginine. P^{-5}'s amino group and sidechain make favorable van der Waals contacts, indicated by blue and green MolProbity dots (Word et al. 1999, Williams et al. 2018, Jou et al. 2024), with CALP's E300 and H301. Its guanidine sidechain also forms H-bonds with E300's carboxyl group. (C) In canonical L-peptide's, a PDZ-domain's hydrophobic pocket is filled by a hydrophobic amino acid at position P^0 (Christensen et al. 2019). In contrast, all of the CALP DPRs fill the pocket with the amino acid at position P^{-1} (see Supplementary Fig. 3). CALP's hydrophobic pocket, defined as the groove between α2 and β2 and involving V345, I295, I293, L291, and L348, is filled by a histidine in position P^{-1}. (D) CALP-PEP9's P^{-2} is arginine, which is predicted to make favorable van der Waals contacts and form H-bonds with β2's S294 and the sidechain of β3's E309. The K* scores and additional structural validation of all the CALP-PEPs can be found in Supplementary Table 3.

arginine reaches across β2 and makes favorable contacts with β3's Glu309 (see Fig. 6D). The magnitude of the predicted improvement in binding ranges from $\Delta\Delta G = -2.0$ kcal/mol for CALP-PEP15 to -3.5 kcal/mol for CALP-PEP9. Next, we validated the CALP-PEPs based on their ability to replicate canonical PDZ domain binding motifs.

To quantify validation criterion (1) (the presence of an H-bond network, see Results - DPR validation criteria bullet point 2), we counted the number of H-bonds formed between the peptide's C-terminal carboxylate and the CBL, and to quantify (2) (the presence of peptide: β2 backbone interactions) we counted the number of H-bonds between the peptide and β2 backbones. We used visual inspection with MolProbity Probe Dots (Word et al. 1999, Williams et al. 2018) in our lab's Protein Design Plugin (Jou et al. 2024) to perform a binary classification for (3) (the ability of the D-peptide to fill the hydrophobic pocket), more specifically, we classified the hydrophobic pocket as *filled* if a D-peptide's sidechain contacted the residues within the pocket. As a point of reference, kCAL01 and the CFTR C-terminal SLiM each have 3 H-bonds with the CBL, 3 H-bonds between the peptide and β2 backbone, and fill CALP's hydrophobic pocket their P^0 amino acid (valine for kCAL01, leucine for CFTR C-terminal SLiM).

9 of the 15 CALP-PEPs (CALP-PEP[4-12]) had 2 or more H-bonds between their C-terminal carboxylate with the CBL. CALP-PEPs derived from the CALP-DPR1 and CALP-DPR5 scaffolds had 1 (or in the case of CALP-PEP1 and CALP-PEP3, 0) H-bonds. Encouragingly, all the CALP-PEPs formed at least 2 backbone H-bonds with CALP's β2 strand (see Fig. 5 and Supplementary Table 3). CALP-PEP9, which we predict to be the tightest binder to CALP, forms 3 C-terminal carboxylate H-bonds with the CBL and 3 backbone H-bonds with CALP's β2 strand, matching the numbers formed by both

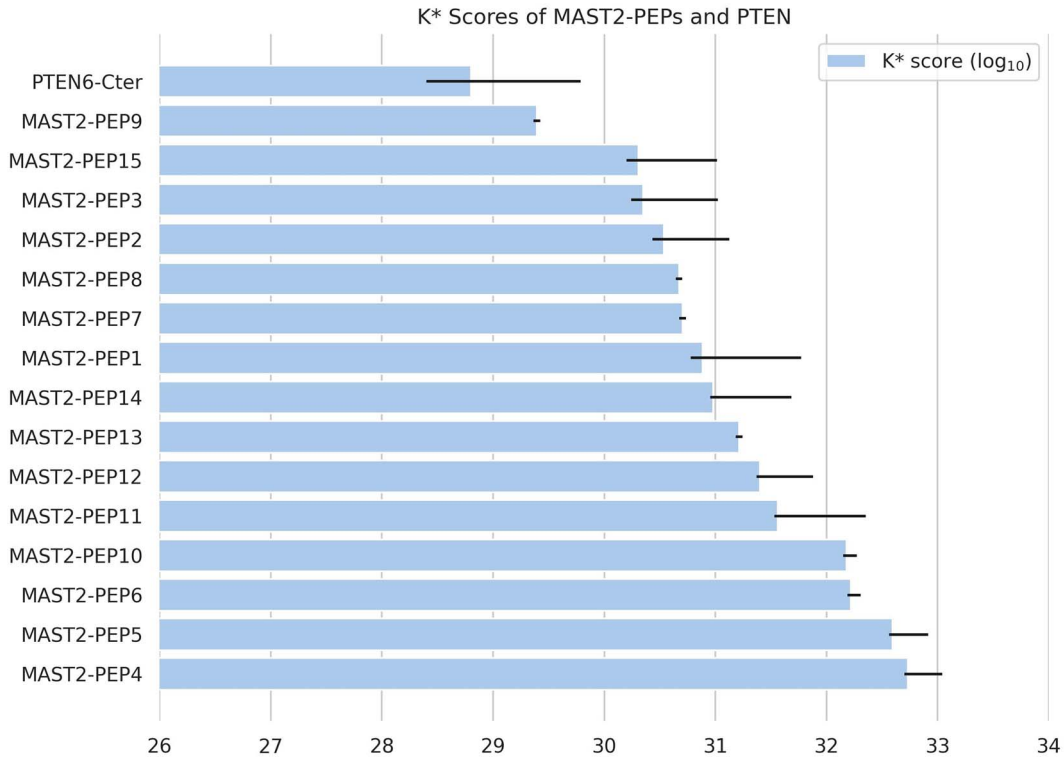

Figure 7. The DexDesign-generated D-peptides are predicted to bind to MAST2 tighter than PTEN. Blue bars show the OSPREY-predicted binding affinity of the MAST-PEPs and PTEN6-Cter (hereafter denoted PTEN6, the 6 C-terminal residues of MAST2's endogenous ligand PTEN). We used the DexDesign algorithm (described in Methods - Algorithm and computational protocol) and novel design techniques (described in Applying DexDesign to CALP and MAST2 - Design techniques) to generate 15 *de novo* D-peptides predicted by the K* algorithm (Lilien et al. 2005, Gainza et al. 2012) to bind MAST2 tighter than PTEN6. Notably, PTEN binds MAST2 209-fold tighter than CFTR binds CALP ($K_D = 1.9 \pm 0.05$ μM *vs.* 420 ± 80 μM) (Terrien et al. 2012, Amacher et al. 2013), and binds MAST2 as tightly as the strongest known inhibitor of CALP, kCAL01 ($K_D = 2.3 \pm 0.2$ μM) (Roberts et al. 2012), indicating a more challenging target of inhibition. OSPREY predicts PTEN6 to bind MAST2 with a $\log_{10}$ K* score (a provably accurate ε-approximations to K_a, see Supplementary Information A) of 28.8. Despite the more difficult target, all the MAST2-PEPs have higher $\log_{10}$ K* scores than PTEN6, ranging from 29.4 for MAST2-PEP9 to 32.7 for MAST2-PEP4, meaning the MAST-PEPs are predicted to outcompete PTEN6 and inhibit PTEN6: MAST2 binding. The best predicted inhibitor, MAST2-PEP4, has a $\Delta\Delta G$ of -1.1 kcal/mol, improving K_D 5-fold compared to the PTEN6 (see Supplementary Information B). The K* scores of the MAST2-PEPs and empirical structures were determined using the K* algorithm (Lilien et al. 2005, Georgiev et al. 2008) in OSPREY (Hallen et al. 2018). Supplementary Information A provides a definition of the K* algorithm and K* score. The error bars on the K* scores show the provable upper- and lower-bound of the K* approximation.

the CFTR C-terminal SLiM and kCAL01 (see Fig. 6). CALP-PEP9 contains arginines at position P^{-2} and P^{-5}, both of which are predicted to make favorable van der Waals contacts with CALP (see Fig. 6B and D). In addition, CALP-PEP's P^{-2} and P^{-5} guanidino groups are predicted to form H-bonds with CALP's S294 and E309, and E300 and H301, respectively (Fig. 6B and D). While it does not appear that the quantity of CBL and backbone H-bonds drives the predicted strength of binding of CALP-PEPs (see Fig. 5), the CBL does play an important role in determining peptide specificity (Amacher et al. 2020), therefore we regard evaluation of D-peptide: PDZ domain H-bonds as necessary components of a larger ensemble of criteria.

In contrast, whether a CALP-PEP fills CALP's hydrophobic pocket is important to designing a tight binder. All the CALP-PEPs saw an increase in their K* scores when they mutated P^{-1} to an amino acid capable of filling the pocket (see Supplementary Table 3). 11 of the 15 CALP-PEPs mutate P^{-1} to histidine, 3 of the 15 to phenylalanine, and CALP-PEP15 is unique with methionine. For example, the mutation to histidine at P^{-1} in CALP-PEP9 fills and favorably interacts with multiple residues in the hydrophobic binding pocket (see Fig. 6C).

MAST2

Using an NMR structure of PTEN bound to MAST2 (PDB ID 2kyl) (Terrien et al. 2010), we used DexDesign to generate 3 DPR scaffolds: MAST2-DPR[1–3] (see Supplementary Table 2). From these 3 DPR scaffolds, we used the design techniques described in Applying DexDesign to CALP and MAST2 - Design techniques to generate 15 peptides, MAST2-PEP[1-15]. Figure 7 shows an overview of the MAST2-PEPs K* scores and how they compare to MAST2's endogenous ligand PTEN. Additional structural information about the MAST2-PEPs, such as which residue fills MAST2's hydrophobic cavity, is in Supplementary Table 4.

PTEN binds MAST2 209-fold tighter than CFTR binds CALP ($K_D = 1.9 \pm 0.05$ μM *vs.* 420 ± 80 μM) (Terrien et al. 2012, Amacher et al. 2013) and binds MAST2 as tightly as the strongest known inhibitor of CALP, kCAL01 ($K_D = 2.3 \pm 0.2$ μM) (Roberts et al. 2012). In other words, to design competitive inhibitors of the MAST2: PTEN interaction requires us to design D-peptide inhibitors with a better affinity than the tightest known L-peptide inhibitor of CALP. Despite the challenge inherent in disrupting the MAST2: PTEN interaction, all the MAST2-PEPs are predicted to bind MAST2 with affinities surpassing PTEN (see Fig. 7).

The $\log_{10}$ K* scores of the MAST-PEPs range from a low of 29.4 for MAST2-PEP9 to a high of 32.7 for MAST2-PEP4. MAST2-PEP4 is the best DexDesign-generated inhibitor and is predicted to bind MAST2 with a normalized Gibbs free energy ΔG of −8.8 kcal/mol, a −1.1 kcal/mol improvement over MAST2: PTEN, resulting in a 5-fold improvement in K_D. In some cases a 5-fold improvement might be considered small, but we have previously shown (Roberts et al. 2012, Rudicell et al. 2014) that differences of this magnitude can have profound effects on biological activity. For example, kCAL01 binds only 6-fold tighter than previous competing peptides, such as iCAL35, that were discovered *via* high-throughput SPOT arrays (Roberts et al. 2012). However, in *ex vivo* assays (see PDZ domains - CFTR-associated ligand) iCAL35 had non-significant biological activity (Roberts et al. 2012), whereas the 6-fold tighter binding kCAL01 had significant biological activity. Since the SLiMs modulate a delicate network of competing affinities and specificities (Lee and Zheng 2010, Amacher et al. 2013, Amacher et al. 2020), 5–7× improvements in affinity (such as that achieved by CALP-PEP4) can make the difference between failure and true biological activity.

The MAST2-PEPs replicate some of the canonical L-peptide PDZ-binding motifs, such as the residue in position P^0 filling the hydrophobic pocket between the PDZ domain's $\alpha 2$ helix and $\beta 2$ strand. 9 out of 15 MAST2-PEPs have residue P^0 filling the hydrophobic pocket (see Supplementary Table 4). This contrasts with the CALP-PEPs, where in all cases the residue at position P^{-1} filled the hydrophobic pocket. In the best predicted inhibitor, MAST2-PEP4, the P^0 leucine fills MAST2's hydrophobic cavity formed by Tyr17, Phe19, Val77, Ile79, and Leu81 (see Fig. 8C and D). In contrast to PTEN's P^0 valine, MAST2-PEP4's P^0 leucine forms favorable interactions with all 5 of the cavity's hydrophobic residues. In addition, a rotation of MAST2-PEP4's C-terminal carboxylate alleviates a steric clash with the carboxylate binding loop present in the MAST2: PTEN complex.

The MAST2-PEPs also exploit novel geometric features of D-peptides not available to their L-counterparts. For example, MAST2-PEP4's P^{-3} glutamate makes favorable van der Waals contacts with MAST2's His73 imidazole side chain (see Fig. 8A). PTEN does not make the analogous interaction, and instead PTEN's P^{-3} isoleucine is oriented towards MAST2's $\beta 2$ strand, and the residue nearest to His73 is a glutamine at P^{-4}, whose amide fails to make van der Waals contacts with His73 (see Fig. 8B). The creation of novel favorable interactions with MAST2 is common in the designed MAST2-PEPs and compensates for the loss of some of the canonical PDZ-domain binding motifs. We discuss the trade-offs between replicating canonical interactions and finding new modes of binding available to D-peptides below.

Discussion

Replication and restitution: a framework for evaluating *de novo* peptides

Though the CALP-PEPs and the MAST2-PEPs both target PDZ domains, the interactions they make with their respective targets can be generally categorized into one of two kinds: replication or restitution. *Replication* means to replicate interactions previously observed in L-peptide inhibitors, for example, in the case of the PDZ domains, the $\beta 2$ strand extension, residue P^0 filling the hydrophobic pocket, and the H-bond network a peptide's terminal carboxylate makes with the CBL. *Restitution*, on the other hand, refers to the process of compensating for typical L-peptide binding motifs by making novel interactions that are now possible to explore due to the change from L- to D-chirality of the peptide. Whereas in some cases the generated peptides are predicted to have improved binding through replicating the canonical PDZ interactions, in other cases, due to the special geometric properties of D-peptides, we instead observe an increase in binding (restitution) due to novel interactions that we observe in the structures that were not available to L-peptides. This suggest the intriguing possibility that some peptides may be stabilized by replicating native-like interactions from L-peptides, whereas others might be stabilized by forming novel interactions, available only to the ligands with D-space configuration of peptides.

In the *de novo* peptides we generated, both the CALP-PEPs and the MAST2-PEPs contained elements both replicating and restituting the binding interactions formed by the endogenous L-peptides from which their backbone conformations are derived (see Fig. 1). In general, the CALP-PEPs relied more on a strategy of replication to improve binding affinity, whereas the MAST2-PEPs exploited a strategy of restitution. Take for example the SLiM's canonical C-terminal carboxylate H-bond network formed with the CBL. Whereas 13 of the 15 CALP-PEPs replicated (to varying degrees, and sometime even exceeding the number of) H-bonds formed with the CBL (see Fig. 5), the MAST2-PEPs' terminal carboxylate tended to have few, if any, H-bonds with the CBL (see, e.g. Fig. 8C). We postulate that the reason for the MAST-PEPs' use of restitution instead of replication for the CBL H-bond network is that the structure of the PTEN: MAST2 complex (PDB ID 2kyl) (Terrien et al. 2010) indicates the existence of steric clashes between the C-terminal carboxylate and the CBL (see Fig. 8D). When using MolProbity (Williams et al. 2018) to evaluate the lowest-energy model of the empirical NMR structure, it indicates there is a bad clash (van der Waal radii overlap of 0.518 Å) between PTEN P^0 valine's OXT atom and the HA atom from Lys16 in MAST2's CBL. OSPREY's energy function (Georgiev and Donald 2009, Gainza et al. 2012, Hallen et al. 2018, Martin and Donald 2024), which uses continuous sidechain minimization in addition to translation and rotation of the peptide to minimize the energy of each conformation evaluated by OSPREY's iMinDEE/K* algorithm (Georgiev et al. 2008, Gainza et al. 2012), pushes and rotates the C-terminal carboxylate away from the CBL to alleviate the steric clash (see Fig. 8C), with the trade-off being the loss of replication of some of canonical PDZ CBL H-bond interactions.

The CALP-PEPs and MAST2-PEPs also exploit restitution to form novel favorable interactions with their target proteins. One type of restitution is the creation of favorable new side chain interactions between the peptides and their targets. For example, CALP-PEP9's P^{-2} arginine forms two new H-bonds with CALP $\beta 2$ strand's S294 and the sidechain of $\beta 3$'s E309 (see Fig. 6D) that are absent in the CALP: kCAL01 structure (PDB ID 6ov7) (Holt et al. 2019). Interestingly, CFTR's P^{-1} arginine forms an analogous H-bond with E59 CALP: CFTR structure (PDB ID 2lob) (Piserchio et al. 2012), providing evidence that this interaction, as restituted in CALP-PEP9 (based on the kCAL01 structure lacking this interaction), is both plausible and favorable *in vitro*. The

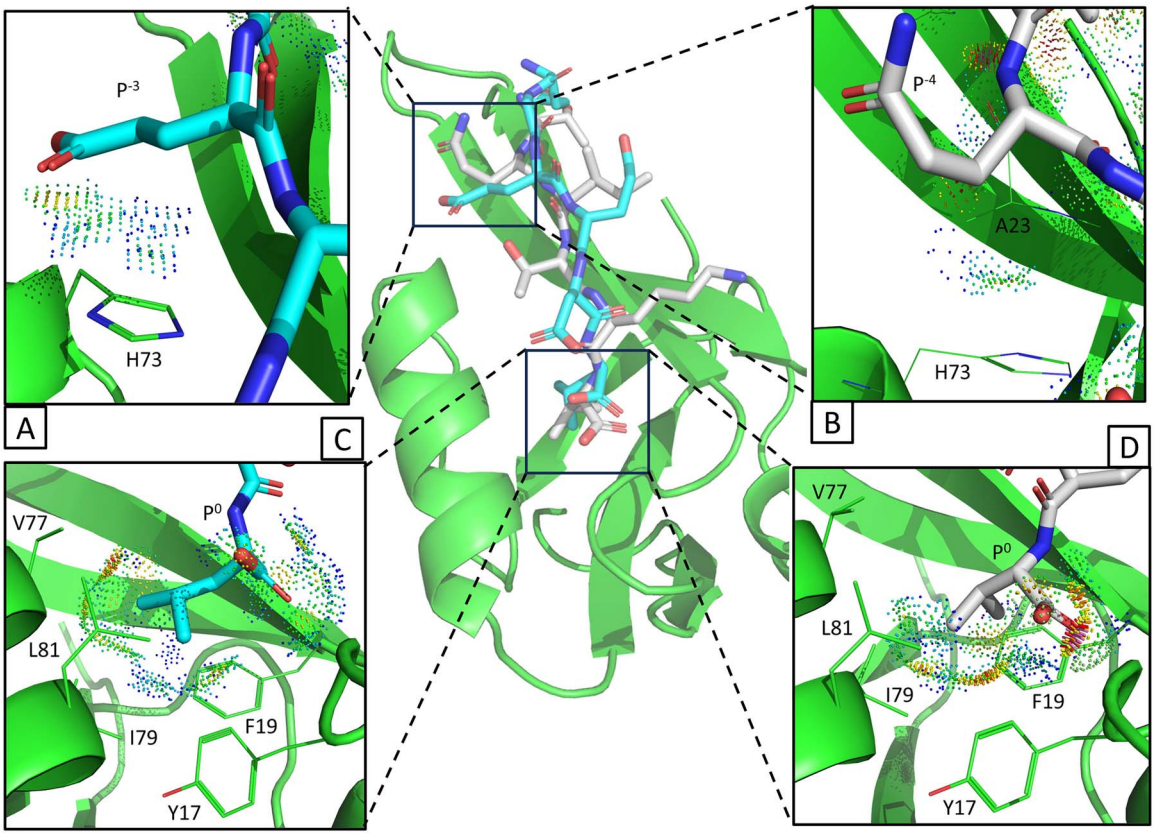

Figure 8. MAST2-PEP4 creates novel favorable interactions with MAST2 not found in PTEN. MAST2-PEP4 (cyan sticks) is the DexDesign-generated *de novo* D-peptide predicted to bind MAST2 (green cartoon and lines) with the tightest affinity. The K* algorithm (Lilien et al. 2005, Gainza et al. 2012) in OSPREY (Hallen et al. 2018) predicts MAST2-PEP4 to bind MAST2 with a $\log_{10}$ K* score (a provably accurate ε-approximations to K_a, see Supplementary Information A) of 32.7, compared to 28.8 for PTEN6 (gray sticks, the 6 C-terminal residues of MAST2's endogenous ligand PTEN). After normalization (see Supplementary Information B), the Gibbs free energy change ΔG of the MAST2: MAST2-PEP4 complex is −8.8 kcal/mol, a −1.1 kcal/mol improvement over MAST2: PTEN6, resulting in a 5-fold improvement in K_D. Center: the lowest-energy conformation from the OSPREY-predicted conformational ensemble of MAST2 (green) bound to MAST2-PTEN6 (cyan) and the lowest-energy model of PTEN6 (grey) from an empirical solution NMR ensemble of the MAST2: PTEN complex (PDB ID 2kyl) (Terrien et al. 2010). In comparison to the binding modes of, e.g. the CALP-PEPs to CALP (see Fig. 6) which largely *recover* canonical PDZ-domain binding interactions, MAST2-PEP4 *restitutes* binding to MAST2 by exploiting novel geometric features of D-peptides not available to their L-counterparts (see Discussion). A: MAST2-PEP4's P^{-3} glutamate makes favorable van der Waals contacts with MAST2's His73 imidazole side chain, as indicated by predominantly green and blue MolProbity dots (Word et al. 1999, Williams et al. 2018, Jou et al. 2024). These favorable contacts are absent in the MAST2: PTEN6 complex. (B) In contrast to (A), PTEN6 cannot make some favorable interactions available to our D-peptides. For example, in MAST2: PTEN6, PTEN6's P^{-3} isoleucine is oriented towards MAST2's $\beta 2$ strand (not shown), and the residue nearest to His73 is P^{-4} glutamine, whose amide fails to make van der Waals contacts with His73. C: MAST2-PEP4's P^0 leucine fills MAST2's hydrophobic cavity (Terrien et al. 2012) formed by Tyr17, Phe19, Val77, Ile79, and Leu81. In contrast to PTEN6's P^0 valine (D), MAST2-PEP4's P^0 leucine forms favorable interactions, as indicated by the green and blue MolProbity dots, with all 5 of the cavity's hydrophobic residues. In addition, a rotation of the C-terminal carboxylate alleviates a steric clash (indicated by the red and pink MolProbity dots in (D) with the carboxylate binding loop present in the MAST2: PTEN6 complex.

MAST2-PEPs likewise restitute novel favorable interactions. For example, MAST2-PEP4's P^{-3} glutamate makes favorable van der Waals contacts with MAST2's His73 imidazole side chain (see Fig. 8A). In contrast, PTEN cannot make some favorable interactions available to MAST2-PEP4. In the MAST2: PTEN structure, PTEN's P^{-3} isoleucine is oriented towards MAST2's $\beta 2$ strand, and the residue nearest to His73 is P^{-4} is glutamine, whose amide fails to make van der Waals contacts with His73 (see Fig. 8B). In the future, we believe that designed D-peptide libraries of binders and inhibitors can be characterized as falling on a spectrum ranging from replication (1) to restitution (−1) and can be visualized as a per-residue replication-restitution score ranging from 1 to −1. In this way, the functional contributions to binding of *de novo* peptides could be mapped into a vector space which can be visualized or exploited as novel features for machine learning design approaches.

Validation of DexDesign scaffold discovery and redesign

To assess DexDesign, we performed a computational experiment to measure the ability of DexDesign to design a *de novo* D-peptide similar to a D-peptide found in an empirical D-peptide: L-protein complex. In our experiment, we began with the crystal structure of a known D-peptide in complex with an L-protein and applied a global reflection to the D-peptide: L-protein complex. Then, we employed MASTER to search the L-database using the resulting, now flipped, L-peptide as the query. The returned L-structures were aligned with the reflected complex and reflected once again to produce D-peptide: L-protein scaffolds ordered by increasing backbone alignment RMSD. The first, and therefore lowest, RMSD backbone alignment was then selected for redesign using OSPREY. We then measured the similarities of the redesigned D-peptide to the empirical D-peptide. While

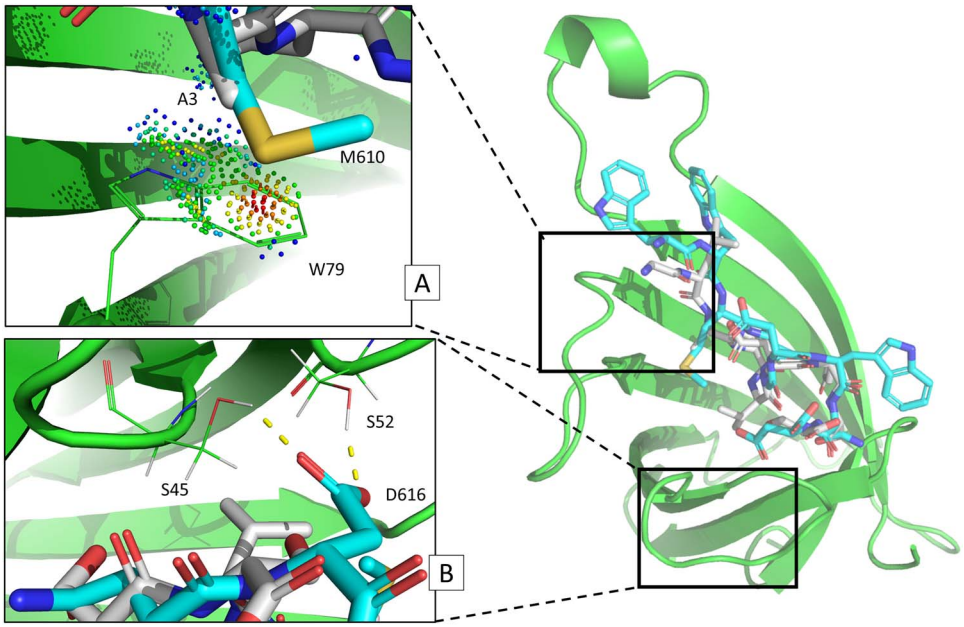

Figure 9. Geometric, chemical, and physical properties of DPRV that drive binding to streptavidin. (A) Similar to GyGlanvdessG Ala3, DRPV Met610 displays favorable hydrophobic and van der Waals contacts with streptavidin Trp79. Streptavidin (green cartoon and lines) displays hydrophobic contributions through inward-facing tryptophan residues of the β barrel, which have been reported as important for ligand binding (Hyre et al. 2006). Favorable van der Waals interactions are shown as green and blue dots between DRPV (cyan) Met610 and streptavidin Trp79. GyGlanvdessG (grey) Ala3 also shares similar hydrophobic and van der Waals contacts with streptavidin Trp79. (B) Differences in C-terminal orientation and interactions between D-peptide GyGlanvdessG: streptavidin and DPRV: streptavidin. Unlike GyGlanvdessG, DRPV's aspartic acid C-terminus makes hydrogen bonds with both streptavidin residues Ser45 and Ser52. However, GyGlanvdessG's C-terminal Ser9 fails to form a hydrogen bond with streptavidin Ser52. Residues number Ser45-Ser52 on streptavidin are known to be important for establishing binding of biotin (Freitag et al. 1997), so these contacts are likely evidence of high-affinity binding of DPRV.

DexDesign is intended for construction of novel *de-novo* D-peptides, the capability to generate a DPR similar to the D-form empirical structure should validate our redesign protocol. Finally, we incorporated significantly more backbone sampling and remodeling that improved both binding affinity and native sequence recovery significantly (see Improved predictions using additional backbone sampling and remodeling below).

We selected a crystal structure of the D-amino acid containing peptide GyGlanvdessG in complex with streptavidin (PDB ID 5n8j) (Lyamichev et al. 2017). Streptavidin is a homotetrameric protein that binds the vitamin biotin with high affinity (Freitag et al. 1997), and is therefore commonly used in Western blotting and immunoassays (Brower et al. 1985). A monomer of streptavidin forms a β barrel, with ligands oriented towards the interior of the barrel. Similar to CBL interactions with CALP and MAST2, streptavidin forms favorable hydrogens bonds with ligands *via* a flexible binding loop (Freitag et al. 1999). Analogously, streptavidin exhibits hydrophobic contributions through inward-facing tryptophan residues of the β barrel (Hyre et al. 2006). Therefore, a high-affinity ligand should establish hydrogen bonds with the binding loop while orienting hydrophobic residues towards β barrel tryptophans. We selected this system due to its comparable D-peptide size and similar chemistry to PDZ domains.

We sourced the lowest backbone RMSD (0.48 Å) of inverted D-amino acid GyGlanvdessG from chain A residues 608 to 616 of ST0929 (PDB ID 3hje) (Cielo et al. 2010), a glycol transferase. After application of Minimum Flexible Set, Inverse Alanine Scanning, and K*-based Mutational Scan to this scaffold (see Applying DexDesign to CALP and MAST2 - Design techniques), we determined the optimal binder, hereafter denoted as DPRV, to have a $\log_{10}$ K* score of 32.8 with the sequence WWMIGDWND. This differed slightly from GyGlanvdessG (GLANVDESS), which has a $\log_{10}$ K* score of 32.2. The sequence similarity between the two peptides is 21.43% (see Improved predictions using additional backbone sampling and remodeling for computational experiments that show how increased backbone sampling and remodeling improved native sequence recovery to 44% and predicted binding affinity to 39.54), a degree of native sequence recovery comparable to reported recovery in popular protein design programs such as Rosetta for L-proteins (Zhou et al. 2020b). This is especially true for NMR structures, such as we used in our MAST2 study (see Results - MAST2). With DPRV, we report that DexDesign generates a D-ligand with chemistry unique to the DPR scaffold.

While DexDesign exhibits comparable performance to state-of-the-art methods (Zhou et al. 2020b), native sequence recovery on a short (9 residue) peptide may be a poor indicator of ligand binding. For example, a 40% sequence recovery equates to 3.6 residues for our redesigned peptide. This is a small number of residues, and likely fails to capture the geometric and chemical features that drive high affinity. To investigate the similarities of GyGlanvdessG and DPRV, we also report the backbone alignment RMSD of DPRV to GyGlanvdessG: streptavidin (0.48 Å), and the geometric, chemical, and biophysical properties of our designed peptide that enable binding (see Fig. 9). Finally, we report the $\log_{10}$ K* scores computed over molecular ensembles as validation of binding competency (above and see Supplementary Table 5).

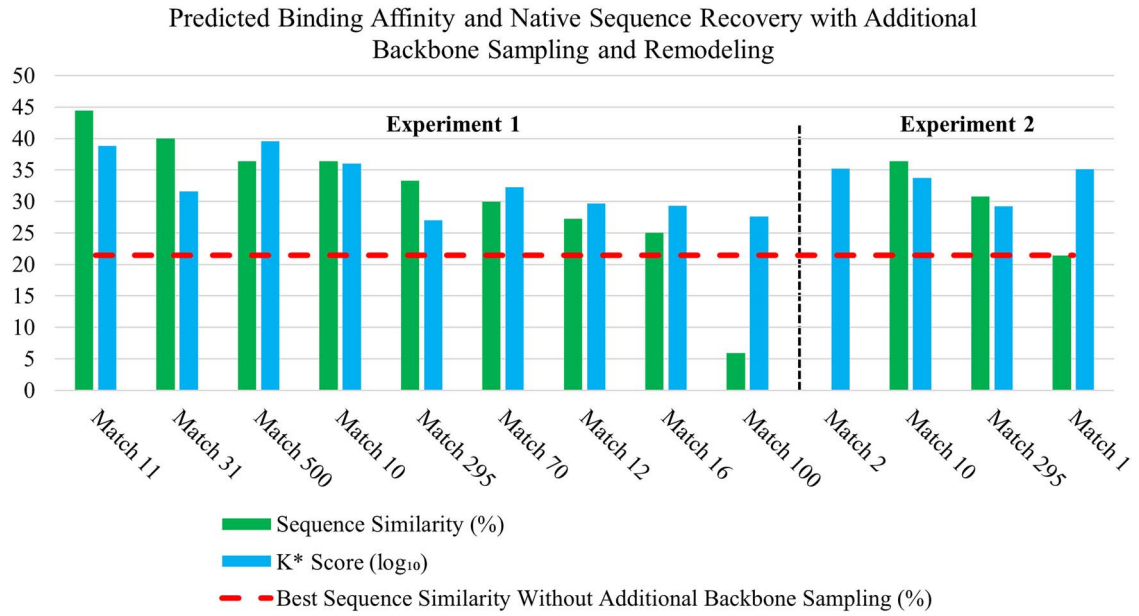

Figure 10. Quantitative analysis of D-peptides with additional backbone sampling and remodeling. Blue bars note the predicted binding affinity (K* score (log_{10}), see Supplementary Information A) while green bars note the native and DexDesign-predicted sequences' similarity. The numbers on the *y*-axis denote both K* score (log_{10}) and sequence similarity (%). A horizontal red dashed line shows comparison of native sequence recovery to DRPV (Discussion - Validation of DexDesign scaffold discovery and redesign). Experiment 1 and 2 data are separated by the vertical dashed black line. See Supplementary Information Tables 6 and 7 for the full dataset. As reported in Discussion - Improved predictions using additional backbone sampling and remodeling, two *in silico* experiments were performed to evaluate DexDesign native sequence recovery and predicted binding affinity using additional backbone sampling and remodeling. These experiments improved native sequence recovery results for DPRV and GyGlanvdessG over the model using a single backbone conformation. Sampling of additional backbone conformations (Experiment 1) resulted in the greatest increase in native sequence recovery from 21.43% (DPRV) to 44.44% (Match 11). This computational experiment also resulted in the greatest improvement in predicted binding affinity, with an improvement in K* score (log_{10}) from 32.8 (DPRV) to 39.54 (Match 500). Experiment 2 indicates the benefit of incorporating diverse scaffolds from homology modeling into DexDesign for native sequence recovery, though Match 1 suggests a potential trade-off between sequence recovery and predicted binding affinity. The median sequence similarity and K* score (log_{10}) for Experiment 1 were 31.7 and 31.9, respectively. The median sequence similarity and K* score (log_{10}) for Experiment 2 were 30.8 and 33.7, respectively. Overall, sampling of additional backbone substructures (Experiment 1) and remodeling of ligand backbone structure (Experiment 2) resulted in an almost universal increase in native sequence recovery and predicted binding affinity compared to the single backbone conformation DPRV native sequence recovery experiments.

We also performed a control experiment wherein we mutated the ST0929 scaffold sequence (RYEEGLFNN) directly to the sequence of the D-peptide GyGlanvdessG, without using any OSPREY-based techniques such as K*-based Mutational Scan and Inverse Alanine Scanning. The purpose of this experiment was to investigate the predicted binding of the GyGlanvdessG sequence on the ST0929 scaffold backbone. This control experiment produced a log_{10} K* score of only 26.6, a difference of −5.7 from GyGlanvdessG. Interestingly, a 100% wildtype sequence recovery mutant yields lower predicted binding despite the selection of a DPR scaffold with the lowest backbone RMSD. Therefore, we conclude that the GyGlanvdessG sequence was not recovered by the full DexDesign protocol because these novel OSPREY-based design techniques would not permit optimal binding on the lowest backbone RMSD scaffold. Instead, the DexDesign techniques (as outlined in Applying DexDesign to CALP and MAST2 - Design techniques) resulted in a novel D-peptide. These results highlight the sensitivity of the peptide design to the starting scaffold; even similar-appearing scaffolds have different degrees of designability due to geometric differences between backbones. Overall, our experiment highlights the utility of DexDesign for generation of novel peptides, as opposed to sequence recovery of known binders.

The difference between amino acid composition of the D-amino acid containing peptide GyGlanvdessG and the redesigned peptide is likely due to subtle differences in scaffold geometry. As shown in Fig. 9 and Supplementary Fig. 4, the backbones of D-amino acid GyGlanvdessG, DPRV, and the ST0929-sourced peptide scaffold mutated to the endogenous ligand (control) vary at residues important for establishing hydrogen bonds. For example, GyGlanvdessG's Glu7 residue makes hydrogen bonds with residues Asn23 and Ser27 of streptavidin (Supplementary Fig. 4A). DPRV's Trp7, which is shifted 1.9 Å away from streptavidin Ser27 in comparison to GyGlanvdessG, does not form either of these hydrogen bonds. However, DRPV's Asp9 facilitates hydrogen bond formation with residues Ser45 and Ser52 on streptavidin (Fig. 9B). These residues belong to the flexible binding loop, where favorable contacts are crucial for high-affinity binding (Freitag et al. 1997). The hydrogen bond formed with flexible loop residue Ser52 is unique to DRPV and is not present in the ST0929 control experiment or GyGlanvdessG. Therefore, a peptide that replicates some characteristics of a known ligand, while restituting novel interactions, may be a much more competent binder.

Improved predictions using additional backbone sampling and remodeling

To further assess DexDesign and DRPV (Discussion - Validation of DexDesign scaffold discovery and redesign), we performed additional computational experiments and extensive data analysis with 13 new designs using additional backbone sampling and remodeling. These results are visualized in Figs 10 and 11 (see Supplementary Tables 6 and 7 for the full dataset). As with the original DexDesign algorithm,

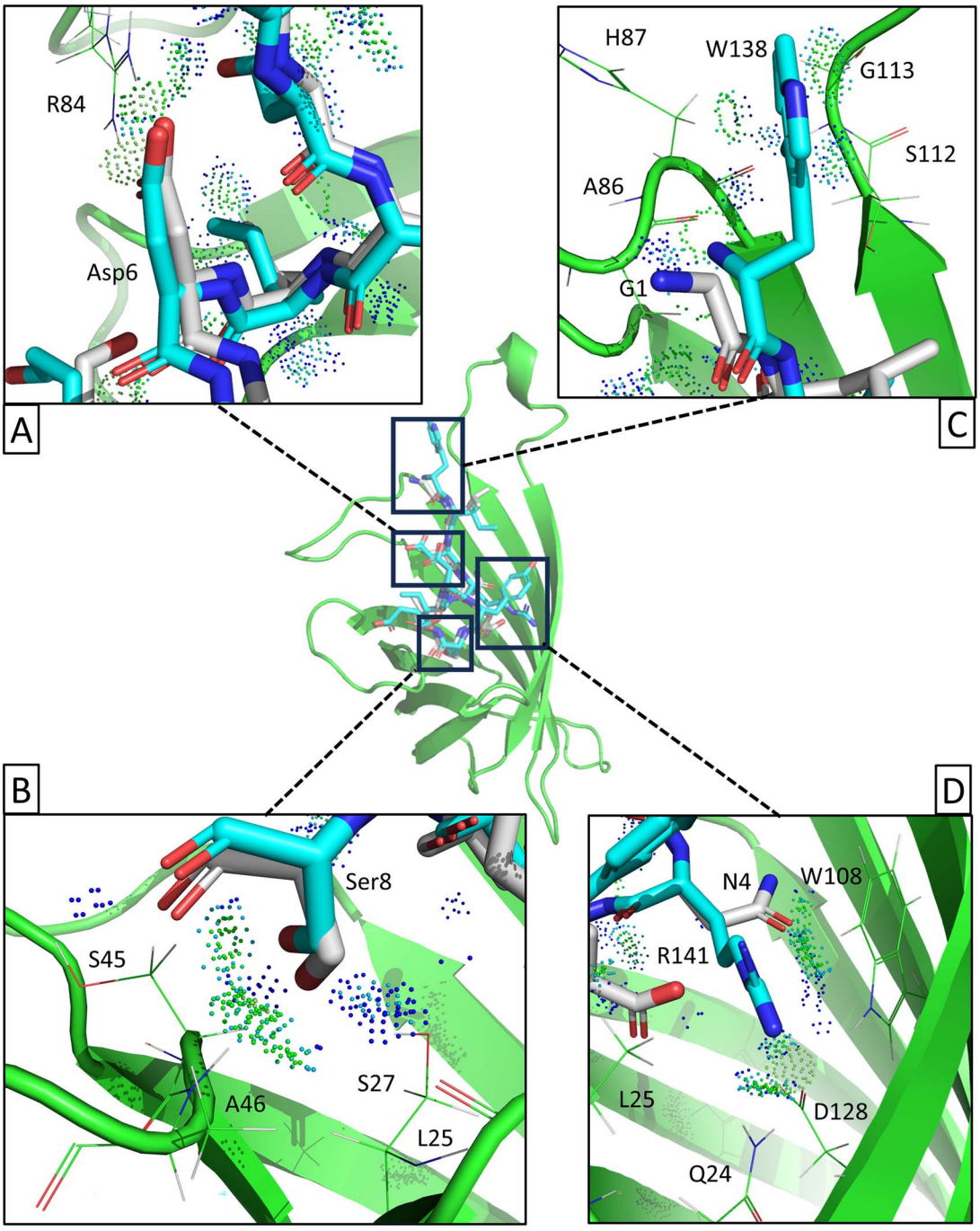

Figure 11. Match 11 creates novel favorable interactions with streptavidin not available to wildtype ligand GyGlanvdessG while replicating endogenous interactions. The wildtype ligand GyGlanvdessG is shown in grey sticks, while Match 11 is shown in cyan. Streptavidin residues are shown in green, with favorable contacts illustrated with green and blue Molprobity Probe Dots (Word et al. 1999, Jou et al. 2023). The designed D-peptide Match 11 was produced in Experiment 1 (Discussion - Improved predictions using additional backbone sampling and remodeling), where predicted binding affinity and native sequence recovery were improved by incorporating additional sampling of MASTER-returned backbones in the DexDesign algorithm. Of the 13 total substructures in computational Experiments 1 and 2, Match 11 (WIDRIDYSE) resulted in the highest native sequence recovery at 44.44%. Of designs in Discussion, Match 11 is predicted to be the 2nd-tightest binder to streptavidin with a $\log_{10}$ K* score of 38.8 (Match 500 is predicted to bind with a $\log_{10}$ K* score of 39.54, but with a native sequence recovery of 36.36%). This is an improvement over designed D-peptide DPRV (no additional sampling or minimization), which had a $\log_{10}$ K* score of 32.2 and native sequence recovery of 21.43%. (A) Match 11 recapitulates the amino acid identity (Asp) and sidechain conformation of GyGlanvdessG at the 6th position from the N-terminus. The recapitulated Asp residue of Match 11 is highly similar to the endogenous (GyGlanvdessG) conformation, resulting in the replication of favorable interactions with streptavidin residue R84. (B) Match 11 recapitulates the amino acid identity (Ser) and sidechain conformation of GyGlanvdessG at the 8th position from the N-terminus. Recapitulating the amino acid type with a similar conformation results in Match 11 replicating favorable interactions with streptavidin residues L25, S27, S45 and A46, as present with the endogenous ligand. (C) Unlike the endogenous GyGlanvdessG ligand (GLANVDESSG), Match 11's residue W138 is predicted to restitute novel interactions with streptavidin. Unlike the inert N-terminal Gly (G1) of GyGlanvdessG, W138 of Match 11 forms favorable van der Waals contacts with streptavidin residues A86, H87, S112, G113. None of these interactions are present in the GyGlanvdessG crystal structure. (D) Similar to W138, Match 11's R141 establishes favorable interactions with streptavidin that are not present in the endogenous GyGlanvdessG ligand. R141 initiates novel, favorable interactions with streptavidin residues Q24, L25, W108, and D128. GyGlanvdessG's N4 is oriented in the opposite direction in the structure, and only recapitulates some of the favorable contact with W108.

computational Experiments 1 and 2 permit ligand translation and rotation when calculating structural molecular ensembles, partition functions, and binding affinity.

In the design community, a consensus has emerged that high native sequence recovery is undesirable: a 25% to 50% native sequence recovery is considered excellent, and higher is considered overfitting (Kuriyan et al. 2012). In the case of native sequence recovery, the community has converged on this desired percentage (Kuhlman and Baker 2000), which is empirically grounded, because computational predictions are being compared to experimental sequences. Therefore, the native sequence recovery of 21% for DRPV is comparable, albeit slightly underperforms, other tools in the computational protein design domain. Moreover, using increased backbone sampling and remodeling, we have increased the native sequence recovery to 44%. Here we include the methods and results for these experiments.

We report analysis as two computational experiments: Experiment 1 and Experiment 2. Experiment 1 (see Supplementary Table 6) incorporates additional sampling of MASTER-returned backbones. That is, more than one substructure was used as a template for redesign. This computational experiment is designed to explore a larger backbone conformational space by performing DexDesign on sequences with backbone alignment RMSD ranging from 0.50–0.70 Å (as returned by MASTER). This method produced the best improvements in native sequence recovery and predicted binding affinity. Experiment 2 (see Supplementary Table 7) includes backbone minimization (*via* molecular dynamics (Case et al. 2022)) before design. Molecular dynamics simulation of atomic backbone movements were performed using the AmberTools23 *sander* (Simulated Annealing with NMR-Derived Restraints) engine (Case et al. 2023). For each backbone in Experiment 2, a non-periodic MD simulation was performed with the AMBER energy function (Case et al. 2022) and generalized Borne (Onufriev and Case 2019) solvation model. Force fields were parameterized for van der Waals and electrostatic interactions, while polar and non-polar solvent contribution terms were added to enable implicit solvation. Each substructure was evaluated for 50 time steps at 298 K. In this design strategy, substructures are minimized before any design techniques (Minimum Flexible Set, Inverse Alanine Scanning, K*-based Mutational Scanning) are performed. Because downstream design decisions are sensitive to MASTER-returned backbones (Discussion - Validation of DexDesign scaffold discovery and redesign), remodeling fragments to more realistic energies before design improved native sequence recovery and predicted binding affinity. Sampling of additional backbone conformations and remodeling before design resulted in improved K* ($\log_{10}$) scoring and native sequence recovery (sequence similarity).

Experiment 1 resulted in improved K* ($\log_{10}$) scoring and native sequence recovery (sequence similarity). Of the 10 system designed in Experiment 1, only 2 peptides, Match 2 and Match 100, have sequence similarity outside of desirable bounds. Match 11 resulted in the best sequence similarity of 44.44%, and the 2nd-highest K* score ($\log_{10}$) of 38.8. Match 500 reports the highest K* score ($\log_{10}$) of 39.54 with the 3rd-highest sequence similarity of 36.36%. Four of the ten designs have a predicted binding affinity greater than DRPV. Five systems have a predicted binding affinity greater than the wildtype GyGlanvdessG, which has a K* score ($\log_{10}$) of 32.2. These results demonstrate the benefit of additional backbone sampling, outline valuable future steps to improve D-peptide design.

Interestingly, in Experiment 2, minimization of Match 1 (the same match used to generate DPRV) before design did not change the sequence similarity of 21.43%, but did improve the K* score ($\log_{10}$) from 32.8 to 35.13. The conformational and energy landscape of the sampled backbones are improved, but perhaps do not optimize for wildtype recovery due to the presence of different D-space amino acids with higher affinity (an observation discussed in Discussion - Validation of DexDesign scaffold discovery and redesign, see ST0929 scaffold). Minimization of Match 10 and 295 increased native sequence recovery by 14.93% and 9.34%, respectively, at the cost of worse K* scores ($\log_{10}$). Therefore, while predicted binding affinity is slightly worse, the native sequence recovery is comparable to computational tools with successful experimental results (Zhou et al. 2020b, Dauparas et al. 2022, Li et al. 2022). This data reports the benefit of incorporating improved backbone sampling and diverse scaffolds from homology modeling into DexDesign for native sequence recovery with a potential trade-off between sequence recovery and predicted binding affinity.

For Experiment 1, the most conserved residues (by median) were at positions 1, 2, 5, 6, and 8 (Supplementary Table 6, Rank columns). At these positions, selection of the GyGlanvdessG wildtype residue were predicted among the top 32.5% of possible selections or better. For Experiment 2 (Supplementary Table 7, Rank columns), the most conserved residues (by median) were at positions 1, 5, 6, and 8. At these positions, selection of the wildtype residue were predicted among the top 20% of possible selections or better. Most notably, conserved regions between Experiment 1 and Experiment 2 capture most nonpolar, hydrophobic residues (GyGlanvdessG residue positions 1–3, 5), which are known to be highly conserved among high-affinity binders with intended folds (Dahiyat and Mayo 1997). Furthermore, for *in vitro*/*in vivo* experimental validation, we, and the community, rarely test only the lowest-energy design. Therefore, computational methods that demonstrate selection within roughly the top third, or better, of highly conserved sequence motifs are a reasonable and effective filter for experimental testing.

Overall, sampling of additional backbone substructures (Experiment 1) and remodeling of ligands (Experiment 2) resulted in an almost universal increase in native sequence recovery and predicted binding affinity. As reported in Discussion - Validation of DexDesign scaffold discovery and redesign, the native sequence recovery (sequence similarity) and K* ($\log_{10}$) score of DRPV without additional backbone sampling and remodeling was 21.43% and 32.8, respectively. The median sequence recovery for Experiment 1 was 31.67%, and for Experiment 2 was 30.77%. Furthermore, the median K* ($\log_{10}$) score for Experiment 1 was 31.91 and for Experiment 2 was 33.74. The best native sequence recovery was seen in Experiment 1, Match 11 with an increase from 21.43% to 44.44%. The best predicted binding affinity resulted from Experiment 1, Match 500, with an improved K* ($\log_{10}$) score from 32.8 to 39.54. Additionally, evaluation of D-amino acid selections relative to design techniques (Inverse Alanine Scanning and K*-based Mutational Scanning) indicates that DexDesign maintains highly-conserved residues while acting as an effective pruning filter for experimental testing.

Conclusions

In this work, we presented an algorithm, DexDesign, to computationally design *de novo* D-peptides. In addition, we have presented three computational protein design techniques, the Minimum Flexible Set, Inverse Alanine Scan, and K*-based Mutational Scan, that are generally applicable to both D- and L-peptide design. The process of developing DexDesign required us to add new capabilities to the OSPREY (Hallen et al. 2018) protein redesign software, including the ability to add arbitrary conformation libraries. This enables exciting new opportunities for the types of chemistries OSPREY can model and the conformations its algorithms can search and optimize. With DexDesign we have added a D-amino acid conformation library to OSPREY by reflecting the L-version of OSPREY's standard protein conformation library. Future designs and algorithms that model non-proteinogenic molecular building blocks, such as non-canonical amino acids or small molecule rotamers, are now substantially easier to implement. We envision providing additional generally useful standard conformation libraries within OSPREY itself in the future, and protein designers with specialized use cases can create their own conformation libraries and import them in their design specification in a trivial and code-free process.

We have used DexDesign to generate and optimize 30 *de novo* D-peptide inhibitors for two biomedically important PDZ targets: CALP and MAST2. We used provable approximations of binding affinity (see Supplementary Information A) and analyzed the OSPREY-predicted low-energy ensembles of the bound D-peptide: target structures to assess the quality of the novel peptides. We employed a novel restitution-replication framework for analyzing the basis upon which our DexDesign-generated D-peptides improved binding compared to their targets' endogenous ligands. Additionally, we performed additional computational experiments and extensive data analysis with 13 new designs using additional backbone sampling and remodeling that resulted in a native sequence recovery of 44%. Future work includes the important *in vitro* experimental validation of the algorithm we have presented here. There are many other peptide-recognizing PDZ domain targets for which one could use DexDesign to design *de novo* D-peptide inhibitors. Furthermore, DexDesign is not restricted to PDZ-domains, it could be applied to design novel antineoplastic, antifungal, or antibiotic D-peptide therapeutics. It is a general algorithm applicable to any target for which there exists structural models of a peptide: target complex. The structural models can be determined experimentally or computationally predicted using machine learning-based algorithms such as AlphaFold (Jumper et al. 2021, Varadi et al. 2022), although the accuracy of the results may be somewhat diminished compared to experimentally determined structures of ligand-target complexes. Thus, DexDesign provides an important tool to the drug discovery community interested in developing novel D-peptide therapeutics.

Acknowledgements

We thank all members of the Donald lab for helpful discussions. We also thank the anonymous reviewers for their helpful suggestions.

Funding

We received funding from the NIH (grants R35-GM144042 to BRD and R01-AI139216 to P.Z.).

Author contributions

Nathan Guerin (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Software [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Henry Childs (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Software [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Pei Zhou (Conceptualization [equal], Funding acquisition [equal], Writing—review & editing [equal]), and Bruce Donald (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Resources [equal], Software [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]).

Supplementary data

Supplementary data is available at *PROENG Journal* online.

Conflict of interest

BRD is a founder of Ten63 Therapeutics, Inc. NSG is currently employed by Ten63 Therapeutics, Inc.

Data availability

OSPREY can be found at github.com/donaldlab/OSPREY3.

References

Amacher JF, Cushing PR, Bahl CD. *et al.* Stereochemical determinants of C-terminal specificity in PDZ peptide-binding domains. *J Biol Chem* 2013;**288**:5114–26. https://doi.org/10.1074/jbc.M112.401588.

Amacher JF, Brooks L, Hampton TH. *et al.* Specificity in PDZ-peptide interaction networks: computational analysis and review. *J Struct Biol* 2020;**4**:100022.

Angelini A, Cendron L, Chen S. *et al.* Bicyclic peptide inhibitor reveals large contact Interface with a protease target. *ACS Chem Biol* 2012;**7**:817–21. https://doi.org/10.1021/cb200478t.

Banting FG. Banting—nobel lecture. In: *Nobel Prize Outreach AB 2023*. 2024. https://www.nobelprize.org/prizes/medicine/1923/banting/lecture/

Berman HM, Westbrook J, Feng Z. *et al.* The protein data Bank. *Nucleic Acids Res* 2000;**28**:235–42. https://doi.org/10.1093/nar/28.1.235.

Bhardwaj G, Mulligan VK, Bahl CD. *et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* 2016;**538**:329–35. https://doi.org/10.1038/nature19791.

Brower MS, Brakel CL, Garry K. Immunodetection with streptavidin-acid phosphatase complex on western blots. *Anal Biochem* 1985;**147**:382–6. https://doi.org/10.1016/0003-2697(85)90286-6.

Caillet-Saguy C, Maisonneuve P, Delhommel F. *et al.* Strategies to interfere with PDZ-mediated interactions in neurons: what we can learn from the rabies virus. *Prog Biophys Mol Biol* 2015;**119**:53–9. https://doi.org/10.1016/j.pbiomolbio.2015.02.007.

Case DA, Aktulga HM, Belfon K. *et al. Amber 2022*, 2022.

Case DA, Aktulga HM, Belfon K. *et al.* AmberTools. *J Chem Inf Model* 2023;**63**:6183–91. https://doi.org/10.1021/acs.jcim.3c01153.

Chen S, Gfeller D, Buth SA. *et al.* Improving binding affinity and stability of peptide ligands by substituting Glycines with D-amino acids. *Chembiochem* 2013;**14**:1316–22. https://doi.org/10.1002/cbic.201300228.

Christensen NR, Čalyševa J, Fernandes EFA. *et al.* PDZ domains as drug targets. *Adv Ther (Weinh)* 2019;**2**:1800143. https://doi.org/10.1002/adtp.201800143.

Cielo CBC, Okazaki S, Suzuki A. *et al.* Structure of ST0929, a putative glycosyl transferase from Sulfolobus tokodaii. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2010;**66**:397–400. https://doi.org/10.1107/S1744309110006354.

Craik DJ, Fairlie DP, Liras S. *et al.* The future of peptide-based drugs. *Chem Biol Drug Des* 2013;**81**:136–47. https://doi.org/10.1111/cbdd.12055.

Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. *Science* 1997;**278**:82–7. https://doi.org/10.1126/science.278.5335.82.

Dauparas J, Anishchenko I, Bennett N. *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 2022;**378**:49–56. https://doi.org/10.1126/science.add2187.

Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. *Trends Biochem Sci* 2011;**36**:159–69. https://doi.org/10.1016/j.tibs.2010.10.002.

Delhommel F, Chaffotte A, Terrien E. *et al.* Deciphering the unconventional peptide binding to the PDZ domain of MAST2. *Biochem J* 2015;**469**:159–68. https://doi.org/10.1042/BJ20141198.

Di L. Strategic approaches to optimizing peptide ADME properties. *AAPS J* 2014;**17**:134–43. https://doi.org/10.1208/s12248-014-9687-3.

Donald BR. *Algorithms in Structural Molecular Biology*. Cambridge, MA: MIT Press, 2011.

Doti N, Mardirossian M, Sandomenico A. *et al.* Recent applications of retro-Inverso peptides. *Int J Mol Sci* 2021;**22**:8677. https://doi.org/10.3390/ijms22168677.

Dougherty PG, Wellmerling JH, Koley A. *et al.* Cyclic peptidyl inhibitors against CAL/CFTR interaction for treatment of cystic fibrosis. *J Med Chem* 2020;**63**:15773–84. https://doi.org/10.1021/acs.jmedchem.0c01528.

Doyle DA, Lee A, Lewis J. *et al.* Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* 1996;**85**:1067–76. https://doi.org/10.1016/S0092-8674(00)81307-0.

Elkin CD, Zuccola HJ, Hogle JM. *et al.* Computational design of D-peptide inhibitors of hepatitis delta antigen dimerization. *J Comput Aided Mol Des* 2000;**14**:705–18. https://doi.org/10.1023/A:1008146015629.

Fosgerau K, Hoffmann T. Peptide therapeutics: current status and future directions. *Drug Discov Today* 2015;**20**:122–8. https://doi.org/10.1016/j.drudis.2014.10.003.

Freitag S, Le Trong I, Klumb L. *et al.* Structural studies of the streptavidin binding loop. *Protein Sci* 1997;**6**:1157–66. https://doi.org/10.1002/pro.5560060604.

Freitag S, Le Trong I, Klumb LA. *et al.* X-ray crystallographic studies of streptavidin mutants binding to biotin. *Biomol Eng* 1999;**16**:13–9. https://doi.org/10.1016/S1050-3862(99)00048-0.

Frey KM, Georgiev I, Donald BR. *et al.* Predicting resistance mutations using protein design algorithms. *Proc Natl Acad Sci* 2010;**107**:13707–12. https://doi.org/10.1073/pnas.1002162107.

Gainza P, Roberts KE, Donald BR. Protein design using continuous rotamers. *PLoS Comput Biol* 2012;**8**:e1002335. https://doi.org/10.1371/journal.pcbi.1002335.

Garton M, Sayadi M, Kim PM. A computational approach for designing D-proteins with non-canonical amino acid optimised binding affinity. *PloS One* 2017;**12**:e0187524. https://doi.org/10.1371/journal.pone.0187524.

Garton M, Nim S, Stone TA. *et al.* Method to generate highly stable D-amino acid analogs of bioactive helical peptides using a mirror image of the entire PDB. *Proc Natl Acad Sci* 2018;**115**:1505–10. https://doi.org/10.1073/pnas.1711837115.

Georgiev I, Donald BR. *OSPREY User Manual v1.0*, 2009. https://donaldlab.cs.duke.edu/software/osprey/osprey1.0/OSPREY.pdf.

Georgiev I, Lilien RH, Donald BR. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *J Comput Chem* 2008;**29**:1527–42. https://doi.org/10.1002/jcc.20909.

Guerin N, Feichtner A, Stefan E. *et al.* Resistor: an algorithm for predicting resistance mutations via Pareto optimization over multistate protein design and mutational signatures. *Cell Systems* 2022;**13**:830–843.e3. https://doi.org/10.1016/j.cels.2022.09.003.

Guerin N, Kaserer T, Donald BR. Protocol for predicting drug-resistant protein mutations to an ERK2 inhibitor using RESISTOR. *STAR Protocols* 2023;**4**:102170. https://doi.org/10.1016/j.xpro.2023.102170.

Hallen MA, Donald BR. CATS (coordinates of atoms by Taylor series): protein design with backbone flexibility in all locally feasible directions. *Bioinformatics* 2017;**33**:i5–i12. https://doi.org/10.1093/bioinformatics/btx277.

Hallen MA, Martin JW, Ojewole A. *et al.* OSPREY 3.0: open-source protein redesign for you, with powerful new features. *J Comput Chem* 2018;**39**:2494–507. https://doi.org/10.1002/jcc.25522.

Harris BZ, Lim WA. Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* 2001;**114**:3219–31. https://doi.org/10.1242/jcs.114.18.3219.

Haugaard-Kedström LM, Clemmensen LS, Sereikaite V. *et al.* A high-affinity peptide ligand targeting Syntenin inhibits glioblastoma. *J Med Chem* 2021;**64**:1423–34. https://doi.org/10.1021/acs.jmedchem.0c00382.

Holt GT, Jou JD, Gill NP. *et al.* Computational analysis of energy landscapes reveals dynamic features that contribute to binding of inhibitors to CFTR-associated ligand. *J Phys Chem B* 2019;**123**:10441–55. https://doi.org/10.1021/acs.jpcb.9b07278.

Holt GT, Gorman J, Wang S. *et al.* Improved HIV-1 neutralization breadth and potency of V2-apex antibodies by in silico design. *Cell Rep* 2023;**42**:112711. https://doi.org/10.1016/j.celrep.2023.112711.

Hyre DE, Le Trong I, Merritt EA. *et al.* Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci* 2006;**15**:459–67. https://doi.org/10.1110/ps.051970306.

Ivarsson Y. Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Lett* 2012;**586**:2638–47. https://doi.org/10.1016/j.febslet.2012.04.015.

Jemth P, Gianni S. PDZ domains: folding and binding. *Biochemistry* 2007;**46**:8701–8. https://doi.org/10.1021/bi7008618.

Jou JD, Holt GT, Lowegard AU. *et al.* Minimization-aware recursive K∗: a novel, provable algorithm that accelerates ensemble-based protein design and provably approximates the energy landscape. *J Comput Biol* 2020;**27**:550–64. https://doi.org/10.1089/cmb.2019.0315.

Jou JD, Guerin N, Roberts KE. *Protein Design Plugin*, 2023; https://github.com/donaldlab/ProteinDesignPlugin.

Jou JD, Guerin N, Roberts KE. *Protein Design Plugin*, 2024; https://github.com/donaldlab/ProteinDesignPlugin.

Jumper J, Evans R, Pritzel A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2.

Kaserer T, Blagg J. Combining mutational signatures, clonal fitness, and drug affinity to define drug-specific resistance mutations in cancer. *Cell Chem Biol* 2018;**25**:1359–1371.e2. https://doi.org/10.1016/j.chembiol.2018.07.013.

Khan Z, Terrien E, Delhommel F. *et al.* Structure-based optimization of a PDZ-binding motif within a viral peptide stimulates neurite outgrowth. *J Biol Chem* 2019;**294**:13755–68. https://doi.org/10.1074/jbc.RA119.008238.

Kugler V, Lieb A, Guerin N. *et al.* Disruptor: computational identification of oncogenic mutants disrupting protein-protein and protein-DNA interactions. *Commun Biol* 2023;**6**:1–6. https://doi.org/10.1038/s42003-023-05089-2.

Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 2000;**97**:10383–8. https://doi.org/10.1073/pnas.97.19.10383.

Kuriyan J, Konforti B, Wemmer D. *The Molecules of Life: Physical and Chemical Principles*. New York, NY: W.W. Norton & Company, 2012. https://doi.org/10.1201/9780429258787.

Lander AJ, Jin Y, Luk LYP. D-peptide and D-protein technology: recent advances, challenges, and opportunities. *Chembiochem* 2023;**24**:e202200537. https://doi.org/10.1002/cbic.202200537.

Lee H-J, Zheng JJ. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal* 2010;**8**:8. https://doi.org/10.1186/1478-811X-8-8.

Li AJ, Sundar V, Grigoryan G. *et al.* TERMinator: a neural framework for structure-based protein design using tertiary repeating motifs-Preprint at http://arxiv.org/abs/2204.13048. 2022.

Lilien RH, Stevens BW, Anderson AC. *et al.* A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase a phenylalanine adenylation enzyme. *J Comput Biol* 2005;**12**:740–61. https://doi.org/10.1089/cmb.2005.12.740.

Liu M, Li C, Pazgier M. *et al.* D-peptide inhibitors of the p53-MDM2 interaction for targeted molecular therapy of malignant neoplasms. *Proc Natl Acad Sci U S A* 2010;**107**:14321–6. https://doi.org/10.1073/pnas.1008930107.

Lovell SC, Word JM, Richardson JS. *et al.* The penultimate rotamer library. *Proteins* 2000;**40**:389–408. https://doi.org/10.1002/1097-0134(20000815)40:3<389::AID-PROT50>3.0.CO;2-2.

Lowegard AU, Frenkel MS, Holt GT. *et al.* Novel, provable algorithms for efficient ensemble-based computational protein design and their application to the redesign of the c-Raf-RBD:KRas protein-protein interface. *PLoS Comput Biol* 2020;**16**:e1007447. https://doi.org/10.1371/journal.pcbi.1007447.

Lyamichev VI, Goodrich LE, Sullivan EH. *et al.* Stepwise evolution improves identification of diverse peptides binding to a protein target. *Sci Rep* 2017;**7**:12116. https://doi.org/10.1038/s41598-017-12440-1.

Martin J, Donald BR. *Conformation Energy:: OSPREY*. OSPREY Documentation, 2024; https://www2.cs.duke.edu/donaldlab/software/osprey/docs/contributing/architecture/conf-energy/.

Melero C, Ollikainen N, Harwood I. *et al.* Quantification of the transferability of a designed protein specificity switch reveals extensive epistasis in molecular recognition. *Proc Natl Acad Sci U S A* 2014;**111**:15426–31. https://doi.org/10.1073/pnas.1410624111.

Mignon D, Panel N, Chen X. *et al.* Computational design of the Tiam1 PDZ domain and its ligand binding. *J Chem Theory Comput* 2017;**13**:2271–89. https://doi.org/10.1021/acs.jctc.6b01255.

Miles JJ, Tan MP, Dolton G. *et al.* Peptide mimic for influenza vaccination using nonnatural combinatorial chemistry. *J Clin Invest* 2018;**128**:1569–80. https://doi.org/10.1172/JCI91512.

Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* 1991;**11**:29–34. https://doi.org/10.1002/prot.340110104.

Nakariyakul S, Liu Z-P, Chen L. A sequence-based computational approach to predicting PDZ domain-peptide interactions. *Biochim Biophys Acta* 2014;**1844**:165–70. https://doi.org/10.1016/j.bbapap.2013.04.008.

Nardella C, Visconti L, Malagrinò F. *et al.* Targeting PDZ domains as potential treatment for viral infections, neurodegeneration and cancer. *Biol Direct* 2021;**16**:15. https://doi.org/10.1186/s13062-021-00303-9.

Noether E. *Gesammelte Abhandlungen-Collected Papers, Springer Collected Works in Mathematics*. Berlin, Heidelberg: Springer, 1983. https://doi.org/10.1007/978-3-642-39990-9.

Oda Y, Saeki K, Takahashi Y. *et al.* Crystal structure of tobacco necrosis virus at 2.25 Å resolution. *J Mol Biol* 2000;**300**:153–69. https://doi.org/10.1006/jmbi.2000.3831.

Ojewole AA, Jou JD, Fowler VG. *et al.* 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* 2018;**25**:726–39. https://doi.org/10.1089/cmb.2017.0267.

Onufriev AV, Case DA. Generalized born implicit solvent models for biomolecules. *Annu Rev Biophys* 2019;**48**:275–96. https://doi.org/10.1146/annurev-biophys-052118-115325.

Opuu V, Sun YJ, Hou T. *et al.* A physics-based energy function allows the computational redesign of a PDZ domain. *Sci Rep* 2020;**10**:11150. https://doi.org/10.1038/s41598-020-67972-w.

Panel N, Villa F, Opuu V. *et al.* Computational design of PDZ-peptide binding. *Methods Mol Biol* 2021;**2256**:237–55. https://doi.org/10.1007/978-1-0716-1166-1_14.

Piserchio A, Fellows A, Madden DR. *et al. PDZ Domain of CAL (Cystic Fibrosis Transmembrane Regulator-Associated Ligand)*. Protein Data Bank (PDB), 2012. https://doi.org/10.2210/pdb2LOB/pdb.

Préhaud C, Wolff N, Terrien E. *et al.* Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. *Sci Signal* 2010;**3**:ra5. https://doi.org/10.1126/scisignal.2000510.

Reeve SM, Gainza P, Frey KM. *et al.* Protein design algorithms predict viable resistance to an experimental antifolate. *Proc Natl Acad Sci* 2015;**112**:749–54. https://doi.org/10.1073/pnas.1411548112.

Reeve SM, Si D, Krucinska J. *et al.* Toward broad spectrum dihydrofolate reductase inhibitors targeting trimethoprim resistant enzymes identified in clinical isolates of methicillin resistant Staphylococcus aureus. *ACS Infect Dis* 2019;**5**:1896–906. https://doi.org/10.1021/acsinfecdis.9b00222.

Renfrew PD, Choi EJ, Bonneau R. *et al.* Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PloS One* 2012;**7**:e32637. https://doi.org/10.1371/journal.pone.0032637.

Roberts KE, Cushing PR, Boisguerin P. *et al.* Computational design of a PDZ domain peptide inhibitor that rescues CFTR activity. *PLoS Comput Biol* 2012;**8**:e1002477. https://doi.org/10.1371/journal.pcbi.1002477.

Rudicell RS, Kwon YD, Ko S-Y. *et al.* Enhanced potency of a broadly neutralizing HIV-1 antibody in vitro improves protection against lentiviral infection in vivo. *J Virol* 2014;**88**:12669–82. https://doi.org/10.1128/JVI.02213-14.

Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 1.8*, 2015.

Skelton NJ, Koehler MFT, Zobel K. *et al.* Origins of PDZ domain ligand specificity: structure determination and mutagenesis of the erbin pdz domain. *J Biol Chem* 2003;**278**:7645–54. https://doi.org/10.1074/jbc.M209751200.

Smith CA, Kortemme T. Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. *J Mol Biol* 2010;**402**:460–74. https://doi.org/10.1016/j.jmb.2010.07.032.

Smith CA, Shi CA, Chroust MK. *et al.* Design of a phosphorylatable PDZ domain with peptide-specific affinity changes. *Structure* 2013;**21**:54–64. https://doi.org/10.1016/j.str.2012.10.007.

Stevens BW, Lilien RH, Georgiev I. *et al.* Redesigning the PheA domain of gramicidin Synthetase leads to a new understanding of the Enzyme's mechanism and selectivity. *Biochemistry* 2006;**45**:15495–504. https://doi.org/10.1021/bi061788m.

Tahti EF, Blount JM, Jackson SN. Additive energetic contributions of multiple peptide positions determine the relative promiscuity of viral and human sequences for PDZ domain targets. *Protein Sci* 2023;**32**:e4611. https://doi.org/10.1002/pro.4611.

Terrien E, Wolff N, Cordier F. *et al. Solution Structure of MAST2-PDZ Complexed with the C-Terminus of PTEN*. Protein Data Bank (PDB), 2010. https://doi.org/10.2210/pdb2KQF/pdb.

Terrien E, Chaffotte A, Lafage M. *et al.* Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. *Sci Signal* 2012;**5**:ra58.

Valiente PA, Wen H, Nim S. *et al.* Computational design of potent D-peptide inhibitors of SARS-CoV-2. *J Med Chem* 2021;**64**:14955–67. https://doi.org/10.1021/acs.jmedchem.1c00655.

Valiente PA, Nim S, Lee J. *et al.* Targeting the receptor-binding motif of SARS-CoV-2 with D-peptides mimicking the ACE2 binding helix: lessons for inhibiting omicron and future variants of concern.

J Chem Inf Model 2022;**62**:3618–26. https://doi.org/10.1021/acs.jcim.2c00500.

Varadi M, Anyango S, Deshpande M. *et al.* AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022;**50**:D439–44. https://doi.org/10.1093/nar/gkab1061.

Wang S. *Computational Protein Design with Non-proteinogenic Amino Acids and Small Molecule Ligands, with Applications to Protein-Protein Interaction Inhibitors, Anti-Microbial Enzyme Inhibitors, and Antibody Design*, Dissertation, Duke University, 2021.

Wang F, Langley R, Gulten G. *et al.* Identification of a type III thioesterase reveals the function of an operon crucial for Mtb virulence. *Chem Biol* 2007;**14**:543–51. https://doi.org/10.1016/j.chembiol.2007.04.005.

Wang H, Feng Z, Xu B. D-amino acid-containing supramolecular nanofibers for potential cancer therapeutics. *Adv Drug Deliv Rev* 2017;**110–111**:102–11. https://doi.org/10.1016/j.addr.2016.04.008.

Wang L, Wang N, Zhang W. *et al.* Therapeutic peptides: current applications and future directions. *Signal Transduct Target Ther* 2022a;**7**:48. https://doi.org/10.1038/s41392-022-00904-4.

Wang S, Reeve SM, Holt GT. *et al.* Chiral evasion and stereospecific antifolate resistance in Staphylococcus aureus. *PLoS Comput Biol* 2022b;**18**:e1009855. https://doi.org/10.1371/journal.pcbi.1009855.

Wei X, Zhan C, Shen Q. *et al.* A D-peptide ligand of nicotine acetylcholine receptors for brain-targeted drug delivery. *Angew Chem Int Ed Engl* 2015;**54**:3023–7. https://doi.org/10.1002/anie.201411226.

Williams CJ, Headd JJ, Moriarty NW. *et al.* MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci* 2018;**27**:293–315. https://doi.org/10.1002/pro.3330.

Word JM, Lovell SC, LaBean TH. *et al.* Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* 1999;**285**:1711–33. https://doi.org/10.1006/jmbi.1998.2400.

Zheng F, Jewell H, Fitzpatrick J. *et al.* Computational design of selective peptides to discriminate between similar PDZ domains in an oncogenic pathway. *J Mol Biol* 2015;**427**:491–510. https://doi.org/10.1016/j.jmb.2014.10.014.

Zhou J, Grigoryan G. Rapid search for tertiary fragments reveals protein sequence–structure relationships. *Protein Sci* 2015;**24**:508–24. https://doi.org/10.1002/pro.2610.

Zhou X, Zuo C, Li W. *et al.* A novel d-peptide identified by mirror-image phage display blocks TIGIT/PVR for cancer immunotherapy. *Angew Chem Int Ed Engl* 2020a;**59**:15114–8. https://doi.org/10.1002/anie.202002783.

Zhou J, Panaitiu AE, Grigoryan G. A general-purpose protein design framework based on mining sequence–structure relationships in known protein structures. *Proc Natl Acad Sci U S A* 2020b;**117**:1059–68. https://doi.org/10.1073/pnas.1908723117.