Title:

De novo design of drug-binding proteins with predictable binding energy and specificity

Authors:

Lei Lu[1], Xuxu Gou[2], Sophia K Tan[1], Samuel I. Mann[1,3], Hyunjun Yang[1], Xiaofang Zhong[4], Dimitrios Gazgalis[5], Jesús Valdiviezo[5], Hyunil Jo[1], Yibing Wu[1], Morgan E. Diolaiti[2], Alan Ashworth[2], Nicholas F. Polizzi[5], William F. DeGrado[1]*

Affiliations:

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94158, USA.
[2]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA 94158, USA.
[3]Department of Chemistry, University of California, Riverside, CA 92521, United States
[4]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA.
[5]Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA.

*Corresponding author. Email: william.degrado@ucsf.edu

Abstract
The *de novo* design of small-molecule-binding proteins has seen exciting recent progress; however, high affinity binding and tunable specificity typically require laborious screening and optimization after computational design. We developed a computational procedure to design a protein that recognizes a common pharmacophore in a series of PARP1 inhibitors. One of three designed proteins bound different inhibitors with affinities ranging from < 5 nM to low μM. X-ray crystal structures confirmed the accuracy of the designed protein-drug interactions. Molecular dynamics simulations informed the role of water in binding. Binding free-energy calculations performed directly on the designed models are in excellent agreement with the experimentally measured affinities. We conclude that *de novo* design of high-affinity small-molecule-binding proteins with tuned interaction energies is feasible entirely from computation.

One Sentence Summary
Informatic sampling enables *de novo* creation of drug-binding proteins with accurate design and prediction of binding affinity.

Main Text

Molecular recognition underlies small molecule binding by and catalytic activity of protein receptors and enzymes. Although we have an advanced understanding of both protein design and molecular interactions, the rational design of *de novo* proteins that specifically bind small molecules with low nM to pM affinity is a major challenge (*1*, *2*) that has not been achieved in *de novo* proteins (*3*, *4*) without experimental screening of large libraries of variants (*5–7*). Even with the application of recent advances in artificial intelligence to facilitate *de novo* design (*8–10*), it has been necessary to screen thousands of independent designs to discover binders with low μM to high nM dissociation constants (K_D) directly from design algorithms (*3*, *11–14*). Proteins with higher affinity are often desirable. Given the low success rate, screening large numbers of designs often relies on biotinylated or fluorescently-labeled versions of their small molecule targets, which restricts the region of the molecule available for binding. The cost of synthesizing thousands of genes and the necessity of synthetic chemistry for conjugation places a practical limitation on the access of these methods to many groups. Moreover, *de novo* design methods rely heavily on structural informatics to guide sampling of protein structure and sequence, as well as scoring functions that rely on a mix of statistical and physical terms without explicit representation of dynamics, conformational entropy, or water (*3–7*). This dependence leaves open the fundamental question of whether our understanding is grounded in physical forces or limited exclusively to advanced pattern recognition (*15*). Here, we asked whether adherence to simple rules based on physical principles might increase the success rate for designing drug-binding proteins, and whether all-atom molecular dynamics simulations with explicit water might be able to recapitulate the experimentally determined binding affinities of high affinity binders, starting with design models that come directly from a computational design.

The recognition of polar groups presents a challenge in *de novo* design of binders, because the polar groups must lose most or all of their highly favorable interactions with water molecules upon binding to the protein. To compensate for this loss in hydration, the drug's polar chemical groups must form highly directional and distance-dependent hydrogen bonds and electrostatic interactions with atomic groups in the protein. These polar interactions are not only required for affinity, but they also provide the specificity of proteins for their substrates over other similarly shaped molecules.

We recently developed a design method known as COMBS (Convergent Motifs for Binding Sites) to enable sampling of only sequences and structures capable of forming such interactions prior to searching for less specific and directional, but energetically favorable, van der Waals and hydrophobic interactions that complete the binding site (*3*). To facilitate this process COMBS uses van der Mers (vdMs) to search for preferred spatial positions where chemical groups can interact with an amino acid (*3*). vdMs are similar to rotamers (*16*), which define favorable positions to arrange an amino acid's sidechain atoms relative to its backbone atoms. However, instead vdMs define favorable positions of an interacting chemical-group fragment relative to a residue's backbone atoms. Each amino acid type can adopt multiple rotamers, which are widely used in design algorithms to position sidechains onto pre-existing backbones. In vdMs we track the positions of chemical groups that can interact with a given type of amino acid. For example, a Gln_{CONH2} consists of a Gln residue that contacts a carboxamide group. Also like rotamers, vdMs can be clustered into similar groups with associated probabilities based on their occurrence in the pdb. The COMBS algorithm then finds multiple positions on a given protein backbone that can simultaneously form favorable van der Waals, aromatic, and/or hydrogen-bonded interactions with

the chemical groups of a target small molecule. The remainder of the sequence is then filled in using flexible backbone sequence design .

COMBS showed promise in identifying protein backbones and creating sites capable of binding the drug apixaban. In previous work, we identified one protein that bound with K_D = 500 nM and a second with K_D = 5 μM after screening only six designed sequences (*3*). In the process, we learned lessons that could improve the use of COMBS in design pipelines. First, during the final steps of sequence design and backbone optimization, sidechains were sometimes introduced that ultimately did not form their intended favorable vdMs. Structural analysis showed that they were not in optimal orientations to interact with the drug, and site directed mutagenesis showed they made small or no contributions to the free energy of binding. We reasoned it should be possible to improve binding by using backbone phi/psi-dependent vdMs and alternating rounds of COMBS with Rosetta flexible-backbone sequence design. Secondly, binding affinity can be optimized by pre-organizing a receptor's conformation so that it loses minimal conformational entropy upon binding. Originally, we analyzed preorganization using Rosetta ab initio folding; now using AlphaFold2, we should be able to reliably determine whether a protein adopted the desired, pre-organized conformation prior to experimental characterization.

Another feature of importance is the need to consider the energetically unfavorable loss of hydrogen-bonds between water molecules and both the drug as well as the protein's binding site, which occurs when the drug binds into the pocket. While COMBS and other algorithms **Please add HBNet reference** considered the need to form hydrogen bonds to compensate for the loss of hydration, here we strive to form a more full set of compensatory ligand-protein hydrogen bonds to every buried polar atom. We sample vdMs between the ligand and the first-shell amino acids, as well as vdMs between the first shell and a second shell of interacting residues, which also assures a favorable geometry for the binding partners. We also opted to bias the orientation of the ligand such that formally charged groups are placed near the surface of the protein, thereby minimizing energic penalties due to Born solvation. Finally, we evaluated MD simulations and free energy calculations, which explicitly consider interactions with bulk and bound water molecules that are not fully considered in protein design algorithms, to assess the usefulness of physics-based methods for evaluating the affinity of the designs.

To demonstrate the utility of our refined methods, we chose to design inhibitors of PARP, a recently developed class of clinically useful anticancer drug (*17*). *De novo* designed binders of PARPi drugs might serve as components in detectors, delivery agents, or detoxification agents for these cytotoxic drugs. The predominant class of PARPi drugs share a tripartite pharmacophore consisting of a fused 5,6-bicyclic core, an amide and a phenyl group bearing a positively charged alkylamine (Fig. 1A). We chose to target rucaparib, the most structurally complex of several related drugs, as our primary target (Fig.1A), as well as a series PARPi analogues. By considering a series of drugs, we at once provide reagents that might be widely useful, while simultaneously testing our understanding of the essential features required for binding.

***De novo* design of high-affinity drug-binding proteins**

We used a recursive version of the COMBS algorithm to design rucaparib-binding sites in a family of mathematically generated four-helix bundle proteins. The key binding residues were introduced using vdMs to identify multiple positions on a given protein backbone that could simultaneously interact favorably with the chemical groups of a target small molecule (Fig. 1B, 1C, Supplementary Methods). Although vdMs are derived from statistics of sidechain and mainchain interactions with one another in the pdbs they can be used to identify binding-site residues capable of forming hydrogen bonds and aromatic interactions with diverse small

molecules, in much the same way that natural proteins bind a wide diversity of small molecules using a set of 20 amino acids. While the energetics of these interactions can vary depending on the specific small molecule bound, the fundamental geometries required to achieve binding remain often remain relatively constant (*18*). Thus, a common set of vdMs should serve to bind a wide range of compounds. We targeted three chemical groups in rucaparib's structure: the indole NH, and the C=O and NH_2 groups of its carboxamide (Fig. 1C). Additionally, COMBS identified Asp58 as a second-shell interaction to the carboxamide of rucaparib (Fig. 1C). It is important to design binding interactions with these groups with sub-Å accuracy to engender specificity and a favorable free energy of association. Next, the remainder of the sequence was designed using Rosetta flexible backbone design (Fig. 1D) (*3, 19*)**,** while retaining the identity of the keystone residues (identified in the COMBS step). The mainchain moved 1 Å rmsd during this step (Fig. S1), so a second round of vdM sampling was performed on the relaxed backbone. This procedure identified three mutants involving drug-contacting residues, including N29D, W90L and N131D (Fig.1E). A second round of flexible backbone sequence design using this backbone and the newly fixed vdMs resulted in converged sequence/structure combinations (Fig. 1E, 2A-B), as a third round of COMBS showed the vdMs were now optimal. The final designs include numerous CH-π and hydrophobic interactions interspersed with specific polar interactions, including four H-bonds (an H-bond donor to the drug's carboxamide oxygen as well as three H-bond acceptors to the drug's carboxamide NH_2, indole NH and charged ammonium group), as well as second shell interactions. (Fig. 2C, Fig. S2).

Throughout the design process, we ensured that the designs would also retain favorable interactions with most of the common pharmacophores of the three other drugs (see supplementary methods for details). However, we predicted that the protein would have lower affinity for niraparib and mefuparib, because they lack the H-bonding group indole NH of rucaparib. Also, we expected veliparib to bind weakly, because it lacks a hydrophobic phenyl group and the position of its charged ammonium group differs from that found in the other three drugs.

The final models were chosen based on multiple criteria: 1) favorable vdMs (highest total vdM cluster scores); 2) satisfaction of all potential buried H-bond donors in the protein and ligand; 3) low Rosetta energy (lowest 50 of the 1000 total designs); and 4) avoidance of clashes with the three other PARP inhibitors, which show structural variability near the amine end of the molecule. The three top-scoring designs were selected for expression including: PiB (PARPi binder) and a variant of this protein, PiB' (Table. S1, Fig. S3), which differs only by the substitution of five solvent-exposed charged residues with Ala to encourage crystallization. The other two (PiB-1 and PiB-2) were less closely related to PiB in structure (r.m.s.d. = 0.93 and 0.79 for PiB-1 and PiB-2 to PiB, respectively) and sequence (41% and 42% identity for PiB-1 and PiB-2 to PiB, respectively, Fig. S4). Circular dichroism spectroscopy showed all four had substantial alpha-helical character (Fig. S5). However, PiB-1 and PiB-2 failed to induce large changes in the fluorescence emission spectrum of rucaparib (Fig. S6). Therefore, we focused our efforts on PiB and PiB' (Fig. S7 - S11).

Spectral titrations showed that PiB and PiB' bound the PARPi drugs with high affinity. Incubation of PiB with equimolar concentrations of rucaparib led to a marked blue shift and an increase in intensity of its fluorescence spectrum, as expected if its indole core were bound in a rigid, solvent-inaccessible site (Fig. S6, S7). NMR spectroscopy of PiB showed that it folded into a well-defined structure, and the addition of a single equivalent of rucaparib led to a new set of peaks, consistent with a stoichiometric, specific complex (Fig. S10, S11). Fluorescently monitored titrations of protein into a solution of rucaparib showed that PiB and PiB' bound with very low to sub-nM affinity (Fig. 2D, 2E, 3A). Even at the lowest experimentally feasible rucaparib

concentration, the binding isotherms show a linear increase in intensity with respect to protein concentration until a single equivalent is added, followed by an abrupt leveling at higher protein concentrations. This behavior is indicative of a dissociation constant that is much lower than the total rucaparib concentration. A non-linear least-squares fit to the data returned a K_D of 2.2 nM for PiB and 0.37 nM For PiB', and a sensitivity analysis showed that the K_D was less than 5 nM for both proteins (Fig. 2D, 2E). Achieving single digit nM to pM binding affinity for *de novo* designed proteins has previously required extensive experimental optimization (*5*).

UV/visible absorption titrations showed that PiB and PiB' also bound to the remaining ligands with affinities that grew increasingly weaker as the drugs' structures diverged from rucaparib (Fig. 3A, Fig. S12). PiB retained sub-μM affinity for mefuparib (K_D = 190 nM and 350 nM for PiB and PiB', respectively) and niraparib (600 and 550 nM). The corresponding K_D values were 14 μM and 24 μM, respectively for the structurally divergent drug veliparib, and no binding was detected for the most divergent drug, olaparib (Fig. S13). This observed trend in binding affinity matches the order expected from the structural differences mentioned above.

We next examined the *in vitro* stability and potency of PiB and PiB' in serum and cellular assays. PiB and PiB' were highly stable in human serum, as are other *de novo* proteins designed for medical applications (Fig. S14, S15) (*3, 20*). PARP inhibitors potently inhibit the viability of cells with certain DNA repair deficiencies, including loss-of-function mutations in *BRCA2* (*17*). To determine whether PiB and PiB' could attenuate the lethal effects of PARPi drugs, we measured their effects on the growth of *BRCA2* mutated DLD-1 cells and SUM149 cells (*21*) over after an 8-day incubation. Dose-response curves were first established in the absence of PiB, then the titration was repeated with PiB or PiB', at varying [protein]/[drug] ratios for each PARPi drug concentration. Addition of a single equivalent of PiB or PiB' resulted in a 4-fold increase in the half maximal inhibitory concentration $(IC)_{50}$ value for rucaparib. Thus, PiB competes effectively for binding of rucaparib to human PARP1, an enzyme reported to bind rucaparib with a dissociation constant of 0.1 to 1 nM in biochemical assays (*22, 23*) (Fig. 3A, 3B, 3D, Fig S16, S17, S18). The potency of PiB and PiB' in the cell viability assay generally tracks with the spectroscopic assays, with the protein showing effects on mefuparib and niraparib intermediate between that for rucaparib and veliparib (Fig. 3B, 3D, Fig. S16, S17, S18). Moreover, PiB and PiB' did not appreciably change the cellular response to olaparib (Fig. 3C, 3D, Fig. S16, S17, S18) in line with spectroscopic data that indicated that PiB and PiB' does not bind this drug.

Structural and mutational validation of designs

The crystallographic structures of PiB' were solved in the absence and presence of the four active compounds at 1.3- to 1.6-Å resolution (Table S2). The protein's conformation is in excellent agreement with the predicted AlphaFold2 model, particularly near the binding site (Cα RMSD of the 60 surrounding residues was 0.2 to 0.5 Å, Fig. 2A, 4A-B, S19, Table. S3). A similar degree of agreement (< 0.5 Å RMSD) was observed comparing the structure of the experimental structure and the designed model. An important aspect of the design was that the binding site should be preorganized. Indeed the binding pocket of PiB' is nearly identical between the experimentally determined apo and drug-bound structures (0.2 – 0.5 Å Cα RMSD; Fig. 4B-4C, S20). Moreover, the sidechains interact precisely as predicted in the design of the rucaparib complex (Fig. 4B, S20): Asp29 makes a direct H-bonded salt bridge to the drug's charged ammonium group. Rucaparib's carboxamide forms a two-coordinate hydrogen bond with Gln54, which in turn is stabilized by a second-shell network of H-bonds predicted in the design; Asp131 formed a solvent mediated H-bond to rucaparib's indole NH group (Fig. 4B, S20). A search of water-mediated Asp sidechains

with related indole and imidazole sidechains showed this bridging interaction is frequently found in the PDB (Fig. S21).

The structures of mefuparib and niraprib bound to PiB' show a similar set of interactions as rucaparib (Fig. S22). However, their aromatic 5-membered azole ring lacks a H-bonding group to interact with Asp131, explaining their decreased affinity for the protein. As expected from its divergent structure, veliparib has a less favorable fit with PiB's binding site, and it lacks a salt bridge to its ammonium group as in other complexes (Fig. S22, S23).

Three residues were changed to improve binding during the second round of COMBS design of PiB. To determine whether these substitutions indeed increased affinity, we evaluated mutants with the second-round substitutions reverted to their identities in the first round of design. These changes each led to one to two orders of magnitude weaker binding affinity for rucaparib Asp29Asn (K_D = 13 nM), Leu90Trp (K_D = 24 nM) and Asp131Asn (K_D = 50 nM) (Fig. 4D, S24A). Thus, iterative vdM selection successfully identified interactions with improved binding affinity. We suggest that vdM guided amino acid optimization might provide a useful alternative to other methods of affinity optimization. We also conducted an alanine scan to probe the energetic contribution of each of the residues that lined the pocket in the rucaparib complex (Fig. 4E, S24B). Each mutation was unfavorable with values of $\Delta\Delta G$ ranging from 1.7 to 3.0 kcal/mol. These values are within the range observed for substitutions of critical binding residues in natural protein binding sites (*24*).

Computational prediction of binding thermodynamics

To provide insight into the experimental results for PiB, PiB', PiB1 and PiB2 we next conducted 2.0-microsecond all-atom molecular dynamics (MD) simulations to compare their structural dynamics on this time scale. The simulations were performed on the designed models (instead of the crystallographic structures) to assess the use of MD as a predictor of experimetnal success. The protein backbone conformations were very stable for all four complexes. However, rucaparib's designed binding pose was stable only in PiB and PiB' (Fig. S23) (as PiB and PiB' behave similar in MD, we only use PiB to illustrate later): it retained its bivalent hydrogen bonding interaction to Gln54 (Fig 5A), and Asp29 and Asp131 showed stable interactions with rucaparib's indole NH and ammonium groups through direct and water-mediated hydrogen bonds, respectively (Fig. 5A). By contrast, PiB-1 and PiB-2 simulations deviated from rucaparib's designed pose, and their key buried H-bonds to Gln54 were broken within 50 nanoseconds in each of three independent calculations (Fig. S25). Moreover, PiB shielded the apolar atoms in rucaparib more efficiently in PiB than PiB-1 and PiB-2, as determined from solvent-accessible surface area calculations within individual MD trajectories (Fig. S26). Furthermore, MD simulations of PiB in complex with niraparib, veliparib, and mefuparib show similar binding-site conformational stability as PiB: rucaparib over 2.0 microseconds (Fig. 5A, S27). MD can thus help rationalize how interactions contribute to stability in predicted complexes and may be a useful tool in design.

We next turned to alchemical and physical-pathway methods to determine whether these methods could predict the absolute binding free energies for PiB and PiB' directly from molecular dynamics simulations, using the computational designs (rather than experimental structures) as the starting models. The alchemical transfer method (*25, 26*) was carried out by one of the authors, who had no knowledge of the experimental results. This method has been shown to be comparable to other alchemical methodologies such as Schrodinger's FEP+ (*25*) or Amber's thermodynamic integration (*26*) given comparable sampling of the configurational space. An initial absolute binding free energy calculation was used to evaluate the energetic contributions of the fused-ring cores of the drugs and to ensure convergence of the calculations. An additional relative binding

free energy calculation was performed to transform each core into the target ligand to estimate the contribution from non-core regions (Fig. S28). Universally, the alchemical transfer method tended to overestimate the binding energy, possibly due to having two sets of restraint potentials. However, this procedure correctly predicted the relative affinities of the four ligands (Table. S4).

We next used potential of mean force calculations, an orthogonal physical-pathway methodology (*27*), to compute absolute binding energies (Fig. S29-S33), and the results were in remarkably good agreement with experiment (Fig. 5B, 5C, Table. S5). The RMS error between the predicted and experimental values is 1.3 kcal/mol, and the correct rank order of affinities was observed. This error is close to the experimental error in the measurement of K_D for rucaparib. We also obtained very good agreement between computation and experiment for a set of four mutants of PiB (Fig. 5B, 5C, Table. S5). This agreement bodes well for the use of alchemical and physical-pathway-based binding free energy calculations to evaluate potential binding energies of *de novo* small-molecule binding proteins.

Discussion

These findings show that it is possible to design very high-affinity proteins for drugs using minimal scaffolds such as the four-helix bundle. vdMs and COMBS were key to this achievement. While they were used in conjunction with Rosetta sequence design, they could be easily adapted to improve sampling and/or scoring of a variety of recently developed protein-design algorithms based on diffusion models (*3, 8–10*). MD simulations of *de novo* proteins have only been occasionally used to provide insight into *de novo* protein design (*28–32*). However, using this method we were able to differentiate successful versus unsuccessful designs, suggesting it should be helpful for prioritizing designs. Although we ran 2 microseconds simulations, the essential features could be gleaned after 100 nanoseconds, suggesting that simulations on this time scale should be useful. Free energy calculations have not previously been applied to designed proteins. Although they are more computationally intensive and require more user-specified parameters, we obtained excellent quantitative agreement between computed and experimentally measured binding free energies using the designed models as the starting structures. These data demonstrate the possibility of designing proteins with high affinity (< 5 nM) to small molecules using fully rational criteria for design and "physics-based" force fields to evaluate the complexes.

Although we designed PiB' without considering the natural binding-site interactions in PARP1, very similar interactions were seen in the experimental structures of the two proteins (Fig. S34): a ligand-Asp salt bridge, solvent-mediated hydrogen bond to rucaparib's indole NH, and a two-coordinate H-bond to the drug's amide. This commonality likely reflects the fact that proteins have only a limited repertoire of functional groups, and that COMBS is capable of identifying highly favorable interactions, similar to those used in nature.

Rucaparib binds to the human PARP1 enzyme with a K_D ranging from 0.1 to 1.5 nM, depending on the experimental conditions, close to the range observed for PiB and PiB'(*22, 23*). Ligand efficiency is often used as a guiding rule in drug discovery to determine whether the affinity of a molecule of a given size is within a range typically seen in highly optimized small-molecule drugs and natural organic ligands for proteins (*33, 34*). As ligands become larger, they have more opportunities to form favorable interactions with their target proteins. Thus, the maximal affinity possible roughly scales with the size of a small molecule, and the ligand efficiency is defined as the free energy of binding (1 M standard state) divided by the number of heavy atoms in the ligand. Most drugs have ligand efficiency around 0.3 kcal/(mol * heavy atom count) (*33, 34*), although higher values are observed for highly optimized drugs such as rucaparib, which has a ligand

efficiency of 0.5 kcal/(mol*heavy atom count). The ligand efficiency of a drug is similarly a good measure of how well optimized a *de novo* protein is for binding to a small molecule. The 0.5 kcal/(mol * heavy atom count) ligand efficiency of PiB is a considerable improvement over the 0.21 to 0.26 ligand efficiency of the first COMBS-designed apixaban binders, demonstrating the importance of incorporating the design principles discussed above. With these and similar refinements, it should be increasingly possible to design high-affinity small molecule-binding proteins with predicable binding energetics for a variety of practical applications in sensing and pharmaceuticals.

References and Notes

1. B. Schreier, C. Stumpp, S. Wiesner, B. Höcker, Computational design of ligand binding is not a solved problem. *Proceedings of the National Academy of Sciences* **106**, 18491–18496 (2009).
2. W. Yang, L. Lai, Computational design of ligand-binding proteins. *Curr Opin Struct Biol* **45**, 67–73 (2017).
3. N. F. Polizzi, W. F. DeGrado, A defined structural unit enables *de novo* design of small-molecule–binding proteins. *Science* **369**, 1227–1233 (2020).
4. F. Thomas, W. M. Dawson, E. J. M. Lang, A. J. Burton, G. J. Bartlett, G. G. Rhys, A. J. Mulholland, D. N. Woolfson, *De Novo* -Designed α-Helical Barrels as Receptors for Small Molecules. *ACS Synth Biol* **7**, 1808–1816 (2018).
5. C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, D. Baker, Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
6. J. Dou, L. Doyle, P. Jr. Greisen, A. Schena, H. Park, K. Johnsson, B. L. Stoddard, D. Baker, Sampling and energy evaluation challenges in ligand binding protein design. *Protein Science* **26**, 2426–2437 (2017).
7. J. Dou, A. A. Vorobieva, W. Sheffler, L. A. Doyle, H. Park, M. J. Bick, B. Mao, G. W. Foight, M. Y. Lee, L. A. Gagnon, L. Carter, B. Sankaran, S. Ovchinnikov, E. Marcos, P.-S. Huang, J. C. Vaughan, B. L. Stoddard, D. Baker, *De novo* design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
8. J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, *De novo* design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
9. J. B. Ingraham, M. Baranov, Z. Costello, K. W. Barber, W. Wang, A. Ismail, V. Frappier, D. M. Lord, C. Ng-Thow-Hing, E. R. Van Vlack, S. Tie, V. Xue, S. C. Cowles, A. Leung, J. V. Rodrigues, C. L. Morales-Perez, A. M. Ayoub, R. Green, K. Puentes, F. Oplinger, N. V. Panwar, F. Obermeyer, A. R. Root, A. L. Beam, F. J. Poelwijk, G. Grigoryan, Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023).
10. J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, D. Baker, Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
11. R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, M. Baek, F. DiMaio, D. Baker, Generalized Biomolecular Modeling and Design with RoseTTAFold All-Atom. bioRxiv, 2023.10.09.561603 [Preprint] (2023). https://doi.org/10.1101/2023.10.09.561603
12. G. R. Lee, S. J. Pellock, C. Norn, D. Tischer, J. Dauparas, I. Anischenko, J. A. M. Mercer, A. Kang, A. Bera, H. Nguyen, I. Goreshnik, D. Vafeados, N. Roullier, H. L. Han, B. Coventry, H. K. Haddox, D. R. Liu, A. H.-W. Yeh, D. Baker, Small-molecule binding and

sensing with a designed protein family. bioRxiv, 2023.11.01.565201 [Preprint] (2023). http://biorxiv.org/lookup/doi/10.1101/2023.11.01.565201
13. L. An, M. Said, L. Tran, S. Majumder, I. Goreshnik, G. R. Lee, D. Juergens, J. Dauparas, I. Anishchenko, B. Coventry, A. K. Bera, A. Kang, P. M. Levine, V. Alvarez, A. Pillai, C. Norn, D. Feldman, D. Zorine, D. R. Hicks, X. Li, M. G. Sanchez, D. K. Vafeados, P. J. Salveson, A. A. Vorobieva, D. Baker, *De novo* design of diverse small molecule binders and sensors using Shape Complementary Pseudocycles. bioRxiv, 2023.12.20.572602 [Preprint] (2023). https://doi.org/10.1101/2023.12.20.572602
14. J. Dauparas, G. R. Lee, R. Pecoraro, L. An, I. Anishchenko, C. Glasscock, D. Baker, Atomic context-conditioned protein sequence design using LigandMPNN. *bioRxiv*, 2023.12.22.573103 [Preprint] (2023). https://doi.org/10.1101/2023.12.22.573103
15. P. Gainza, F. Sverrisson, F. Monti, E. Rodolà, D. Boscaini, M. M. Bronstein, B. E. Correia, Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods* **17**, 184–192 (2020).
16. M. V. Shapovalov, R. L. Dunbrack, A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **19**, 844–858 (2011).
17. C. J. Lord, A. Ashworth, PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
18. A. Koehl, M. Jagota, D. D. Erdmann-Pham, A. Fung, Y. S. Song, Transferability of Geometric Patterns from Protein Self-Interactions to Protein-Ligand Interactions. *Pac Symp Biocomput* **27**, 22–33 (2022).
19. S. I. Mann, A. Nayak, G. T. Gassner, M. J. Therien, W. F. DeGrado, *De Novo* Design, Solution Characterization, and Crystallographic Structure of an Abiological Mn–Porphyrin-Binding Protein Capable of Stabilizing a Mn(V) Species. *J Am Chem Soc* **143**, 252–259 (2021).
20. D.-A. Silva, S. Yu, U. Y. Ulge, J. B. Spangler, K. M. Jude, C. Labão-Almeida, L. R. Ali, A. Quijano-Rubio, M. Ruterbusch, I. Leung, T. Biary, S. J. Crowley, E. Marcos, C. D. Walkey, B. D. Weitzner, F. Pardo-Avila, J. Castellanos, L. Carter, L. Stewart, S. R. Riddell, M. Pepper, G. J. L. Bernardes, M. Dougan, K. C. Garcia, D. Baker, *De novo* design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
21. T. Hucl, C. Rago, E. Gallmeier, J. R. Brody, M. Gorospe, S. E. Kern, A Syngeneic Variance Library for Functional Annotation of Human Variation: Application to *BRCA2*. *Cancer Res* **68**, 5023–5030 (2008).
22. J. Rudolph, G. Roberts, K. Luger, Histone Parylation factor 1 contributes to the inhibition of PARP1 by cancer drugs. *Nat Commun* **12**, 736 (2021).
23. H. D. Thomas, C. R. Calabrese, M. A. Batey, S. Canan, Z. Hostomsky, S. Kyle, K. A. Maegley, D. R. Newell, D. Skalitzky, L.-Z. Wang, S. E. Webber, N. J. Curtin, Preclinical selection of a novel poly(ADP-ribose) polymerase inhibitor for clinical trial. *Mol Cancer Ther* **6**, 945–956 (2007).
24. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Macmillan, 1999).
25. L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel,

Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J Am Chem Soc* **137**, 2695–2703 (2015).
26. T.-S. Lee, B. K. Allen, T. J. Giese, Z. Guo, P. Li, C. Lin, T. D. McGee, D. A. Pearlman, B. K. Radak, Y. Tao, H.-C. Tsai, H. Xu, W. Sherman, D. M. York, Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J Chem Inf Model* **60**, 5595–5623 (2020).
27. C. Chipot, Free energy methods for the description of molecular processes. *Annu Rev Biophys* **52**, 113–138 (2023).
28. E. P. Barros, J. M. Schiffer, A. Vorobieva, J. Dou, D. Baker, R. E. Amaro, Improving the Efficiency of Ligand-Binding Protein Design with Molecular Dynamics Simulations. *J Chem Theory Comput* **15**, 5703–5715 (2019).
29. G. Ulas, T. Lemmin, Y. Wu, G. T. Gassner, W. F. DeGrado, Designed metalloprotein stabilizes a semiquinone radical. *Nat Chem* **8**, 354–359 (2016).
30. M. Gill, M. E. McCully, Molecular dynamics simulations suggest stabilizing mutations in a *de novo* designed α/β protein. *Protein Engineering, Design and Selection* **32**, 317–329 (2019).
31. A. Chevalier, D.-A. Silva, G. J. Rocklin, D. R. Hicks, R. Vergara, P. Murapa, S. M. Bernard, L. Zhang, K.-H. Lam, G. Yao, C. D. Bahl, S.-I. Miyashita, I. Goreshnik, J. T. Fuller, M. T. Koday, C. M. Jenkins, T. Colvin, L. Carter, A. Bohn, C. M. Bryan, D. A. Fernández-Velasco, L. Stewart, M. Dong, X. Huang, R. Jin, I. A. Wilson, D. H. Fuller, D. Baker, Massively parallel *de novo* protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).
32. M. C. Childers, V. Daggett, Insights from molecular dynamics simulations for computational protein design. *Mol Syst Des Eng* **2**, 9–33 (2017).
33. A. L. Hopkins, G. M. Keserü, P. D. Leeson, D. C. Rees, C. H. Reynolds, The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* **13**, 105–121 (2014).
34. I. D. Kuntz, K. Chen, K. A. Sharp, P. A. Kollman, The maximal affinity of ligands. *Proceedings of the National Academy of Sciences* **96**, 9997–10002 (1999).
35. L. Lu, *De novo* design of drug-binding proteins with predictable binding energy and specificity, *Zenodo* (2024). https://doi.org/https://doi.org/10.5281/zenodo.10653015.
36. G. Grigoryan, W. F. DeGrado, Probing Designability via a Generalized Model of Helical Bundle Geometry. *J Mol Biol* **405**, 1079–1100 (2011).
37. J. Zhou, G. Grigoryan, Rapid search for tertiary fragments reveals protein sequence–structure relationships. *Protein Science* **24**, 508–524 (2015).
38. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
39. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).
40. R. Wu, F. Ding, R. Wang, R. Shen, X. Zhang, S. Luo, C. Su, Z. Wu, Q. Xie, B. Berger, High-resolution *de novo* structure prediction from primary sequence. *BioRxiv*, 2022–2027 [preprint] (2022). https://doi.org/10.1101/2022.07.21.500999

41. M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. A. Hochberg, J. L. P. Benesch, C. V. Robinson, Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal Chem* **87**, 4370–4376 (2015).
42. J.-X. He, M. Wang, X.-J. Huan, C.-H. Chen, S.-S. Song, Y.-Q. Wang, X.-M. Liao, C. Tan, Q. He, L.-J. Tong, Y.-T. Wang, X.-H. Li, Y. Su, Y.-Y. Shen, Y.-M. Sun, X.-Y. Yang, Y. Chen, Z.-W. Gao, X.-Y. Chen, B. Xiong, X.-L. Lu, J. Ding, C.-H. Yang, Z.-H. Miao, Novel PARP1/2 inhibitor mefuparib hydrochloride elicits potent *in vitro* and *in vivo* anticancer activity, characteristic of high tissue distribution. *Oncotarget* **8**, 4156–4168 (2017).
43. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden,R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang,S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J.Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R.Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. *Amber 2018* (University of California, San Francisco, 2018).
44. R. Salomon-Ferrer, D. A. Case, R. C. Walker, An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci* **3**, 198–210 (2013).
45. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. *Gaussian 09, Revision A.02* (Gaussian, Inc., Wallingford CT, 2016).
46. U. C. Singh, P. A. Kollman, An approach to computing electrostatic charges for molecules. *J Comput Chem* **5**, 129–145 (1984).
47. B. H. Besler, K. M. Merz, P. A. Kollman, Atomic charges derived from semiempirical methods. *J Comput Chem* **11**, 431–439 (1990).
48. J. Wang, W. Wang, P. A. Kollman, D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **25**, 247–260 (2006).
49. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. *J Comput Chem* **25**, 1157–1174 (2004).
50. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935 (1983).
51. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713 (2015).
52. J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, B. O. Brandsdal, Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem Phys Lett* **384**, 288–294 (2004).

53. J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **23**, 327–341 (1977).
54. S. Miyamoto, P. A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* **13**, 952–962 (1992).
55. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* **98**, 10089–10092 (1993).
56. R. Gowers, M. Linke, J. Barnoud, T. Reddy, M. Melo, S. Seyler, J. Domański, D. Dotson, S. Buchoux, I. Kenney, O. Beckstein, "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations" in *Proceedings of the 15th Python in Science Conference* (SciPy Austin, TX, 2016; https://conference.scipy.org/proceedings/scipy2016/oliver_beckstein.html)vol. 98, pp. 98–105.
57. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319–2327 (2011).
58. A. Bakan, L. M. Meireles, I. Bahar, *ProDy* : Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **27**, 1575–1577 (2011).
59. W. Kabsch, *XDS*. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132 (2010).
60. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *Phaser* crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
61. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).
62. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
63. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, *electronic Ligand Builder and Optimization Workbench* (*eLBOW*): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074–1080 (2009).
64. A. Shrake, J. A. Rupley, Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79**, 351–371 (1973).
65. J.-D. Chai, M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics* **10**, 6615 (2008).
66. J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, T. D. Crawford, Psi4: an open-source *ab initio* electronic structure program. *Wiley Interdiscip Rev Comput Mol Sci* **2**, 556–565 (2012).
67. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* **132** (2010).
68. S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **32**, 1456–1465 (2011).

69. C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* **97**, 10269–10280 (1993).
70. A. Alenaizan, L. A. Burns, C. D. Sherrill, Python implementation of the restrained electrostatic potential charge model. *Int J Quantum Chem* **120**, e26035 (2020).
71. S. Azimi, S. Khuttan, J. Z. Wu, R. K. Pal, E. Gallicchio, Relative Binding Free Energy Calculations for Ligands with Diverse Scaffolds with the Alchemical Transfer Method. *J Chem Inf Model* **62**, 309–323 (2022).
72. J. Z. Wu, S. Azimi, S. Khuttan, N. Deng, E. Gallicchio, Alchemical Transfer Approach to Absolute Binding Free Energy Estimation. *J Chem Theory Comput* **17**, 3309–3319 (2021).
73. E. Gallicchio, J. Xia, W. F. Flynn, B. Zhang, S. Samlalsingh, A. Mentes, R. M. Levy, Asynchronous replica exchange software for grid and heterogeneous computing. *Comput Phys Commun* **196**, 236–246 (2015).
74. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzesh, G. Giambaşu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, Q. Zhu, and P.A. Kollman. *Amber 2023* (University of California, San Francisco, 2023).
75. X. He, V. H. Man, W. Yang, T.-S. Lee, J. Wang, A fast and high-quality charge model for the next generation general AMBER force field. *J Chem Phys* **153** (2020).
76. P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, V. S. Pande, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).
77. S. Khuttan, S. Azimi, J. Z. Wu, E. Gallicchio, Alchemical transformations for concerted hydration free energy estimation with explicit solvation. *J Chem Phys* **154** (2021).
78. T.-S. Lee, B. K. Allen, T. J. Giese, Z. Guo, P. Li, C. Lin, T. D. McGee, D. A. Pearlman, B. K. Radak, Y. Tao, H.-C. Tsai, H. Xu, W. Sherman, D. M. York, Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J Chem Inf Model* **60**, 5595–5623 (2020).
79. Z. Tan, E. Gallicchio, M. Lapelosa, R. M. Levy, Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J Chem Phys* **136** (2012).
80. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
81. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J Chem Phys* **126** (2007).
82. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*, NA-NA (2009).

83. A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J Chem Theory Comput* **8**, 527–541 (2012).
84. Y. Zhao, D. G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* **120**, 215–241 (2008).
85. T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J Chem Phys* **90**, 1007–1023 (1989).
86. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. *Gaussian 16, Revision A.03* (Gaussian, Inc., Wallingford CT, 2016).
87. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh Ewald method. *J Chem Phys* **103**, 8577–8593 (1995).
88. W. You, Z. Tang, C. A. Chang, Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J Chem Theory Comput* **15**, 2433–2443 (2019).
89. J. A. Lemkul, D. R. Bevan, Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics. *J Phys Chem B* **114**, 1652–1660 (2010).
90. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52**, 7182–7190 (1981).
91. G. M. Torrie, J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* **23**, 187–199 (1977).
92. J. S. Hub, B. L. de Groot, D. van der Spoel, g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J Chem Theory Comput* **6**, 3713–3720 (2010).
93. J. K. Myers, C. N. Pace, J. M. Scholtz, Helix Propensities Are Identical in Proteins and Peptides. *Biochemistry* **36**, 10923–10929 (1997).
94. K. T. O'Neil, W. F. DeGrado, A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science* **250**, 646–651 (1990).

Acknowledgments
We thank members of the DeGrado lab for useful discussions. Beamline 8.3.1 of the Advanced Light Source, a DOE Office of Science User Facility under Contract No. DE-AC02-05CH11231, is supported in part by the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01. We are grateful to George Meigs from LBNL helping with the setup.

Funding
We acknowledge research support from grant from NIH (2R35GM122603). N.P. acknowledges support from NIH grant R00GM135519.

Author Contributions
L.L. and W.F.D conceived the idea for the project, L.L. and N.P. developed computer code, and L.L. performed the computational design. L.L., X.G., S.I.M., X.Z., H.J. performed the experiments. L.L., H.Y., X.G., Y.W. performed the data analysis. N.P., S.T., D.G., J.V. ran and analyzed MD stimulations. X.G., M.D., A.A., and W.F.D. designed the cell assay experiments. L.L., N.P., and W.F.D. wrote the paper with input from all authors.

Competing financial interests
A.A. is a co-founder of Tango Therapeutics, Azkarra Therapeutics, Ovibio Corporation and Kytarro, a member of the board of Cytomx and Cambridge Science Corporation, a member of the scientific advisory board of Genentech, GLAdiator, Circle, Bluestar, Earli, Ambagon, Phoenix Molecular Designs, Yingli, ProRavel, Oric, Hap10 and Trial Library, a consultant for SPARC, ProLynx, Novartis and GSK, receives research support from SPARC, and holds patents on the use of PARP inhibitors held jointly with AstraZeneca from which he has benefited financially (and may do so in the future).

Data and materials availability
Coordinates and structure files have been deposited to the Protein Data Bank (PDB) with accession codes: 8TN1 (apo-PiB), 8TN6 (rucaparib-bound PiB), 8TNB (mefuparib-bound PiB), 8TNC (niraparib-bound PiB), 8TND (veliparib-bound PiB). Computational code and design scripts are available in the supplementary materials and at Zenodo (*35*). All other relevant data are available in the main text or the supplementary materials.

Supplementary Materials
Materials and Methods
Figs. S1 to S34
Tables S1 to S5
Supplementary Text
References (36-94)
MDAR Reproducibility Checklist
Data S1

Figures

Figure 1. The computational design of poly(ADP-ribose) polymerase inhibitors (PARPi) protein binders. (A) The PARPi analogues. The shared chemical features are marked in orange. Olaparib is used as a negative control in the design and binding assay. (B – E) The overall design strategy. (B, C) We first define the pharmacophore and use COMBS to sample vdMs on the selected protein backbones. At the outset of the design, we chose chemical groups that should form hydrogen bonds when the drug bound to the binding site. These groups included rucaparib's indole NH and carboxamide groups. The carboxamide group is present in our vdM library. However, there were relatively few examples of indole NH vdMs in the database, so we used imidazole as a proxy for the indole's pyrole ring. We then used COMBS to discover sidechains at different positions of a four-helix bundle template that could simultaneously form hydrogen bonds to the indole and carboxamide chemical groups of the drug. (In brief, the COMBS algorithm samples vdMs on a protein backbone and then performs superpositions of a ligand onto the chemical groups of the vdMs; next, COMBS finds all the vdMs with nearby chemical groups to each superposed ligand; and finally, COMBS computes a unique combination of vdMs for each ligand that optimizes a score, such as the vdM prevalence or cluster score). We discovered a solution in which the carboxamide formed bidentate hydrogen bonds with sidechain of Gln54, and the drug's indole NH interacted with the Asn131 (C, carbon atoms of protein green, those of rucaparib are orange). A second-shell interaction to Q54 that was discovered by COMBS was Asp58 (carbons brown). (D) We applied flexible backbone sequence design with a custom Rosetta script while fixing the interactions selected from COMBS. (Rucaparib in purple). (E) Then we search vdM again based on the design output from the previous sequence step. The slightly different (~ 1 Å Ca RMSD) backbone now preferred different vdMs at some locations (higher cluster scores) and these mutations were made. Three residues at 29, 90, 131 (deep blue) were changed based on COMBS results.

Figure 2. Assessing the computational model and experimental binding of PiB to rucaparib. (A) The AlphaFold2 model agrees with the designed PiB very well, with the binding site Cα RMSD of 0.41 Å, the upper fold Cα RMSD of 0.49 Å and overall Cα RMSD of 0.67 Å. (B) The predicted local distance difference test scores (pLDDTs) concur with the trend of RMSD difference of the design model. For example, the N-terminal, C-terminal and the middle loop with low pLDDTs (<90) showed higher Cα RMSD. (C) The design model showing the polar groups of rucaparib are all hydrogen-bonded. (D) (E) A fluorescence titration shows that PiB and PiB' bind rucaparib with K_D < 5 nM. The fluorescence emission intensity at 420 nm of rucaparib (excitation wavelength 355 nm) was measured after titrating aliquots of PiB (D) or PiB' (E) to a final concentration indicated in the abscissa. The data are well described by a single-site protein-ligand binding model, and a non-linear least squares fit to the data returned values of K_D of 2.2 (± 0.9) nM for PIB, and 0.37 (± 0.29) nM for PIB'. Although the fitting error was relatively small, a sensitivity analysis, in which the value of K_D was held constant at various values, showed that the data for both proteins were fit within experimental error so long as the K_D is less than 5 nM. Therefore, while the most probable binding constants were 2 and 0.4 nM, respectively, we can confidently conclude that the values for PiB and PiB' are less than 5 nM. The titration was carried out in in buffer containing 50 mM Tris, 100 mM NaCl (pH 7.4).

Figure 3. Spectral titrations and cell viability assay of PiB with PARPi. (A) The values of K_D of various drugs for PiB as obtained from global fit of a single-site binding model to the fluorescence changes (A, from Fig. 3) or absorbance changes as a function of the concentration of PiB. Indicated wavelengths for the titration were chosen to maximize the difference in absorption for the free versus bound drug. (B) Seven-day growth assays in DLD-1 *BRCA2* mutated cells show that PiB alleviates the effects of rucaparib, mefuparib, niraparib and veliparib toxicity in a dose-dependent manner. The PARP inhibitors were pre-incubated with PiB in media at room temperature for 5 minutes at multiple concentration ratios (ligand : protein) of 1:0, 1:0.2, 1:1, 1:2.5, 1:5 and 1:10. (C) Cell viability assay as in Figure 4B showing that PiB had no effect on the olaparib dose response. (D) Table showing IC_{50} values for the inhibition of cell proliferation by PARPi drugs in the presence of increasing mole ratios of added PiB protein.

Figure 4. The structure of drug-bound PiB' agrees with the design. (A) The design model agrees well with the rucaparib-bound PiB' crystal structure, with binding site (Fig. 3A) Cα RMSD range between 0.38-0.46 Å for the three monomers in the asymmetric unit. (B) The binding site of PiB'. A 2mFo-DFc composite omit map contoured at 1.6 σ. The map was generated from a model that omitted coordinates of rucaparib. Overlay of the design (gray) and the structure (protein in orange, rucaparib in pink). The sidechains of the binding pocket in rucaparib-bound PiB' agrees with the design. Asp131 interacts with the indole NH via a bridging water as in MD simulations. (C) The structure of apo-PiB' shows a preorganized open pocket filled with multiple waters, which are displaced in holo structure. (D) Reversal of the three designed substitutions from the vdM optimization procedure led to lower binding affinity (higher K_D) for rucaparib by fluorescence emission titrations. (E) Alanine mutations of the direct binding residues decreased binding affinities confirmed by fluorescence emission titrations.

Figure 5. The MD simulations of PiB, PiB', and mutants. (A) Using unbiased molecular dynamics simulations in Amber, we calculated (in triplicate) the frequency with which the intermolecular hydrogen bonds formed between the protein scaffold and the bound drug molecule. PiB was found to form a hydrogen bond between Gln54 and the targeted drug carboxamide in 100% of all simulations for each drug complex. The charged ammonium groups of rucaparib and mefuparib interacted with Asp29 through a combination of direct and water-mediated hydrogen bonds, totaling to more than half of the full simulation time, which contrasts niraparib and veliparib's inabilities to form equivalent hydrogen bonds (due to changes in chemical structure around the ammonium tail of the ligand). In a small fraction of each rucaparib and veliparib trajectory, Asp131 engaged in water-mediated hydrogen bonds to the drugs. (B) Using biased simulations in GROMACS, we calculated binding free energies for each ligand and found that ranked affinity for each drug is consistent with experimental results. (C) Comparison of ΔG binding from the GROMACS calculation with the experimental value from spectral titrations.