

**Title:**

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

**Authors:**

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

**Affiliations:**

<sup>1</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94158, USA.

<sup>2</sup>Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA 94158, USA.

<sup>3</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA.

<sup>4</sup>Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA.

\*Corresponding author. Email: [william.degrado@ucsf.edu](mailto:william.degrado@ucsf.edu)

## Abstract

The de novo design of small-molecule-binding proteins has seen exciting recent progress; however, the ability to achieve exquisite affinity for binding small molecules while tuning specificity has not yet been demonstrated directly from computation. Here, we develop a computational procedure that results in the highest affinity binders to date with predetermined relative affinities, targeting a series of PARP1 inhibitors. Two of four designed proteins bound with affinities ranging from  $< 5$  nM to low  $\mu$ M, in a predictable manner. 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, suggesting that the de novo design of small-molecule-binding proteins with tuned interaction energies is now feasible entirely from computation. We expect these methods to open many opportunities in biomedicine, including rapid sensor development, antidote design, and drug delivery vehicles.

## One Sentence Summary

We use informatic sampling to design low nM drug-binding proteins, and physics-based calculations to accurately predict affinities.

## Main Text

Molecular recognition underlies the binding and catalysis of small molecules by protein receptors and enzymes (1–5). While 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 that has not been achieved in *de novo* proteins without experimental screening of large libraries of variants (4–12). The few successful examples of micromolar binders have relied on structural informatics to guide sampling of protein structure, sequence, and interactions, and scoring functions that rely on a mix of statistical and physical terms without explicit representation of dynamics, conformational entropy or water (6–11). From a practical perspective, the reliance on imperfect empirical and statistical information has limited the success of the process, and more fundamentally leaves open the question of whether our understanding is grounded in physical forces or limited exclusively to advanced pattern recognition (13–17). Here, we show how fully physics-based methods can be used to evaluate small molecule binders. We also introduce enhancements on existing sampling methods to better account for molecular complementarity between the ligand and protein. Finally, we show that all-atom molecular dynamics simulations are sufficient to describe the conformations, dynamics, binding interactions and free energies of association of a *de novo* protein designed to bind a series of poly(ADP-ribose) polymerase (PARP) inhibitors (PARPi) (18). These studies show the feasibility of combining informatics-based approaches for sampling and physics-based approaches for evaluation of top-scoring designs.

Inhibitors of PARP are a recently developed class of clinically useful anticancer drugs. *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 new recursive version of the COMBS algorithm to design rucaparib-binding sites in a family of mathematically generated four-helix bundle proteins into which key binding residues were introduced using van der Mers (vdMs) (10, 19, 20). A vdM is an element of protein structure that identifies the preferred positions of interacting chemical groups relative to a residue's backbone atoms (10). The COMBS algorithm finds 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 (Fig. 1B, 1C). To initiate the design process, we used COMBS to identify sidechains to bind rucaparib's key polar groups, which included H-bonded interactions involving the indole NH, and the carboxamide's C=O and NH<sub>2</sub> groups (Fig. 1C). Additionally, COMBS identified Asp131 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 flexible backbone design (Fig. 1D) (10, 21, 22) 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- $\pi$  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, 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 significantly from that found in the other three drugs.

The final models were chosen based on multiple criteria: 1) low Rosetta energy (lowest 50 of the 1000 total designs); 2) favorable vdMs (highest total vdM cluster scores); 3) satisfaction of all buried H-bond donors in the protein and ligand; and 4) avoidance of clashes with the three other PARP inhibitors, which show significant structural variability near the amine end of the molecule.

Four designs were selected for expression: two designated as PiB (PARPi binder) and PiB', respectively, are closely related (Table. S1, Fig. S3). PiB' differs from PiB only by the substitution of five solvent-exposed charged residues with Ala to enhance the crystallinity of the protein for X-ray crystallography. The other two (PiB-1 and PiB-2) were less closely related to PiB in structure and sequence (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) (10). Thus, PiB and PiB' are the first reported *de novo* protein that bind a small molecule with very low, single digit nM to pM binding affinity prior to extensive experimental optimization.

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- $\mu$ M 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  $\mu$ M and 24  $\mu$ M, 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 our hypothesized order.

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) (10, 23, 24). PARP inhibitors potently inhibit the viability of cells with certain DNA repair deficiencies, including loss-of-function mutations in *BRCA2*. 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 (25) 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)<sub>50</sub> 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 (26, 27) (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 (26) (Fig. 3B, 3D, Fig. S16, S17, S18). Moreover, PiB and PiB' did not significantly 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.

### High-resolution crystal structures of complexes are in remarkable agreement with the designs.

The crystallographic structures of PiB' were solved in the absence and presence of the four active compounds at 1.3-1.6 Å resolution (Table S2). The protein's conformation is in excellent agreement with the predicted AlphaFold2 model (Table. S3), particularly near the binding site (Cα RMSD of the 60 surrounding residues was 0.2- 0.5 Å, Fig. 2A, 4A-B, S19, Table. S3). The binding pocket of PiB' is very well pre-organized, and the binding of the drugs leads to only 0.2 – 0.5 Å Cα RMSD changes in the binding-site residues (Fig. 4A-4C, S20, Table. S3). The conformations of the sidechains in the unbound structure are almost entirely as in the design, and they 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). Interestingly, 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 niraparib 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). In summary, the structures are in excellent agreement with the design and confirm our structural prediction of the affinity differences between drugs.

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 (28).

### **Molecular dynamics and free energy calculations confirm mode of binding and accurately predict binding thermodynamics.**

We performed 2.0-microsecond all-atom molecular dynamics (MD) simulations to compare the structural stability of PiB, PiB', PiB1, and PiB2 in complex with rucaparib. The simulations were performed on the designed models (instead of the crystallographic structures) to assess the usefulness of incorporating MD into a design pipeline. The protein backbone conformations were very stable for all three 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 exhibited significant deviation 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 PiB1 and PiB2, as determined from solvent-accessible surface area calculations (29) within individual MD trajectories (Fig. S26). Furthermore, MD simulations of PiB in complex with niraparib, veliparib, and mefuparib show similar binding-site stability as PiB : rucaparib over 2.0 microseconds (Fig. 5A, S27). Thus, MD appears to be a useful tool in assessing the stability of interactions in designed complexes.

We next turned to alchemical and physical-pathway methods to calculate absolute binding free energies to determine whether the dissociation constants for binding of the drugs to PiB and PiB' might be predicted directly from molecular dynamics simulations. The alchemical transfer method (29–31) was carried out starting with the design models by two 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+ (30) or Amber's thermodynamic integration (31) 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 (32), to compute absolute binding energies, and found that the results were in remarkably good agreement with experiment (Fig. S29). 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 (Fig. 5B, 5C, Fig.

S30-S33, Table. S5). 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.

## Conclusion

The *de novo* design of proteins with very high affinity for small-molecule ligands has been a challenging endeavor, because the design process generally relies on sampling and scoring functions that are optimized for predicting protein structure rather than binding to the diversity of functionality encountered in small molecules. In this work we enhanced the COMBS method, which uses van der Mers to sample realistic binding poses and key interactions within a protein framework. In this procedure, vdMs are used to identify binding-site residues capable of forming key interactions such as hydrogen bonds and aromatic interactions to enable binding of a 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 optimal binding remain relatively constant. Thus, a common set of vdMs should serve to bind a wide range of compounds.

A number of other methods for designing ligand-binding proteins are appearing in the recent literature (12). However, at the end of any design procedure, the designer is presented with a list of candidates for further consideration. With sufficient resources it is feasible to synthesize genes encoding thousands of candidates and evaluate them individually using robotics or yeast display to identify a few binders. Additionally, because so many variants are screened, the characterization of the binders has relied on organic synthesis of fluorescently labeled or biotinylated versions of the ligand, again increasing the barriers to widespread economical use of computational design.

Here we identify a series of physically grounded steps to help identify the tightest binders in a computational library of potential proteins (Fig. S34). 1) We prioritized the principle of preorganization: AlphaFold2 predicted a preorganized site, and we find greater discrimination using all-atom MD simulations. Even in the absence of the drug, the conformation of the backbone and binding-site residues of PiB were stable on the nanosecond/microsecond timescale. While we conducted relatively long 2  $\mu$ sec simulations, we found that 100 nsec simulations were sufficient to identify the successful binder. 2) We focus on designs that maximize burial of apolar atoms based on calculation of solvent-accessible surface area in the bound versus free state. 3) Importantly, the H-bond potential of buried polar atoms should be maximized for both the bound drug as well as the interacting sidechains, as the loss of water interactions have to be recouped by drug–protein H-bonding, a highly unfavorable process (33). This is apparent if we compare PiB to ABLE, the first COMBS-designed drug binder (10). The PiB:rucaparib (app. 1 nM) binding site maximizes the H-bonding potential of both the drug and the first-shell liganding sidechains, whereas the ABLE:apixaban (5  $\mu$ M) binding site has four unsatisfied hydrogen-bonding sites (Fig. S35). Water-bridging H-bonds can also be identified by vdMs (that include explicit bridging waters) and MD simulations. 4) We placed the charged ammonium group of rucaparib and interacting sidechains at a surface location that minimizes the Born solvation energy (34), simplifying the task of recognizing charged groups.

Finally, while MD simulations of *de novo* proteins (14, 34–36) can help screen designs (14, 34–36), 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 for the first time 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 complex.

It is interesting to compare the mode of interaction and interaction energy of rucaparib for PiB versus the natural protein PARP1. Although we designed PiB without considering the natural binding-site interactions in RARP1, COMBS identified a very similar set of interactions to bind rucaparib (Fig. S36): 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 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, within the range observed for PiB and PiB’(26, 27). Rucaparib is a third-generation drug that was discovered only after synthesizing hundreds of analogues in multiple groups. Thus, it is noteworthy that we were able to design a protein with similar affinity to rucaparib as PARP1 after screening only a few proteins. 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 developed small-molecule drugs and natural organic ligands for proteins (37, 38). 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) (37, 38), 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 an considerable improvement over the 0.21 to 0.26 ligand efficiency of the first COMBS-designed apixaban binders, demonstrating the significance of incorporating the design principles discussed.

In nature, small molecule-binding proteins serve a range of functions, and we might expect a similarly broad spectrum of practical applications of designed binders for sensing and pharmaceutical applications. For example, *de novo* proteins might be used as antidotes to neutralize drugs that have accumulated to toxic levels, or to reverse the action of anti-coagulants prior to surgery to reverse the risk of bleeding. In other cases, proteins could be used as drug carriers to tune their pharmacokinetic properties. Drugs might also be released in the vicinity of tumors or sites of infection by introducing proteolytic sites into flexible loops that allow cleavage by proteases enriched in these environments (39). Furthermore, *de novo* proteins might be targeted to tumor environments by fusing sequences (e.g., disulfide-linked cyclic RGD peptides (40) or peptide hormones such as somatostatin (41) to bind to proteins enriched on the surface of cancer cells. Finally, designed proteins can be used to sequester environmental toxins or as components in detection devices.

## References and Notes

1. J. K. Lassila, H. K. Privett, B. D. Allen, S. L. Mayo, Combinatorial methods for small-molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences* **103**, 16710–16715 (2006).
2. D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
3. H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences* **109**, 3790–3795 (2012).
4. W. Yang, L. Lai, Computational design of ligand-binding proteins. *Curr Opin Struct Biol* **45**, 67–73 (2017).
5. B. Schreier, C. Stumpf, 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).
6. N. F. Polizzi, Y. Wu, T. Lemmin, A. M. Maxwell, S.-Q. Zhang, J. Rawson, D. N. Beratan, M. J. Therien, W. F. DeGrado, De novo design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy. *Nat Chem* **9**, 1157–1164 (2017).
7. 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).
8. 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).
9. 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).
10. N. F. Polizzi, W. F. DeGrado, A defined structural unit enables de novo design of small-molecule–binding proteins. *Science* **369**, 1227–1233 (2020).
11. 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).
12. 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 (2023).
13. M. C. Childers, V. Daggett, Insights from molecular dynamics simulations for computational protein design. *Mol Syst Des Eng* **2**, 9–33 (2017).
14. 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).

15. Y. Mou, P.-S. Huang, L. M. Thomas, S. L. Mayo, Using Molecular Dynamics Simulations as an Aid in the Prediction of Domain Swapping of Computationally Designed Protein Variants. *J Mol Biol* **427**, 2697–2706 (2015).
16. 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).
17. G. Kiss, D. Röthlisberger, D. Baker, K. N. Houk, Evaluation and ranking of enzyme designs. *Protein Science* **19**, 1760–1773 (2010).
18. C. J. Lord, A. Ashworth, PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
19. P.-S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
20. G. Grigoryan, W. F. DeGrado, Probing Designability via a Generalized Model of Helical Bundle Geometry. *J Mol Biol* **405**, 1079–1100 (2011).
21. B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302**, 1364–1368 (2003).
22. 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).
23. L. Cao, I. Goreshnik, B. Coventry, J. B. Case, L. Miller, L. Kozodoy, R. E. Chen, L. Carter, A. C. Walls, Y.-J. Park, E.-M. Strauch, L. Stewart, M. S. Diamond, D. Veesler, D. Baker, *De novo* design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
24. 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).
25. 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).
26. J. Rudolph, G. Roberts, K. Luger, Histone Parylation factor 1 contributes to the inhibition of PARP1 by cancer drugs. *Nat Commun* **12**, 736 (2021).
27. 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).
28. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Macmillan, 1999).
29. A. Shrake, J. A. Rupley, Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79**, 351–371 (1973).
30. 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).

31. 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).

32. C. Chipot, Free energy methods for the description of molecular processes. *Annu Rev Biophys* **52**, 113–138 (2023).

33. P. W. Kenny, Hydrogen-Bond Donors in Drug Design. *J Med Chem* **65**, 14261–14275 (2022).

34. G. Ulas, T. Lemmin, Y. Wu, G. T. Gassner, W. F. DeGrado, Designed metalloprotein stabilizes a semiquinone radical. *Nat Chem* **8**, 354–359 (2016).

35. M. Gill, M. E. McCully, Molecular dynamics simulations suggest stabilizing mutations in a de novo designed  $\alpha/\beta$  protein. *Protein Engineering, Design and Selection* **32**, 317–329 (2019).

36. 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).

37. 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).

38. 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).

39. J. Vandooren, G. Opdenakker, P. M. Loadman, D. R. Edwards, Proteases in cancer drug delivery. *Adv Drug Deliv Rev* **97**, 144–155 (2016).

40. E. Ruoslahti, RGD AND OTHER RECOGNITION SEQUENCES FOR INTEGRINS. *Annu Rev Cell Dev Biol* **12**, 697–715 (1996).

41. C. Susini, L. Buscail, Rationale for the use of somatostatin analogs as antitumor agents. *Annals of Oncology* **17**, 1733–1742 (2006).

42. J. Zhou, G. Grigoryan, Rapid search for tertiary fragments reveals protein sequence–structure relationships. *Protein Science* **24**, 508–524 (2015).

43. 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).

44. 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).

45. 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 (2022).

46. 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).
47. 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).
48. 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. Salomon-Ferrer, 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).
49. 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).
50. 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).
51. U. C. Singh, P. A. Kollman, An approach to computing electrostatic charges for molecules. *J Comput Chem* **5**, 129–145 (1984).
52. B. H. Besler, K. M. Merz, P. A. Kollman, Atomic charges derived from semiempirical methods. *J Comput Chem* **11**, 431–439 (1990).
53. 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).
54. 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).
55. 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).
56. 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).
57. 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).

58. 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).
59. 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).
60. 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).
61. 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](https://conference.scipy.org/proceedings/scipy2016/oliver_beckstein.html)) vol. 98, pp. 98–105.
62. 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).
63. A. Bakan, L. M. Meireles, I. Bahar, *ProDy* : Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **27**, 1575–1577 (2011).
64. W. Kabsch, *XDS. Acta Crystallogr D Biol Crystallogr* **66**, 125–132 (2010).
65. 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).
66. P. D. Adams, P. V. Afonine, G. Bunkóczki, 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).
67. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
68. 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).
69. 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).
70. 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).
71. 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).
72. S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **32**, 1456–1465 (2011).
73. 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).

74. A. Alenaizan, L. A. Burns, C. D. Sherrill, Python implementation of the restrained electrostatic potential charge model. *Int J Quantum Chem* **120**, e26035 (2020).
75. 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).
76. 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).
77. 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).
78. 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).
79. 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).
80. 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).
81. 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).
82. 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).
83. 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).
84. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J Chem Phys* **126** (2007).
85. 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).
86. 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).
87. 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).

88. 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).
89. 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).
90. 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).
91. 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).
92. 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).
93. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52**, 7182–7190 (1981).
94. G. M. Torrie, J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* **23**, 187–199 (1977).
95. 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).
96. J. K. Myers, C. N. Pace, J. M. Scholtz, Helix Propensities Are Identical in Proteins and Peptides. *Biochemistry* **36**, 10923–10929 (1997).
97. 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 discussion. 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 GitHub. All other relevant data are available in the main text or the supplementary materials.

## Supplementary Materials

Materials and Methods

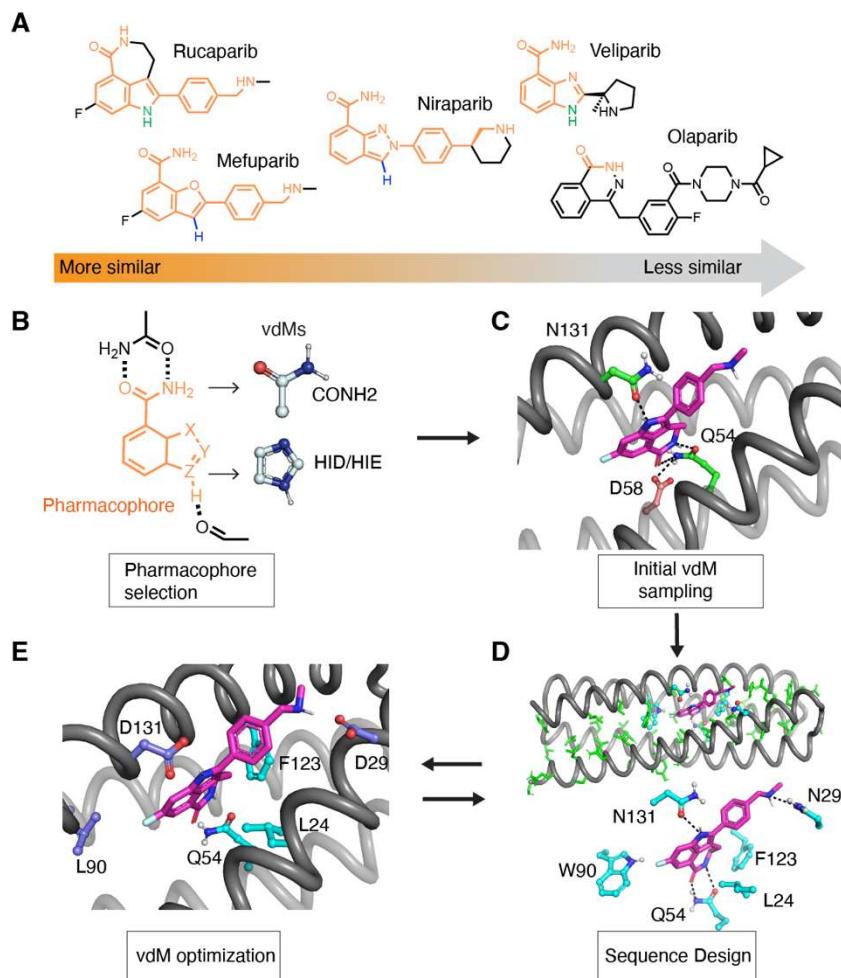
Figs. S1 to S36

Tables S1 to S5

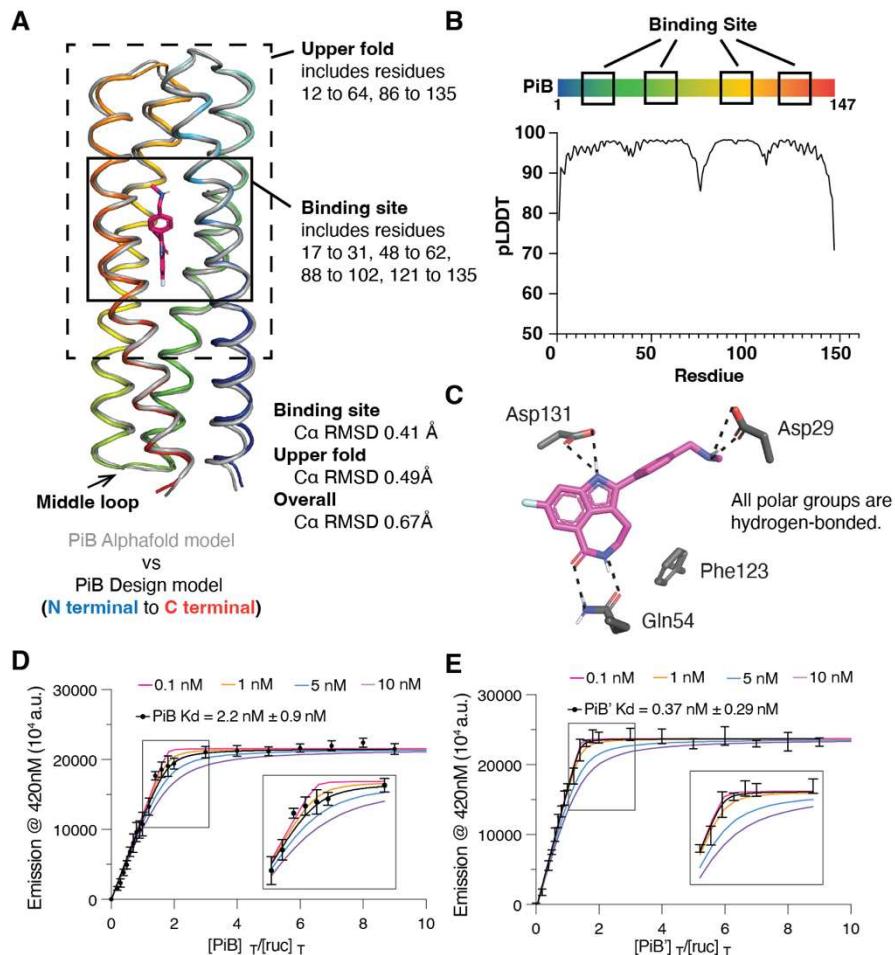
Supplementary Text

References

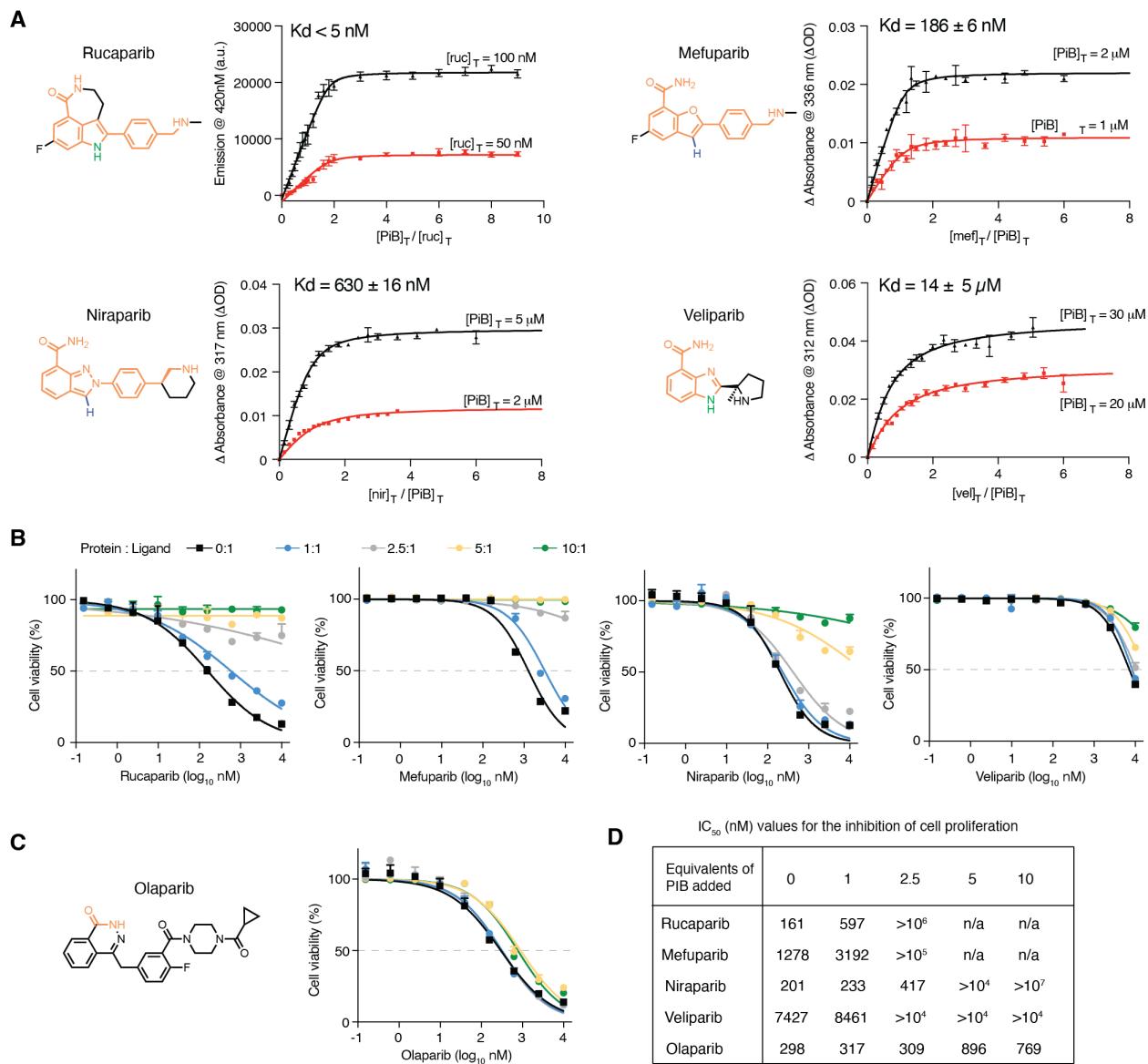
## Figures



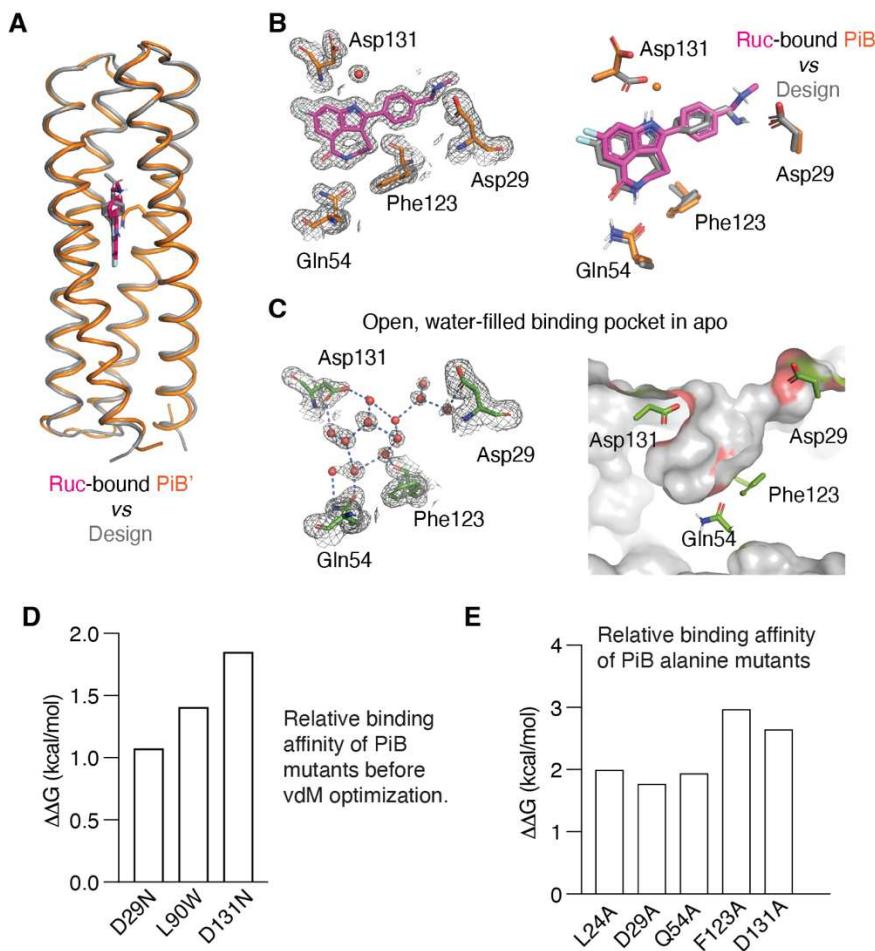
**Figure 1. The computational design of poly(ADP-ribose) polymerase inhibitors (PARPi) protein binders.** (A) The PARPi analogues. The shared chemical features are marked in orange. Olaparib is used as a negative control in the design and binding assay. (B – E) The overall design strategy. (B, C) We first define the pharmacophore and use COMBS to sample vdMs on the selected protein backbones. We initially targeted the indole and carboxamide of the drug, and used COMBS to discover sidechains that would form first and second-shell hydrogen bonds to both of these chemical groups. 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 purple). 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. (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.



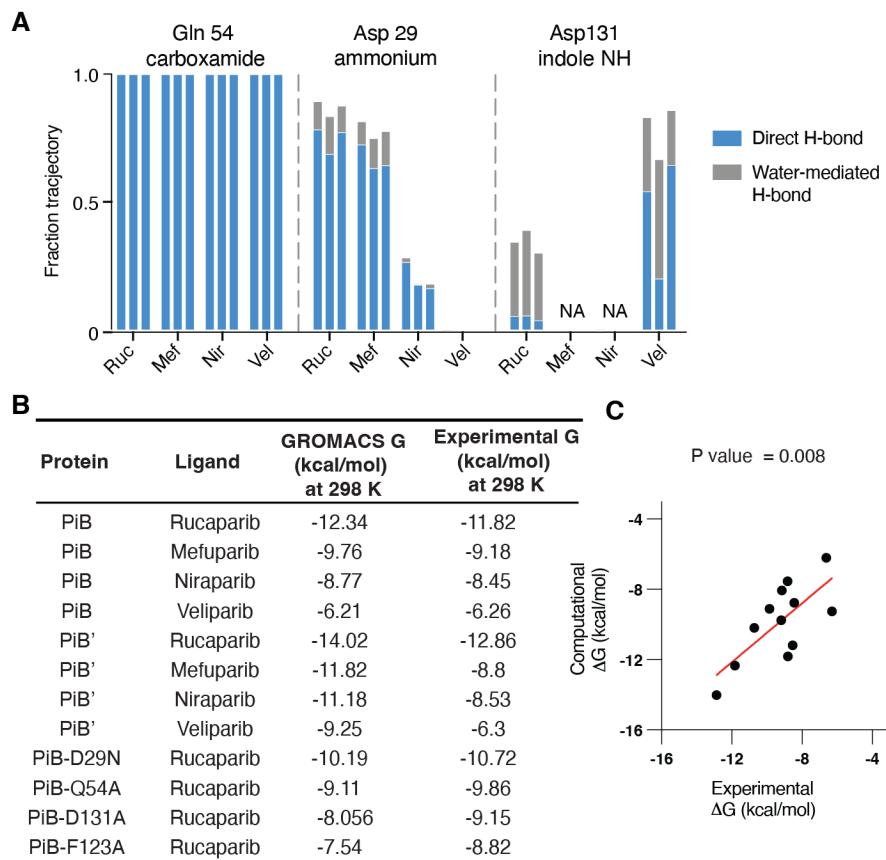
**Figure 2. Assessing the computational model and experimental binding of PiB to rucaparib.** (A) The AlphaFold2 model agrees with the designed PiB very well, with the binding site  $\text{Ca RMSD}$  of  $0.41 \text{ \AA}$ , the upper fold  $\text{Ca RMSD}$  of  $0.49 \text{ \AA}$  and overall  $\text{Ca RMSD}$  of  $0.67 \text{ \AA}$ . (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  $\text{Ca RMSD}$ . (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  $K_D < 5 \text{ nM}$ . The fluorescence emission intensity at  $420 \text{ nm}$  of rucaparib (excitation wavelength  $355 \text{ nm}$ ) 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  $2.2 (\pm 0.9) \text{ nM}$  for PiB, and  $0.37 (\pm 0.29) \text{ nM}$  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  $5 \text{ nM}$ . Therefore, while the most probable binding constants were  $2$  and  $0.4 \text{ nM}$ , respectively, we can confidently conclude that the values for PiB and PiB' are less than  $5 \text{ nM}$ . The titration was carried out in buffer containing  $50 \text{ mM Tris}, 100 \text{ mM NaCl}$  (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<sub>50</sub> 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) C $\alpha$  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  $\sigma$ . 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 inability 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  $\Delta G$  binding from the GROMACS calculation with the experimental value from spectral titrations.