Optimizing stability in dynamic small-molecule binding proteins

Marc Scherer [1,2], Mark Kriegel [1], Birte Höcker [1,*], Sarel J. Fleishman [2,*]

[1] University of Bayreuth, Department of Biochemistry, Bayreuth, Germany

[2] Weizmann Institute of Science, Department of Biomolecular Sciences, Rehovot, Israel

[*] Corresponding authors

Keywords: protein dynamics, computational design, PROSS, Gaussian network models, small-molecule binding

Abstract

The function of dynamic proteins is determined by the stability of distinct conformational states and the energy barriers that separate these states. For most dynamic proteins, the molecular details of the energy barriers are not known, implying a fundamental limit to the ability of protein design methods to engineer beneficial mutations without disrupting activity. We hypothesized that designing mutations that are compatible with structurally distinct equilibrium conformations may enable reliable stability design. We focus on periplasmic binding proteins (PBPs), a superfamily of dynamic proteins that change conformation from open to closed states in response to binding their small-molecule ligands. We find that the space of allowed mutations computed for one conformation is incompatible with the other. We therefore constrained putative conformational hinge points and interface residues and filtered out incompatible mutations. Starting from four different PBPs, we designed a total of 16 stabilized variants with 7-28 mutations each. Our results show that designs with a subset of mutations compatible with both conformations and structural constraints reliably enhance the thermal stability of PBPs while mitigating functional trade-offs. Our work demonstrates a potentially straightforward path for the one-shot design of thermostable and responsive biosensors.

Introduction

Many proteins are dynamic, alternating between several conformational states in equilibrium or in response to binding other molecules.[1,2] Protein dynamics enable transitions between functionally different states, allowing for complex outcomes such as allosteric regulation and cooperativity.[3] The folding landscapes of dynamic proteins may have been tuned by evolution to facilitate rapid (nanosecond-to-second) exchange between functional states while avoiding non-functional ones.[4–6] Notwithstanding their importance, the energy landscapes governing protein conformational equilibria are complex, remain poorly understood, and are difficult to estimate computationally, except in the case of small, fast-folding proteins.[7–9]

Our paper focuses on improving the stability of dynamic binding proteins. In the case of proteins that exhibit limited dynamics, reliable methods that combine phylogenetic analysis and energy calculations applied to a single conformation, such as PROSS (protein repair one-stop shop) and FireProt, have shown high reliability and broad scope.[10–13] When applied to dynamic proteins, however, the results of stability design have been mixed. For instance, PROSS has been successfully applied to improve the stability, expressibility and tendency to form crystals of human estrogen receptor α (hERα) based on a single ligand- and coactivator-bound crystal structure[14] and to improve the expressibility of a voltage-gated potassium channel (Kv channel);[15] in both cases, without a discernible drop in activity. Notably, interfaces and putative hinge points were restricted from mutation. In contrast, applying PROSS to the periplasmic binding protein (PBP) DalS[16] and the amino ester hydrolase QVH[17] (without considering hinge points) decreased activity. PBPs may be an exceptionally challenging target for computational design as a previous design study[18] showed unpredicted conformational changes, loss of stability, and undesired functional profiles.[19]

We focus on PBPs because of the general interest in improving the stability of dynamic proteins and the specific difficulties observed in previous applications of design to these proteins. PBPs share a two-lobe architecture connected by a hinge region and employ a 'venus flytrap' mechanism switching from an open to a closed conformation to bind various small molecules.[6,20,21] Despite a shared structural fold, PBPs possess diverse sequences,[22,23] and their energy landscapes are complex, featuring multiple meta-stable states.[24,25] In addition, PBPs have potential applications as biosensors.[26–28]

Results

Conformation impacts stability design choices

We began by searching structural databases for PBPs that undergo large conformational changes upon ligand binding and have been structurally characterized in both open and closed states. We selected four PBPs (PotF, TphC, MBP, and LAO) that bind to diverse ligands, ranging from aromatic dicarboxylic acids to sugars (Figure 1A; PDB IDs can be found in Table S1). To serve as a reference, we applied PROSS design calculations with standard parameters on the open and closed structures of each PBP (Figure 1B, Design X.1 & X.2; see Methods for details).

We compared the $\Delta\Delta G$ values for all mutations of the open and closed-state designs. This analysis showed that some mutations were stabilizing in one and destabilizing in the other state (Figure 2A). Generally, the open-state energies tended to be lower than the closed-state ones, likely due to the high packing density in the latter. Inspection of the residues around PROSS mutations that showed state-dependent sign differences in energy indicated subtle differences in backbone and side-chain configurations (Figure 2B). We categorized the mutations into surface, boundary and core layers, finding that the energies of surface mutations were similar in both states, likely reflecting reduced sensitivity due to low packing density (Figure S1). By contrast, boundary and core mutations, especially in PotF and TphC, showed sign energy differences. We concluded that PBP open-to-closed transitions are caused by global backbone changes and that mutations in the core of the protein are especially sensitive to the conformational state.

Two additional sets of designs were generated: To avoid changes in the conformational equilibrium of PBPs, mutations that showed state-dependent sign differences in energy were excluded in the third set (Figure 1B, Design X.3). Still, mutations in the interface and hinge region were present. As those areas showed the strongest changes in local backbone density as seen by Gaussian Network Model (GNM) calculations (Figure 2C), in the fourth set of designs we also eliminated mutations in the hinge or lobe interface (Methods; Figure 1B, Figure S2, Design X.4). In total, 16 PBP designs (4 designs per wild type protein) were selected for experimental testing and compared to their respective wild type (WT) (Table 1 and Table S2).

Two-state filter and structural constraints enable reliable stability optimization

WT proteins and designs were expressed in E. coli and purified. First, the folding of all PBP designs was checked with circular dichroism (CD) spectroscopy (Figure S3-4). Subsequent

thermal melt (Tmelt) experiments revealed that all PBP designs exhibited improved thermal stability with apparent ΔT_M values ranging between 4 - 17 °C (Figure 3A and S4, Table 1). For comparison, the apparent T_M values of PotF, TphC, MBP and LAO WT were 63 °C, 61 °C, 63 °C and 47 °C, respectively. Remarkably, in case of PotF, TphC and LAO, the increase in T_M was not correlated with the number of mutations opposing a trend seen in previous work.[29] Next, we investigated whether the PROSS designs bind their cognate ligands. For that, Tmelt experiments were performed in the presence of 100 µM of the cognate substrates, putrescine, terephthalic acid, maltose or L-lysine in case of PotF, TphC, MBP and LAO designs, respectively. Using a threshold of +1 °C ΔT_M, all PBP designs bound their cognate ligands but Des1.4 (Table S4). Except for three TphC designs, PBP PROSS designs exhibited slightly lower ΔT_M (ligand) values than the wild-type proteins (Figure 3B). These results indicate that the binding affinity of most PBP variants to their respective ligands might be decreased. To investigate the effect of the mutations on the ligand binding behavior in more detail, isothermal titration calorimetry (ITC) measurements of the PBP designs were conducted and compared with the binding affinities (K_D) of their respective WT (Figure 3C and Figures S6-9). The binding affinities of the PBP WT proteins were determined to 42 ± 3 nM for PotF, 1180 ± 250 nM for TphC, 1240 ± 102 nM for MBP, and 93 ± 35 nM for LAO which were in accordance with previously published binding constants: 68 ± 40 nM for PotF[25], 364 nM for TphC[30], 1102 ± 65 nM for MBP[31], and 232 ± 106 nM for LAO[32]. Three out of eight single-state designs exhibited a decreased affinity of fivefold or more; two of these were designed based on the open state. For the single-state MBP design Des3.2, which is based on the closed state, no proper binding data was obtainable (see Table 1 and Figures S8). The two-state filtered designs without hinge constraints, on the other hand, showed decreased binding affinity only in one out of four cases, while all two-state filtered designs with hinge constraints showed WT-like affinity. The binding affinity was marginally improved (less than fivefold) for TphC Des2.2-2.4 and MBP Des3.3. Taken together, these data suggest that designing PBPs from only one state is more likely to change the binding affinity. Two-state mutation filtering, especially with constraints on the hinge region, on the other hand, reliably optimized the thermal stability of all PBPs while retaining their ligand-binding affinity.

Discussion

Optimizing dynamic proteins without prior knowledge of the underlying energy landscape is a challenging task, as conformational states and function are usually not independent. Our thermal melt and binding data revealed that evolutionary constraints on the sequence space in single-state design of PBPs were not sufficient to maintain WT-like binding affinity. Fine balancing is likely needed to allow PBP open-to-closed transitions in fast time-scales (low energy barriers) but also stabilize an open state ensemble for the ligand to enter. Mutations in the binding pocket of PBPs can have an effect on the conformational ensemble and accessibility of closed states.[33,34] Similarly, mutations in the hinge area can impair ligand-binding[31,35] by differentially stabilizing particular backbone conformations. Taking advantage of structural information of a second state improved the reliability of the design approach but still led to decreased binding affinity in the case of the LAO protein (Des4.3). Additional constraints on residues of the hinge and interface were necessary to achieve success in all cases (DesX.4) while still improving thermal stability at a similar level as designs with more than twice the number of mutations. Hence, the approach presented here allowed reliable optimization of PBPs at a low mutational load. These designs can serve as starting points for biosensor engineering.

While our design approach worked well on four PBPs, it remains to be seen how generalizable this method is. For most dynamic proteins, high-quality structural information of all functional states is not available. If modelling of structural ensembles of dynamic proteins can be achieved with high precision using machine learning-based approaches such as AlphaFold[36,37] or Chai[38], the applicability of the design approach could be broadened. We saw that at least one out of five models (default setting) generated by AlphaFold3 resembled the open and closed conformations of the PBP wildtypes studied here. Still, it must be ensured that the chosen structural representations constitute functional states of the protein (depending on the system, maybe even the transition states) as the stabilization of non-functional conformations will negatively impact function. The study of PBP-like folds provides initial steps towards designing complex dynamic proteins.

Methods

Sequence design with PROSS

PROSS[10,11] uses the crystal structure and amino acid sequence of a protein as input. First, a multiple sequence alignment (MSA) is generated that serves as a basis to compute a position-specific scoring matrix (PSSM) reflecting the likelihood of an amino acid to occur at each position of the protein in the natural diversity. Filtering out all amino acid identities that do not occur frequently in the MSA (PSSM values < 0), allows to focus the sequence space on regions that evolution has explored and can be considered more likely to maintain the protein fold. Subsequently, Rosetta atomistic design calculations of each single amino acid substitution identified stabilizing mutations (Rosetta FilterScan functionality calculates $\Delta\Delta G$ values of mutations compared to WT) which were then considered for a final combinatorial design step based on nine energy cut-offs. The PROSS algorithm was run with default parameters using open or closed state crystal structures of each PBP as input. Generally, the binding pocket residues 6 Å around the ligands in the closed structures were held fixed in all designs to ensure that interactions with the ligands are kept intact. In the case of DesX.4, hinge point and interface residues were fixed (see Methods below - Gaussian network models). Designs with the same Rosetta energy cut-off and 4-10 % mutational load were chosen. Each mutation was inspected by eye and hand-picked mutations were excluded if mutations were too close to the termini or no rationale of stabilizing the protein could be found. If the $\Delta\Delta G$ of a mutation was stabilizing ($\Delta\Delta G < 0$) in both the open and closed state designs, this mutation was included in the two-state design variants (two-state filter). Categorization of mutations was done with the Rosetta LayerSelector using default settings.

Gaussian network models

PDBs of the open and closed ligand-bound structure were used (Table S1). Ligands were deleted and in the case of TphC, the residue numbering of the closed state was adjusted to that of the open state. The per-residue square fluctuations were obtained by GNM calculations using the first five modes with the distance cut-off set to 10 Å and the spring constant set to 1 using the ProDy python implementation.[39,40] To obtain the hinge residues, GNM was calculated based on only the 1st mode. The ProDy python implementation provides a function to output hinge residues. The protein lobes were then split along the hinge residues and the interface residues between the lobes were determined with a 6 Å distance cut-off using the PyMol[41] python implementation.

Cloning, expression & purification

All genes were synthesized by Twist Bioscience HQ including restriction sites for *Nde*I and *Xho*I on the N-terminal and C-terminal flanking sites, respectively (see Table S2 for amino acid sequences). The DNA fragments were first digested and then ligated into a pET21b(+)-vector which included a C-terminal His-tag. In case of TphC fragments, a N-terminal His-tag and TEV-cleavage site was included in the synthesized gene as described elsewhere.[30] Hence, a stop codon was inserted between the gene and the *Xho*I cleavage site by PCR (primers listed in Table S3) before cloning into the expression vector. Top10 *E. coli* cells were transformed with the generated vectors and resulting clones were verified by Sanger sequencing. For expression, BL21 (DE3) cells were transformed with the plasmids and plated out on LB agar plates containing 100 µg/mL ampicillin. Each 25 mL LB media-based overnight culture supplemented with 100 µg/mL ampicillin was inoculated with a single colony. 0.5 L TB media was then inoculated with 5 mL of the overnight culture and incubated at 37 °C until the OD_{600} reached a value of ~ 0.7. Next, overexpression was induced by adding isopropyl-β-thiogalactoside (IPTG) to a final concentration of 1 mM and cultures were incubated for ~ 18 h at 18 °C. Cells were harvested by centrifugation (Beckman Coulter, Avanti JXN-26, JLA8.1000 rotor) at 4,000 g, 4 °C for 20 min. The cell pellets were washed with 30 mL lysis buffer (25 mM NaH_2PO_4/Na_2HPO_4, 500 mM NaCl, 20 mM Imidazole, pH 7.4) supplemented with protease inhibitor mix HP (Serva) and centrifuged again (Eppendorf 5920R) at 4,000 g, 4 °C for 10 min. The cell pellets were resuspended in lysis buffer, sonicated (Branson 6.3 mm tip, 2 x 3 min, 40 % duty cycle, output power 4) and centrifuged (Beckman Coulter, Avanti JXN-26, JA25.50 rotor) at 18,000 g, 4 °C for 1 h. The supernatant was loaded onto a His-Trap HP 5 mL column (GE Healthcare) that was equilibrated with the lysis buffer beforehand using a peristaltic pump (Pump P-1, Pharmacia). After loading, the column was washed with 10 column volumes (CV) of lysis buffer. Next, the protein was unfolded by adding 5 CV 6 M Guanidine Hydrochloride (GdHCl) and incubation for 1 h. Endogenous ligands were removed by washing with 5 CV GdHCl. The protein was refolded by washing with 10 CV of lysis buffer. Elution was performed in a stepwise manner on an ÄKTA system using elution buffer (25 mM NaH_2PO_4/Na_2HPO_4, 500 mM NaCl, 500 mM Imidazole, pH 7.4). The protein fractions were pooled and concentrated using centrifugal concentrators to a final volume of 10 mL and loaded onto a preparative size exclusion column (HiLoad Superdex 75 26/60, GE Healthcare) equilibrated with running buffer (25 mM NaH_2PO_4/Na_2HPO_4, 500 mM NaCl, pH 7.4). Monomeric protein fractions were pooled and concentrated for the subsequent measurements.

Concentrations were checked using a NanoDrop Spectrophotometer (Eppendorf BioPhotometer D30). Expressions and purifications were validated by SDS-PAGE.

Circular dichroism spectroscopy

To remove salts, protein samples were dialyzed overnight against 2 L of CD buffer (25 mM NaH_2PO_4/Na_2HPO_4, pH 7.4) using a 10 kDa molecular weight cut-off. Aggregated protein was removed by centrifugation at 15,000 g for 5 min using a tabletop centrifuge (Eppendorf 5427R). Next, the protein concentration was adjusted to 0.3 mg/mL and verified photometrically. 200 µL of protein samples were mixed with either 100 µL CD buffer or 100 µL of ligand solution (300 µM stocks of ligand solubilized in CD buffer) to reach a final concentration of 100 µM for the ligands and 0.2 mg/ml for the protein. CD spectra and thermal melt experiments were obtained on a JASCO J-1500 using a glass cuvette with a 1 mm path length. For CD spectra, following settings were used: wavelengths 240-190 nm, bandwidth 1 nm, response time 2 s, data pitch 0.1 nm, scanning speed 100 nm/min and ten accumulations (for buffer controls only one accumulation was used). For Tmelt measurements, the following settings were used: wavelength of 222 nm, temperature range from 20-95 °C controlled by a Pelletier AWC100 (Julabo) using a slope of 1 °C/min, a bandwidth of 1 nm and a response time of 2 s. Raw CD data was normalized by subtraction of the respective buffer data and subsequent conversion of the measured signal in millidegree to mean residue molar ellipticity.[42] To obtain the apparent melting temperature T_M, thermal melt data was fit using a sigmoidal equation according to Niklasson et al.[43] The error plotted stems from the fit function uncertainty.

Isothermal titration calorimetry

Ligands were dissolved in the specific buffer that was used in the size-exclusion chromatography (SEC) run of each protein purification (stored at - 20 °C and thawed on the day of use). All protein and ligand samples were degassed and temperature equilibrated to 25 °C for at least 10 min using a degassing station (TA Instruments). 400 µL of protein samples were loaded into the sample cell of a low volume AffinityITC machine with gold cell (TA Instruments) and ligand solutions were added into the injection syringe (protein and ligand concentrations can be found in Table S5). ITC measurements were performed at 25 °C with 20 injections of 2.5 µL (one initial injection of 0.5 µL was excluded from analysis), an adaptive injection interval and stirring rate of 125 rpm. Subtractions of constant heat of dilution values for each ligand, peak integration and fitting with a one-site binding model were done with

NanoAnalyze (TA Instruments). The reported errors reflect the standard deviation of three technical replicas. Thermograms and binding isotherms were plotted in Figures S5-8.

Author Contributions

Conceptualization – M.S., B.H. & S.J.F., Investigation – M.S., Formal Analysis – M.S. & M.K., Manuscript writing – all authors.

Acknowledgments

We thank Ariel Tennenhouse for adding suggestions on the manuscript. We thank Sabrina Wischt and Pascal Kröger for experimental support. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 955623 (to B.H. and S.J.F.) and the European Synergy grant agreement No 951375 (ArtMotor) (to B.H.).

Abbreviations

PBP, periplasmic binding protein; PotF, putrescine binding protein; TphC, terephthalate binding protein; MBP, maltose binding protein; LAO, L-lysine/L-arginine/L-ornithine binding protein.

References

(1) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254* (5038), 1598–1603. https://doi.org/10.1126/science.1749933.

(2) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450* (7172), 964–972. https://doi.org/10.1038/nature06522.

(3) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), 331–339. https://doi.org/10.1038/nature13001.

(4) Chisholm, L. O.; Orlandi, K. N.; Phillips, S. R.; Shavlik, M. J.; Harms, M. J. Ancestral Reconstruction and the Evolution of Protein Energy Landscapes. *Annu. Rev. Biophys.* **2024**, *53* (Volume 53, 2024), 127–146. https://doi.org/10.1146/annurev-biophys-030722-125440.

(5) Gershenson, A.; Gierasch, L. M.; Pastore, A.; Radford, S. E. Energy Landscapes of Functional Proteins Are Inherently Risky. *Nat. Chem. Biol.* **2014**, *10* (11), 884–891. https://doi.org/10.1038/nchembio.1670.

(6) Gouridis, G.; Muthahari, Y. A.; de Boer, M.; Griffith, D. A.; Tsirigotaki, A.; Tassis, K.; Zijlstra, N.; Xu, R.; Eleftheriadis, N.; Sugijo, Y.; Zacharias, M.; Dömling, A.; Karamanou, S.; Pozidis, C.; Economou, A.; Cordes, T. Structural Dynamics in the Evolution of a Bilobed Protein Scaffold. *Proc. Natl. Acad. Sci.* **2021**, *118* (49), e2026165118. https://doi.org/10.1073/pnas.2026165118.

(7) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning. *Science* **2019**, *365* (6457), eaaw1147. https://doi.org/10.1126/science.aaw1147.

(8) Barducci, A.; Bonomi, M.; Parrinello, M. Linking Well-Tempered Metadynamics Simulations with Experiments. *Biophys. J.* **2010**, *98* (9), L44–L46. https://doi.org/10.1016/j.bpj.2010.01.033.

(9) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330* (6002), 341–346. https://doi.org/10.1126/science.1187409.

(10) Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **2016**, *63* (2), 337–346. https://doi.org/10.1016/j.molcel.2016.06.012.

(11) Weinstein, J. J.; Goldenzweig, A.; Hoch, S.; Fleishman, S. J. PROSS 2: A New Server for the Design of Stable and Highly Expressed Protein Variants. *Bioinformatics* **2021**, *37* (1), 123–125. https://doi.org/10.1093/bioinformatics/btaa1071.

(12) Musil, M.; Stourac, J.; Bendl, J.; Brezovsky, J.; Prokop, Z.; Zendulka, J.; Martinek, T.; Bednar, D.; Damborsky, J. FireProt: Web Server for Automated Design of Thermostable Proteins. *Nucleic Acids Res.* **2017**, *45* (W1), W393–W399. https://doi.org/10.1093/nar/gkx285.

(13) Musil, M.; Jezik, A.; Horackova, J.; Borko, S.; Kabourek, P.; Damborsky, J.; Bednar, D. FireProt 2.0: Web-Based Platform for the Fully Automated Design of

Thermostable Proteins. *Brief. Bioinform.* **2024**, *25* (1), bbad425. https://doi.org/10.1093/bib/bbad425.

(14) Kriegel, M.; Wiederanders, H. J.; Alkhashrom, S.; Eichler, J.; Muller, Y. A. A PROSS-Designed Extensively Mutated Estrogen Receptor α Variant Displays Enhanced Thermal Stability While Retaining Native Allosteric Regulation and Structure. *Sci. Rep.* **2021**, *11* (1), 10509. https://doi.org/10.1038/s41598-021-89785-1.

(15) Weinstein, J. J.; Saikia, C.; Karbat, I.; Goldenzweig, A.; Reuveny, E.; Fleishman, S. J. One-Shot Design Elevates Functional Expression Levels of a Voltage-Gated Potassium Channel. *Protein Sci.* **2024**, *33* (6), e4995. https://doi.org/10.1002/pro.4995.

(16) Vongsouthi, V.; Whitfield, J. H.; Unichenko, P.; Mitchell, J. A.; Breithausen, B.; Khersonsky, O.; Kremers, L.; Janovjak, H.; Monai, H.; Hirase, H.; Fleishman, S. J.; Henneberger, C.; Jackson, C. J. A Rationally and Computationally Designed Fluorescent Biosensor for D-Serine. *ACS Sens.* **2021**, *6* (11), 4193–4205. https://doi.org/10.1021/acssensors.1c01803.

(17) Lagerman, C. E.; Joe, E. A.; Grover, M. A.; Rousseau, R. W.; Bommarius, A. S. Improvement of α-Amino Ester Hydrolase Stability via Computational Protein Design. *Protein J.* **2023**, *42* (6), 675–684. https://doi.org/10.1007/s10930-023-10155-z.

(18) Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. Computational Design of Receptor and Sensor Proteins with Novel Functions. *Nature* **2003**, *423* (6936), 185–190. https://doi.org/10.1038/nature01556.

(19) Schreier, B.; Stumpp, C.; Wiesner, S.; Höcker, B. Computational Design of Ligand Binding Is Not a Solved Problem. *Proc. Natl. Acad. Sci.* **2009**, *106* (44), 18491–18496. https://doi.org/10.1073/pnas.0907950106.

(20) Tang, C.; Schwieters, C. D.; Clore, G. M. Open-to-Closed Transition in Apo Maltose-Binding Protein Observed by Paramagnetic NMR. *Nature* **2007**, *449* (7165), 1078–1082. https://doi.org/10.1038/nature06232.

(21) Mao, B.; Pear, M. R.; McCammon, J. A.; Quiocho, F. A. Hinge-Bending in L-Arabinose-Binding Protein. The "Venus's-Flytrap" Model. *J. Biol. Chem.* **1982**, *257* (3), 1131–1133. https://doi.org/10.1016/S0021-9258(19)68161-5.

(22) Dwyer, M. A.; Hellinga, H. W. Periplasmic Binding Proteins: A Versatile Superfamily for Protein Engineering. *Curr. Opin. Struct. Biol.* **2004**, *14* (4), 495–504. https://doi.org/10.1016/j.sbi.2004.07.004.

(23) Scheepers, G. H.; Lycklama a Nijeholt, J. A.; Poolman, B. An Updated Structural Classification of Substrate-Binding Proteins. *FEBS Lett.* **2016**, *590* (23), 4393–4401. https://doi.org/10.1002/1873-3468.12445.

(24) Marinelli, F.; Kuhlmann, S. I.; Grell, E.; Kunte, H.-J.; Ziegler, C.; Faraldo-Gómez, J. D. Evidence for an Allosteric Mechanism of Substrate Release from Membrane-Transporter Accessory Binding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (49), E1285–E1292. https://doi.org/10.1073/pnas.1112534108.

(25) Kröger, P.; Shanmugaratnam, S.; Ferruz, N.; Schweimer, K.; Höcker, B. A Comprehensive Binding Study Illustrates Ligand Recognition in the Periplasmic Binding Protein PotF. *Structure* **2021**, *29* (5), 433-443.e4. https://doi.org/10.1016/j.str.2020.12.005.

(26) Hellweg, L.; Pfeifer, M.; Tarnawski, M.; Thing-Teoh, S.; Chang, L.; Bergner, A.; Kress, J.; Hiblot, J.; Wiedmer, T.; Superti-Furga, G.; Reinhardt, J.; Johnsson, K.; Leippe, P. AspSnFR: A Genetically Encoded Biosensor for Real-Time Monitoring of Aspartate in Live Cells. *Cell Chem. Biol.* **2024**, *31* (8), 1529-1541.e12. https://doi.org/10.1016/j.chembiol.2024.05.002.

(27) Marvin, J. S.; Schreiter, E. R.; Echevarría, I. M.; Looger, L. L. A Genetically Encoded, High-Signal-to-Noise Maltose Sensor. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (11), 3025–3036. https://doi.org/10.1002/prot.23118.

(28) Kröger, P.; Stiel, A. C.; Stüven, B.; Shanmugaratnam, S.; Schoch, S.; Wachten, D.; Höcker, B. A Fluorescent Biosensor for the Visualization of Agmatine. bioRxiv July 27, 2024, p 2024.07.27.605435. https://doi.org/10.1101/2024.07.27.605435.

(29) Peleg, Y.; Vincentelli, R.; Collins, B. M.; Chen, K.-E.; Livingstone, E. K.; Weeratunga, S.; Leneva, N.; Guo, Q.; Remans, K.; Perez, K.; Bjerga, G. E. K.; Larsen, Ø.; Vaněk, O.; Skořepa, O.; Jacquemin, S.; Poterszman, A.; Kjær, S.; Christodoulou, E.; Albeck, S.; Dym, O.; Ainbinder, E.; Unger, T.; Schuetz, A.; Matthes, S.; Bader, M.; de Marco, A.; Storici, P.; Semrau, M. S.; Stolt-Bergner, P.; Aigner, C.; Suppmann, S.; Goldenzweig, A.; Fleishman, S. J. Community-Wide Experimental Evaluation of the PROSS Stability-Design Method. *J. Mol. Biol.* **2021**, *433* (13), 166964. https://doi.org/10.1016/j.jmb.2021.166964.

(30) Gautom, T.; Dheeman, D.; Levy, C.; Butterfield, T.; Alvarez Gonzalez, G.; Le Roy, P.; Caiger, L.; Fisher, K.; Johannissen, L.; Dixon, N. Structural Basis of Terephthalate Recognition by Solute Binding Protein TphC. *Nat. Commun.* **2021**, *12* (1), 6244. https://doi.org/10.1038/s41467-021-26508-0.

(31) Seo, M.-H.; Park, J.; Kim, E.; Hohng, S.; Kim, H.-S. Protein Conformational Dynamics Dictate the Binding Affinity for a Ligand. *Nat. Commun.* **2014**, *5* (1), 3724. https://doi.org/10.1038/ncomms4724.

(32) Clifton, B. E.; Jackson, C. J. Ancestral Protein Reconstruction Yields Insights into Adaptive Evolution of Binding Specificity in Solute-Binding Proteins. *Cell Chem. Biol.* **2016**, *23* (2), 236–245. https://doi.org/10.1016/j.chembiol.2015.12.010.

(33) Kröger, P.; Shanmugaratnam, S.; Scheib, U.; Höcker, B. Fine-Tuning Spermidine Binding Modes in the Putrescine Binding Protein PotF. *J. Biol. Chem.* **2021**, *297* (6). https://doi.org/10.1016/j.jbc.2021.101419.

(34) Scheib, U.; Shanmugaratnam, S.; Farías-Rico, J. A.; Höcker, B. Change in Protein-Ligand Specificity through Binding Pocket Grafting. *J. Struct. Biol.* **2014**, *185* (2), 186–192. https://doi.org/10.1016/j.jsb.2013.06.002.

(35) van den Noort, M.; de Boer, M.; Poolman, B. Stability of Ligand-Induced Protein Conformation Influences Affinity in Maltose-Binding Protein. *J. Mol. Biol.* **2021**, *433* (15), 167036. https://doi.org/10.1016/j.jmb.2021.167036.

(36) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(37) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(38) Discovery, C.; Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. Chai-1: Decoding the Molecular Interactions of Life. bioRxiv October 15, 2024, p 2024.10.10.615955. https://doi.org/10.1101/2024.10.10.615955.

(39) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27* (11), 1575–1577. https://doi.org/10.1093/bioinformatics/btr168.

(40) Zhang, S.; Krieger, J. M.; Zhang, Y.; Kaya, C.; Kaynak, B.; Mikulska-Ruminska, K.; Doruker, P.; Li, H.; Bahar, I. ProDy 2.0: Increased Scale and Scope after 10 Years of Protein Dynamics Modelling with Python. *Bioinformatics* **2021**, *37* (20), 3657–3659. https://doi.org/10.1093/bioinformatics/btab187.

(41) Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8, 2015.

(42) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2006**, *1* (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.

(43) Niklasson, M.; Andresen, C.; Helander, S.; Roth, M. G. L.; Zimdahl Kahlin, A.; Lindqvist Appell, M.; Mårtensson, L.-G.; Lundström, P. Robust and Convenient Analysis of Protein Thermal and Chemical Stability. *Protein Sci.* **2015**, *24* (12), 2055–2062. https://doi.org/10.1002/pro.2809.

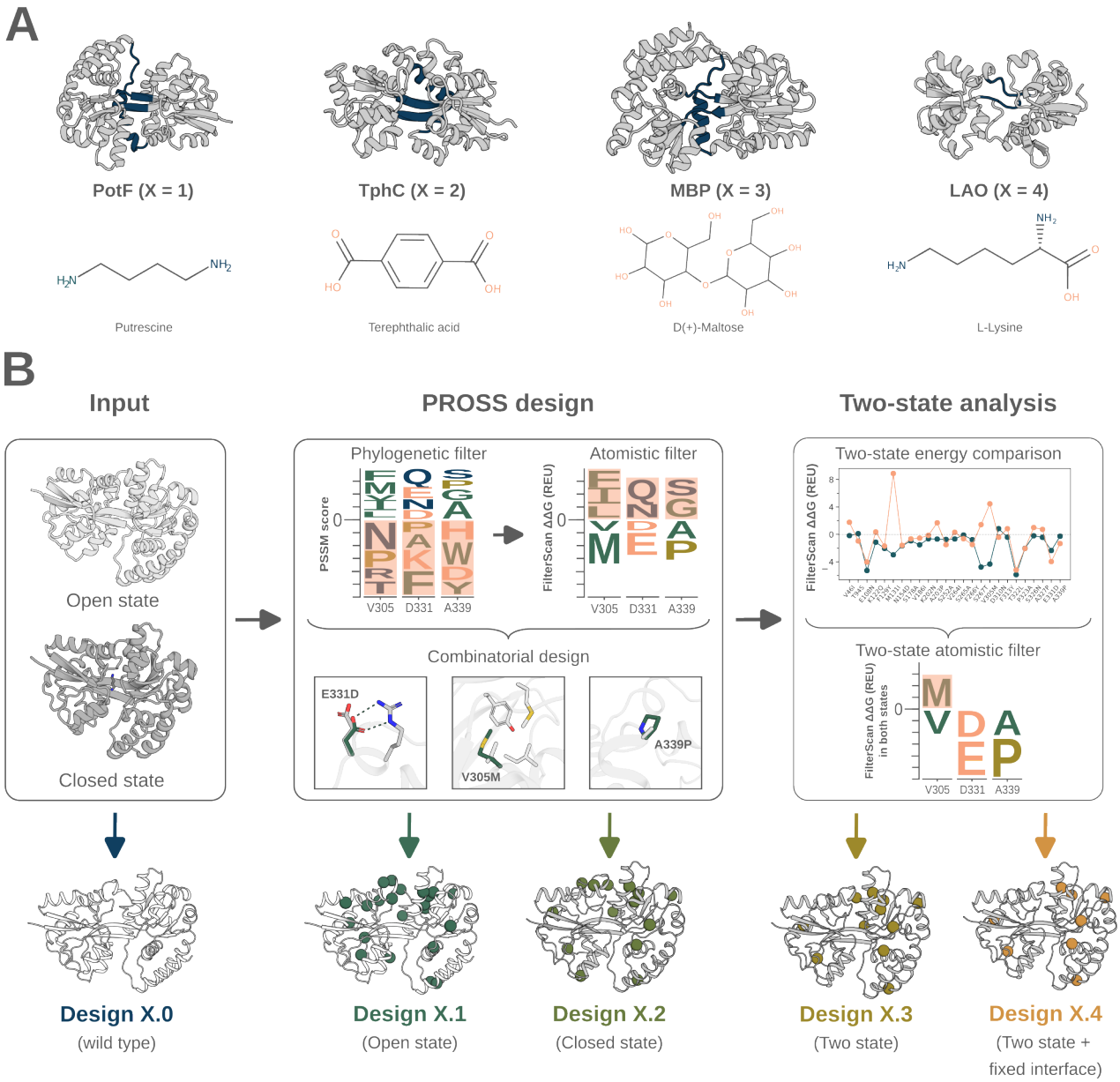


Figure 1. Design strategy for optimising PBP stability with single- and two-state PROSS design. (A) Four PBPs were chosen: putrescine-binding protein PotF; terephthalic acid-binding protein TphC; maltose-binding protein MBP; and lysine-binding protein LAO. Closed-state crystal structures (gray) with hinge regions (blue) are shown as cartoons. (B) The design strategy used the open- and closed-state crystal structures of PotF, TphC, MBP and LAO as input. The WT sequences of all PBPs were used as control (Design X.0). PROSS design calculations were run with default parameters, generating Designs X.1 and X.2 based on the open and closed structures, respectively. The impact of mutations on the energy (ΔΔG) depends on the backbone conformation between the two states. Hence, mutations that were destabilizing (positive energies) in any state were filtered out leading to Design X.3. Design X.4 was generated by fixing the hinge region and lobe interface before running PROSS design calculations and the two-state energy filter.

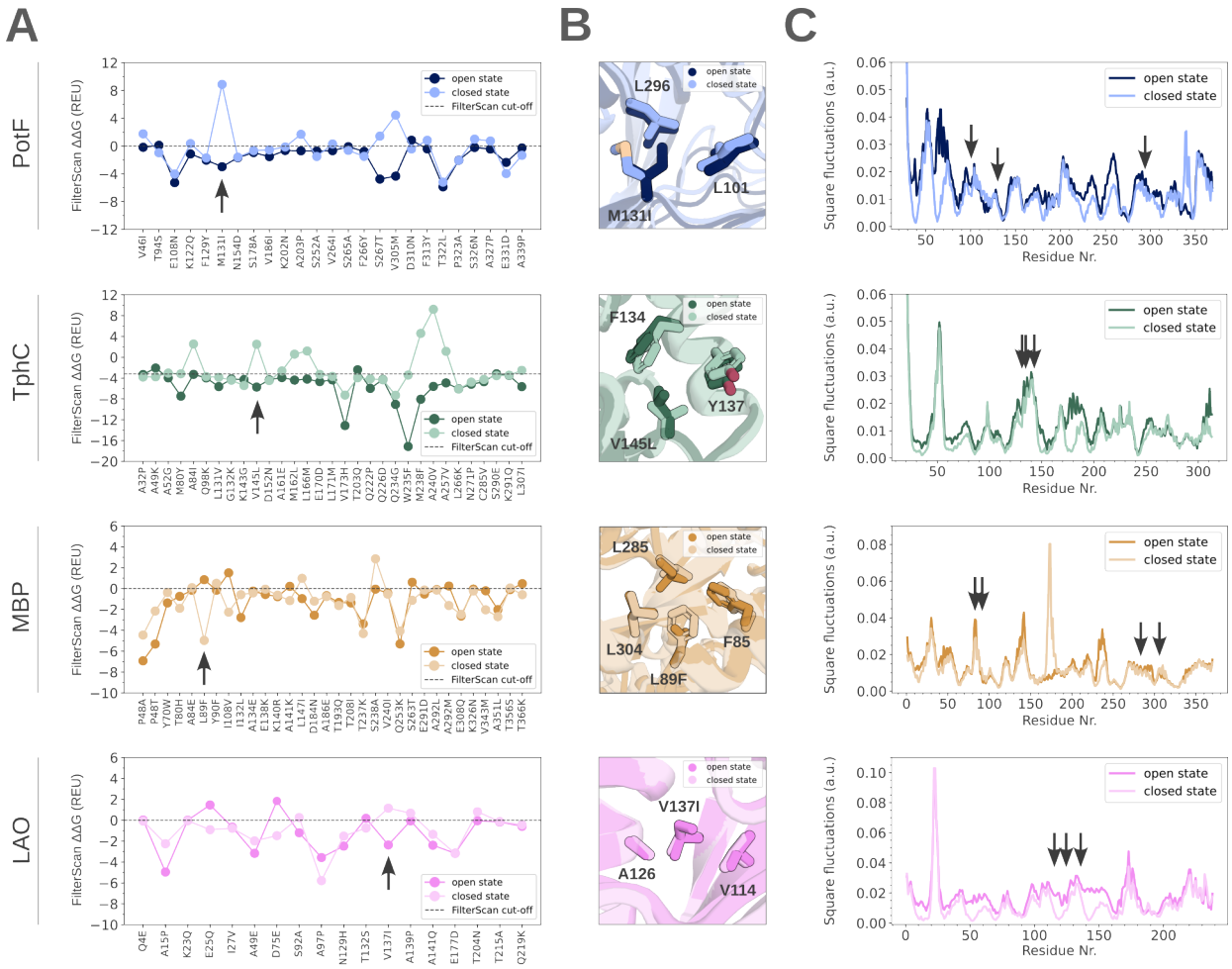


Figure 2. Rosetta $\Delta\Delta G$ and packing density analysis of PBP single-state PROSS mutations. (A) $\Delta\Delta G$ values of PROSS mutations generated based on the open and closed state of PotF (blue), TphC (green), MBP (orange), and LAO (pink), and cut-off energies (dotted lines). Energies are plotted in Rosetta energy units (REU). Arrows indicate mutations visualized in (B). (B) Subtle packing differences of residues were observed around four representative mutations (shown as sticks) of PotF (blue), TphC (green), MBP (orange) and LAO (pink) designs, respectively. Mutations were shown to have stabilizing $\Delta\Delta G$ values in one state and destabilizing values in the other state as part of the analysis in (A). (C) Backbone square fluctuations (inverse of the packing density) of PotF (blue), TphC (green), MBP (orange) and LAO (pink) based on GNMs. Arrows indicate position of residues shown in (B).

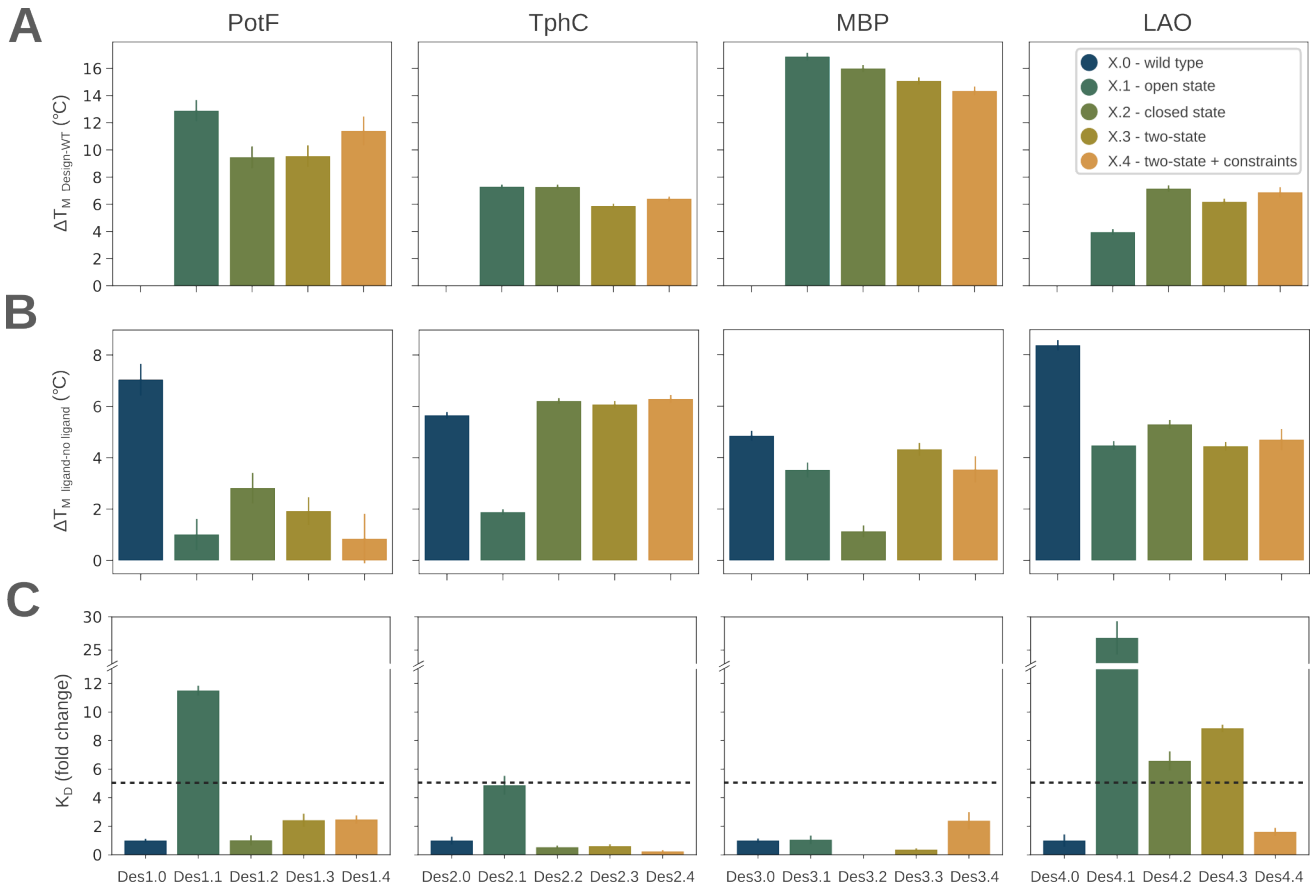


Figure 3. Thermostability and binding affinity of PBP WTs and designs. (A) The change in apparent melting temperature ΔT_M was calculated relative to the WT. (B) The change in the apparent ΔT_M upon addition of 100 µM ligand was calculated by subtracting the respective sample with the apo protein control (no ligand). (C) Fold change of K_D against WT. Error bars reflect the standard deviation of three independent repeats. Putrescine, terephthalic acid, maltose and L-lysine were used for PotF, TphC, MBP and LAO binding studies, respectively. A fivefold decreased affinity compared to the WT is visualized by a dashed line.

Table 1. Thermal stability and binding parameters for the PBP designs and wild types

Protein target	Design ID	Nr. of mutations	Mutational load	T_M in °C	K_D (nM)
PotF	Des1.0	0	0 %	63 ± 0.4	42 ± 3
	Des1.1	22	6 %	76 ± 0.3	479 ± 12
	Des1.2	14	4 %	73 ± 0.3	42 ± 13
	Des1.3	12	4 %	73 ± 0.3	101 ± 17
	Des1.4	9	3 %	75 ± 0.6	103 ± 10
TphC	Des2.0	0	0 %	61 ± 0.1	1180 ± 250
	Des2.1	28	10 %	68 ± 0.0	5750 ± 709
	Des2.2	20	7 %	68 ± 0.1	627 ± 69
	Des2.3	16	5 %	66 ± 0.1	716 ± 91
	Des2.4	9	3 %	67 ± 0.1	282 ± 39
MBP	Des3.0	0	0 %	63 ± 0.1	1240 ± 102
	Des3.1	25	7 %	80 ± 0.1	1310 ± 293
	Des3.2	25	7 %	79 ± 0.1	In µM range
	Des3.3	18	5 %	78 ± 0.1	450 ± 44
	Des3.4	14	4 %	77 ± 0.2	2960 ± 675
LAO	Des4.0	0	0 %	47 ± 0.1	93 ± 35
	Des4.1	13	5 %	51 ± 0.1	2496 ± 211
	Des4.2	13	5 %	55 ± 0.1	612 ± 56
	Des4.3	9	4 %	54 ± 0.1	824 ± 18
	Des4.4	7	3 %	54 ± 0.2	150 ± 21

Supporting Information

Optimizing stability in dynamic small-molecule binding proteins

Marc Scherer [1,2], Mark Kriegel [1], Birte Höcker [1,*], Sarel J. Fleishman [2,*]

[1] University of Bayreuth, Department of Biochemistry, Bayreuth, Germany

[2] Weizmann Institute of Science, Department of Biomolecular Sciences, Rehovot, Israel

Supporting Tables

Supporting Table 1 - Original organism and PDB identifiers of the PBPs

Supporting Table 2 - Amino acid sequences of PBP wildtypes and designs

Supporting Table 3 - Primers used for cloning

Supporting Table 4 - Thermal melt data of PBP wildtypes and designs

Supporting Table 5 - Protein and ligand concentrations for ITC measurements

Supporting Table 6 - ITC data of PBP wildtypes and designs

Supporting Figures

Supporting Figure 1 - FilterScan $\Delta\Delta G$ values of PROSS mutations categorized by layer

Supporting Figure 2 - PBP interface residues fixed in PROSS design

Supporting Figure 3 - CD spectra of PBP wildtypes and designs

Supporting Figure 4 - CD thermal melts of PBP wildtypes and designs without ligand

Supporting Figure 5 - CD thermal melts of PBP wildtypes and designs with ligand

Supporting Figure 6-9 - ITC data of PBP wildtypes and designs

Protein target	Original organism	PDB ID (open)	PDB ID (closed)
PotF	Escherichia coli (strain K12)	6YED [1]	6YE0 [1]
TphC	Comamonas sp.	7NDR [2]	7NDS [2]
MBP	Escherichia coli (strain K12)	1OMP [3]	3MBP [4]
LAO	Salmonella typhimurium (strain LT2/SGSC1412/ATCC 700720)	2LAO [5]	1LAF [6]

Supporting Table 1. Original organism and PDB[7] identifiers of the PBP crystal structures used.

Protein target	Design ID	Amino acid sequences
PotF	Des1.0	MAEQKTLHIYNWSDYIAPDTVANFEKETGIKVVYDVFDSNEVLEGKL MAGSTGFDLVVPSASFLERQLTAGVFQPLDKSKLPEWKNLDPELLKLV AKHDPDNKFAMPYMWATTGIGYNVDKVKAVLGENAPVDSWDLILKP ENLEKLKSCGVSFLDAPEEVFATVLNYLGKDPNSTKADDYTGPATDLL LKLRPNIRYFHSSQYINDLANGDICVAIGWAGDVWQASNRAKEAKNG VNVSFSIPKEGAMAFFDVFAMPADAKNKDEAYQFLNYLLRPDVVAHI SDHVFYANANKAATPLVSAEVRENPGIYPPADVRAKLFTLKVQDPKID RVRTRAWTKVKSGKLEHHHHHH
	Des1.1	MAEQKTLHIYNWSDYIAPDTIANFEKETGIKVVYDVFDSNEVLEGKL MAGSTGFDLVVPSASFLERQLTAGVFQPLDKSKLPNWKNLDPELLKL VAQHDPDNKYAIPYMWATTGIGYNVDKVKAVLGEDAPVDSWDLILKP ENLEKLKSCGVAFLDAPEEIFATVLNYLGKDPNSTNPDDYTGPATDLLL KLRPNIRYFHSSQYINDLANGDICVAIGWAGDVWQAANRAKEAKNGV NIAYTIPKEGAMAFFDVFAMPADAKNKDEAYQFLNYLLRPDVMAHIS DHVYYANANKAALPLVNPEVRDNPGIYPPPDVRAKLFTLKVQDPKID RVRTRAWTKVKSGKLEHHHHHH
	Des1.2	MAEQKTLHIYNWSDYIAPDTVANFEKETGIKVVYDVFDSNEVLEGKL MAGSTGFDLVVPSASFLERQLTAGVFQPLDKSKLPNWKNLDPELLKL VAKHDPDNKYAMPYMWATTGIGYNVDKVKAVLGEDAPVDSWDLILK PENLEKLKSCGVAFLDAPEEIFATVLNYLGKDPNSTNADDYTGPATDL LLKLRPNIRYFHSSQYINDLANGDICVAIGWAGDVWQAANRAKEAKN GVNVAYSIPKEGAMAFFDVFAMPADAKNKDEAYQFLNYLLRPDVVA HISNHVFYANANKAATSLVSAEVRDNPGIYPPPDVRAKLFTLKVQDPK IDRVRTRAWTKVKSGKLEHHHHHH
	Des1.3	MAEQKTLHIYNWSDYIAPDTVANFEKETGIKVVYDVFDSNEVLEGKL MAGSTGFDLVVPSASFLERQLSAGVFQPLDKSKLPNWKNLDPELLKL VAKHDPDNKYAMPYMWATTGIGYNVDKVKAVLGEDAPVDSWDLILK PENLEKLKSCGVAFLDAPEEIFATVLNYLGKDPNSTNADDYTGPATDL LLKLRPNIRYFHSSQYINDLANGDICVAIGWAGDVWQAANRAKEAKN GVNVAYSIPKEGAMAFFDVFAMPADAKNKDEAYQFLNYLLRPDVVA HISDHVFYANANKAATPLVSAEVRDNPGIYPPPDVRAKLFTLKVQDPK IDRVRTRAWTKVKSGKLEHHHHHH
	Des1.4	MAEQKTLHIYNWSDYIAPDTVANFEKETGIKVVYDVFDSNEVLEGKL MAGSTGFDLVVPSASFLERQLTAGVFQPLDKSKLPNWKNLDPELLKL VAKHDPDNKYAMPYMWATTGIGYNVDKVKAVLGEDAPVDSWDLILK PENLEKLKSCGVAFLDAPEEIFATVLNYLGKDPNSTNADDYTGPATDL LLKLRPNIRYFHSSQYINDLANGDICVAIGWAGDVWQAANRAKEAKN GVNVAFSIPKEGAMAFFDVFAMPADAKNKDEAYQFLNYLLRPDVVAH ISDHVFYANANKAATALVSAEVRENPGIYPPADVRAKLFTLKVQDPKI DRVRTRAWTKVKSGKLEHHHHHH
		MGSSHHHHHHGSGENLYFQSNQPLKIVVPFSAGGTADVLPRLVAEKIR ADYAGGVIIENKPGAGGNIGADLVFRAPPDGMTVLASPPGPIAINHNL YQKLSFDPTRWVPVTILATVPNVLVINPKLPVKSLGEFIAYAKANPKKV

TphC	Des2.0	TVATQGDGSTSHLTAAMFMQLTGTELTVIPYKGTAPALIDLIGGNVDVF FDNISSSATYHQAGKVRILAVADEQRSQILPQVPTFAEQQWPAMQAVT FFSVVAPPGTSAEIAQKLQKQMALALSSNDIRKHFQEQGAVPCGWDPS KTAQFIRQETEKWKKVLKAANVKL
	Des2.1	MGSSHHHHHHGSGENLYFQSNQPLKIVVPFSPGGTADVLPRLVAEKIR ADYGGGVIIENKPGAGGNIGADLVFRAPPDGYTVLISPPGPIAINHNLY KKLSFDPTRWVPVTILATVPNVLVINPKLPVKSVKEFIAYAKANPGKLT VATQGNGSTSHLTAELFMQMTGTDMTHIPYKGTAPALIDLIGGNVDVF FDNISSSATYHQAGKVRILAVADEQRSPILPDVPTFAEQGFPAFQVVTFF SVVAPPGTSAEIVQKLQKQIAKALSSPDIRKHFQEQGAVPVGWDPSQT AQFIRQETEKWKKVIKAANVKL
	Des2.2	MGSSHHHHHHGSGENLYFQSNQPLKIVVPFSPGGTADVLPRLVAEKIR KDYAGGVIIENKPGAGGNIGADLVFRAPPDGMTVLASPPGPIAINHNL YKKLSFDPTRWVPVTILATVPNVLVINPKLPVKSVKEFIAYAKANPGK VTVATQGNGSTSHLTAAMFMQLTGTDMTHIPYKGTAPALIDLIGGNVD VFFDNISSSAQYHQAGKVRILAVADEQRSPILPDVPTFAEQGFPAMQAV TFFSVVAPPGTSAEIAQKLQKQMAKALSSPDIRKHFQEQGAVPVGWDP EQTAQFIRQETEKWKKVLKAANVKL
	Des2.3	MGSSHHHHHHGSGENLYFQSNQPLKIVVPFSPGGTADVLPRLVAEKIR ADYAGGVIIENKPGAGGNIGADLVFRAPPDGMTVLASPPGPIAINHNL YKKLSFDPTRWVPVTILATVPNVLVINPKLPVKSVKEFIAYAKANPGK VTVATQGNGSTSHLTAAMFMQLTGTDMTHIPYKGTAPALIDLIGGNVD VFFDNISSSATYHQAGKVRILAVADEQRSPILPDVPTFAEQGFPAMQAV TFFSVVAPPGTSAEIAQKLQKQMAKALSSPDIRKHFQEQGAVPVGWDP SQTAQFIRQETEKWKKVLKAANVKL
	Des2.4	MGSSHHHHHHGSGENLYFQSNQPLKIVVPFSAGGTADVLPRLVAEKIR ADYAGGVIIENKPGAGGNIGADLVFRAPPDGYTVLASPPGPIAINHNLY QKLSFDPTRWVPVTILATVPNVLVINPKLPVKSLKEFIAYAKANPGKVT VATQGDGSTSHLTAAMFMQLTGTELTVIPYKGTAPALIDLIGGNVDVFF DNISSSATYHQAGKVRILAVADEQRSQILPQVPTFAEQGFPAMQAVTFF SVVAPPGTSAEIAQKLQKQMAKALSSPDIRKHFQEQGAVPCGWDPEQ TAQFIRQETEKWKKVLKAANVKL
MBP	Des3.0	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEK FPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKG KSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAK AGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNID TSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYL LTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMP NIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITKLEHHHHHH
	Des3.1	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEK FAQVAATGDGPDIIFWAHDRFGGWAQSGLLAEIHPDKEFQDKLFPFTW DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEELPELDKKLRAK GKSAIMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNNEGA KAGLQFLVDLIKNKHMNADIDYSIAEAAFNKGETAMTINGPWAWSNI DKAKINYGVTVLPTFKGKPSKPFVGVLSAGINAASPNKELAKEFLENY LLTDEGLDLVNKDKPLGAVALKSYQEELAKDPRIAATMENAQNGEIM

		PNIPQMSAFWYAMRTAVINALSGRQSVDEALKDAQTRITKLEHHHHH H
	Des3.2	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEK FTQVAATGDGPDIIFWAHDRFGGWAQSGLLAEIHPDKAFQDKFYPFTW DAVRYNGKLIAYPVAVEALSLIYNKDLLPNPPKTWEELPELDKELRKK GKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNNEGA KAGLQFLVDLIKNKHMNADIDYSIAEAAFNKGETAMTINGPWAWSNI DKAKINYGVTVLPTFKGKPSKPFVGVLTAGINAASPNKELAKEFLENY LLTDEGLDMVNKDKPLGAVALKSYQEELAKDPRIAATMENAQNGEIM PNIPQMSAFWYAMRTAVINALSGRQTVDEALKDAQKRITKLEHHHHH H
	Des3.3	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEK FAQVAATGDGPDIIFWAHDRFGGWAQSGLLAEIHPDKAFQDKLYPFTW DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEELPELDKELRAK GKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNNEGA KAGLQFLVDLIKNKHMNADIDYSIAEAAFNKGETAMTINGPWAWSNI DKAKINYGVTVLPTFKGKPSKPFVGVLSAGINAASPNKELAKEFLENY LLTDEGLDAVNKDKPLGAVALKSYQEELAKDPRIAATMENAQNGEIM PNIPQMSAFWYAMRTAVINALSGRQTVDEALKDAQTRITKLEHHHHH H
	Des3.4	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEK FAQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKFYPFTW DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEELPALDKELRAK GKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNNEGA KAGLQFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNI DKAKINYGVTVLPTFKGKPSKPFVGVLSAGINAASPNKELAKEFLENY LLTDEGLDAVNKDKPLGAVALKSYQEELAKDPRIAATMENAQKGEIM PNIPQMSAFWYAMRTAVINALSGRQTVDEALKDAQTRITKLEHHHHH H
LAO	Des4.0	MALPQTVRIGTDTTYAPFSSKDAKGEFIGFDIDLGNEMCKRMQVKCT WVASDFDALIPSLKAKKIDAIISSLSITDKRQQEIAFSDKLYAADSRLIA AKGSPIQPTLESLKGKHVGVLQGSTQEAYANDNWRTKGVDVVAYAN QDLIYSDLTAGRLDAALQDEVAASEGFLKQPAGKEYAFAGPSVKDKK YFGDGTGVGLRKDDTELKAAFDKALTELRQDGTYDKMAKKYFDFN VYGDLEHHHHHH
	Des4.1	MALPQTVRIGTDTTYPPFSSKDAKGEFVGFDIDLGNEMCKRMQVKCT WVESDFDALIPSLKAKKIDAIISSLSITDKRQQEIAFSDKLYAADARLIA PKGSPIQPTLESLKGKHVGVLQGSTQEAYANDHWRTKGVDIVPYQNQ DLIYSDLTAGRLDAALQDEVAASEGFLKQPAGKDYAFAGPSVKDKKY FGDGTGVGLRKDDNELKAAFDKALAELRKDGTYDKMAKKYFDFNV YGDLEHHHHHH
	Des4.2	MALPQTVRIGTDTTYPPFSSKDAQGQFVGFDIDLGNEMCKRMQVKCT WVESDFDALIPSLKAKKIDAIISSLSITEKRQQEIAFSDKLYAADSRLIAP KGSPIQPTLESLKGKHVGVLQGSTQEAYANDHWRSKGVDVVAYQNQ DLIYSDLTAGRLDAALQDEVAASEGFLKQPAGKDYAFAGPSVKDKKY FGDGTGVGLRKDDTELKAAFDKALAELRKDGTYDKMAKKYFDFNV YGDLEHHHHHH

	Des4.3	MALPQTVRIGTDTTYPPFSSKDAKGEFVGFDIDLGNEMCKRMQVKCT WVESDFDALIPSLKAKKIDAIISSLSITDKRQQEIAFSDKLYAADSRLIAP KGSPIQPTLESLKGKHVGVLQGSTQEAYANDHWRTKGVDVVAYQNQ DLIYSDLTAGRLDAALQDEVAASEGFLKQPAGKDYAFAGPSVKDKKY FGDGTGVGLRKDDTELKAAFDKALAELRKDGTYDKMAKKYFDFNV YGDLEHHHHHH
	Des4.4	MALPQTVRIGTDTTYAPFSSKDAKGEFVGFDIDLGNEMCKRMQVKCT WVESDFDALIPSLKAKKIDAIISSLSITDKRQQEIAFSDKLYAADSRLIAP KGSPIQPTLESLKGKHVGVLQGSTQEAYANDHWRTKGVDVVAYANQ DLIYSDLTAGRLDAALQDEVAASEGFLKQPAGKDYAFAGPSVKDKKY FGDGTGVGLRKDDTELKAAFDKALAELRKDGTYDKMAKKYFDFNV YGDLEHHHHHH

Supporting Table 2. Amino acid sequences of all proteins tested in this study. DesX.0 reflects the sequence of the respective WT of PotF, TphC, MBP and LAO.

Primer Name	**DNA sequence**
TWIST amp for	CAATCCGCCCTCACTACAACCG
TphC_des2.0_term_rev	TTCGCTCACCCTCGAGTTACAACTTTAC
TphC_des2.1_term_rev	TTCGCTCACCCTCGAGTTATAACTTCAC
TphC_des2.2_term_rev	TTCGCTCACCCTCGAGTTAAAGCTTTAC
TphC_des2.3_term_rev	TTCGCTCACCCTCGAGTTATAACTTTAC
TphC_des2.4_term_rev	TTCGCTCACCCTCGAGTTAGAGCTTGAC

Supporting Table 3. Primers used in this study.

Protein target	Design ID	T_M in ℃	ΔT_M in ℃ (- WT)	T_M in ℃ (+ ligand)	ΔT_M in ℃ (+ ligand)
PotF	Des1.0	63.21 ± 0.40	0	70.25 ± 0.19	7.04 ± 0.59
	Des1.1	76.11 ± 0.32	12.90 ± 0.72	77.12 ± 0.26	1.01 ± 0.58
	Des1.2	72.67 ± 0.34	9.46 ± 0.74	75.49 ± 0.22	2.82 ± 0.56
	Des1.3	72.75 ± 0.34	9.54 ± 0.74	74.67 ± 0.17	1.92 ± 0.51
	Des1.4	74.61 ± 0.61	11.40 ± 1.01	75.46 ± 0.33	0.85 ± 0.94
TphC	Des2.0	60.57 ± 0.06	0	66.22 ± 0.04	5.65 ± 0.10
	Des2.1	67.86 ± 0.04	7.29 ± 0.10	69.74 ± 0.04	1.88 ± 0.08
	Des2.2	67.85 ± 0.05	7.28 ± 0.11	74.06 ± 0.03	6.21 ± 0.08
	Des2.3	66.44 ± 0.06	5.87 ± 0.12	72.51 ± 0.05	6.07 ± 0.11
	Des2.4	66.98 ± 0.05	6.41 ± 0.11	73.27 ± 0.07	6.29 ± 0.12
MBP	Des3.0	63.10 ± 0.09	0	67.95 ± 0.07	4.85 ± 0.16
	Des3.1	79.97 ± 0.13	16.87 ± 0.22	83.49 ± 0.13	3.52 ± 0.26
	Des3.2	79.10 ± 0.10	16.00 ± 0.19	80.23 ± 0.10	1.13 ± 0.20
	Des3.3	78.18 ± 0.12	15.08 ± 0.21	82.50 ± 0.10	4.32 ± 0.22
	Des3.4	77.45 ± 0.17	14.35 ± 0.26	80.99 ± 0.31	3.54 ± 0.48
LAO	Des4.0	47.47 ± 0.10	0	55.84 ± 0.07	8.37 ± 0.17
	Des4.1	51.42 ± 0.06	3.95 ± 0.16	55.89 ± 0.08	4.47 ± 0.14
	Des4.2	54.62 ± 0.08	7.15 ± 0.18	59.91 ± 0.06	5.29 ± 0.14
	Des4.3	53.66 ± 0.07	6.19 ± 0.17	58.10 ± 0.07	4.44 ± 0.14
	Des4.4	54.35 ± 0.22	6.88 ± 0.32	59.05 ± 0.16	4.70 ± 0.38

Supporting Table 4. Thermal melt data of all PBP variants.

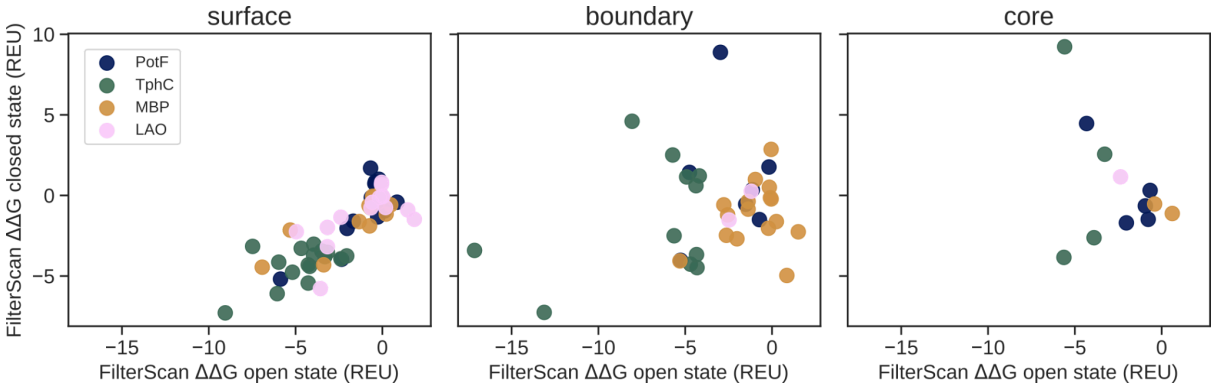
Protein target	Design ID	Experimental repeat	Protein concentration (µM)	Ligand concentration (µM)
PotF	buffer-putrescine	1 - 3	-	600
	Des1.0	1 - 3	10.5	95
	Des1.1	1	15.2	120
		2 - 3	14.7	120
	Des1.2	1 - 3	14.8	120
	Des1.3	1 - 3	15.7	120
	Des1.4	1 - 3	16.2	120
TphC	buffer-terephthalate	1 - 3	-	278
	Des2.0	1 - 3	38.8	278
	Des2.1	1 - 2	59.0	278
		3	52.5	320
	Des2.2	1 - 3	51.1	278
	Des2.3	1 - 3	48.0	278
	Des2.4	1 - 3	31.2	200
MBP	buffer-maltose	1 - 3	-	1000
	Des3.0	1 - 3	104.0	1000
	Des3.1	1 - 2	70.4	500
		3	57.5	500
	Des3.2	1	98.9	900
		2	73.9	800
		3	97.9	1000
	Des3.3	1 - 3	99.9	1000
	Des3.4	1 - 3	96.3	1000

LAO	buffer-lysine	1 - 3	-	300
	Des4.0	1	24	200
		2 - 3	17	150
	Des4.1	1 - 3	71	500
	Des4.2	1	69	500
		2 - 3	38	500
	Des4.3	1	68	500
		2 - 3	60	500
	Des4.4	1 - 3	16	150

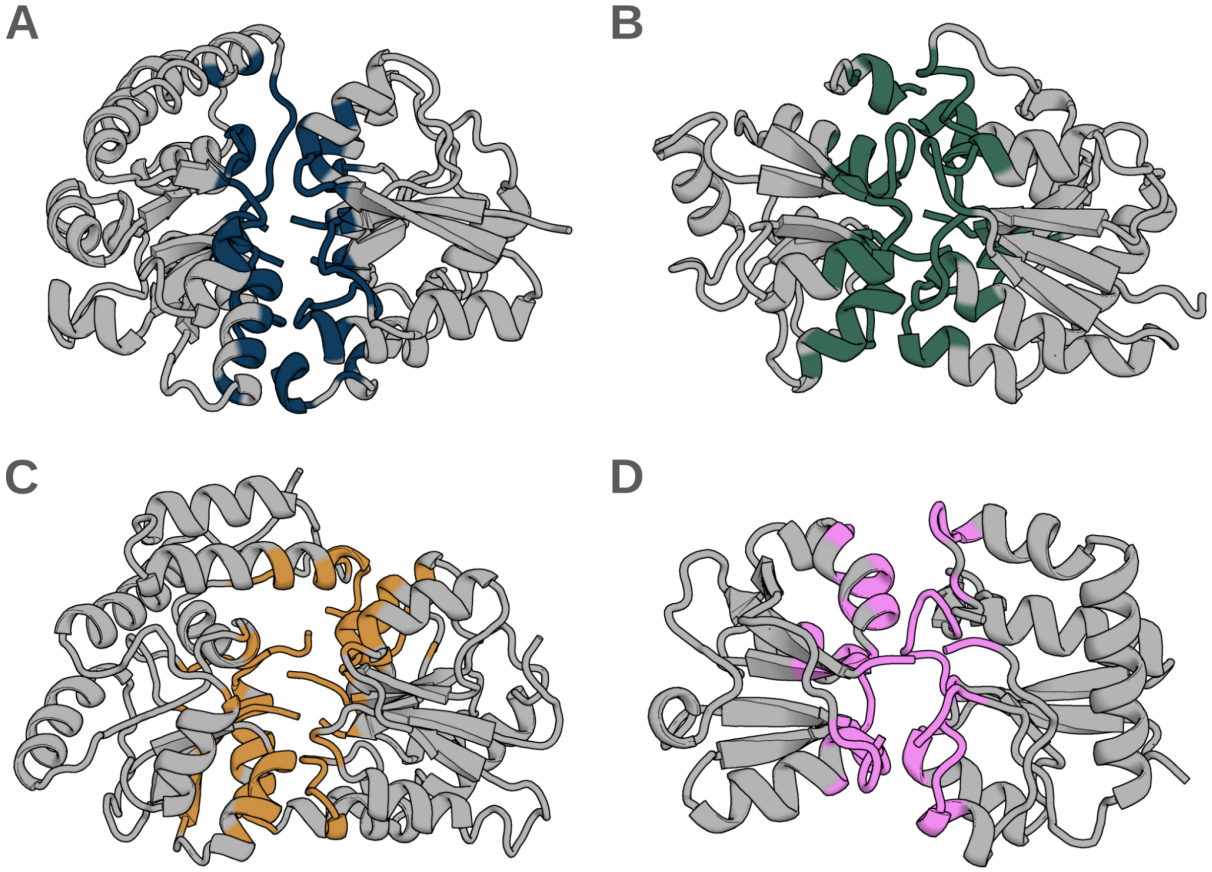
Supporting Table 5. Protein and ligand concentrations for ITC measurements

Protein target	Design ID	K_D (nM)	Δ*H* (kJ/mol)	-TΔ*S* (kJ/mol)	Δ*G* (kJ/mol)	N
PotF	Des1.0	41.66 ± 2.69	-88.92 ± 1.07	46.79 ± 1.23	-42.13 ± 0.16	0.92 ± 0.02
	Des1.1	479.1 ± 11.85	-76.17 ± 1.53	40.10 ± 1.59	-36.07 ± 0.06	0.97 ± 0.05
	Des1.2	42.27 ± 12.46	-87.26 ± 1.74	45.08 ± 2.53	-42.18 ± 0.80	0.92 ± 0.09
	Des1.3	101.0 ± 16.62	-82.66 ± 1.45	42.71 ± 1.08	-39.95 ± 0.43	0.95 ± 0.01
	Des1.4	103.1 ± 9.54	-88.82 ± 1.44	48.94 ± 0.43	-39.88 ± 0.22	0.88 ± 0.01
TphC	Des2.0	1180 ± 250	-43.07 ± 1.08	9.19 ± 1.53	-33.88 ± 0.53	0.95 ± 0.04
	Des2.1	5750 ± 709	-29.77 ± 3.89	-0.15 ± 4.13	-29.93 ± 0.30	0.88 ± 0.03
	Des2.2	627.0 ± 69.4	-36.05 ± 0.28	0.64 ± 0.55	-35.42 ± 0.27	0.92 ± 0.03
	Des2.3	716.0 ± 91.3	-36.65 ± 1.03	1.56 ± 1.28	-35.09 ± 0.31	0.92 ± 0.01
	Des2.4	281.7 ± 38.6	-45.58 ± 0.38	8.16 ± 0.68	-37.41 ± 0.35	0.97 ± 0.07
MBP	Des3.0	1240 ± 102	8.09 ± 0.01	-41.82 ± 0.20	-33.73 ± 0.21	0.96 ± 0.04
	Des3.1	1310 ± 293	-4.02 ± 0.27	-29.60 ± 0.81	-33.62 ± 0.54	1.15 ± 0.01
	Des3.2	in µM range	-	-	-	0.66 ± 0.03
	Des3.3	450 ± 43.6	3.83 ± 0.05	-40.07 ± 0.21	-36.24 ± 0.24	0.94 ± 0.03
	Des3.4	2960 ± 675	5.99 ± 0.32	-37.60 ± 0.37	-31.60 ± 0.60	1.02 ± 0.03
LAO	Des4.0	93.13 ± 34.54	-55.09 ± 0.89	14.86 ± 1.64	-40.23 ± 0.86	1.02 ± 0.01
	Des4.1	2496 ± 211.4	-20.23 ± 1.36	-11.76 ± 1.30	-31.99 ± 0.22	0.90 ± 0.00
	Des4.2	612.0 ± 56.29	-26.10 ± 0.22	-9.38 ± 0.12	-35.48 ± 0.22	0.91 ± 0.01
	Des4.3	824.0 ± 18.19	-25.44 ± 0.15	-9.29 ± 0.20	-34.73 ± 0.05	0.95 ± 0.03
	Des4.4	149.7 ± 20.60	-43.69 ± 0.52	4.71 ± 0.85	-38.97 ± 0.35	0.93 ± 0.04

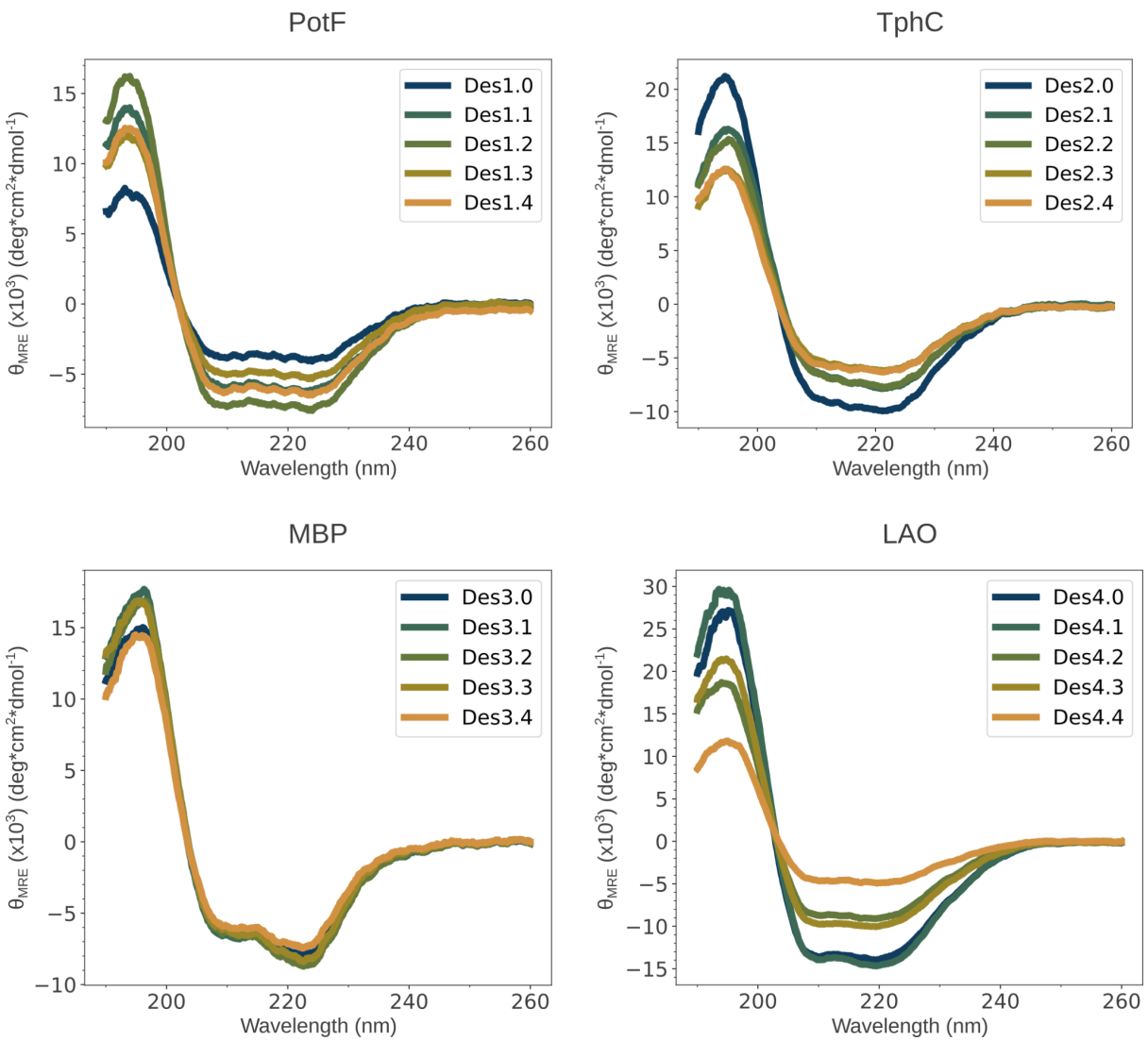
Supporting Table 6. ITC data of all PBP variants. Averages and standard deviations were calculated based on three technical replicas.



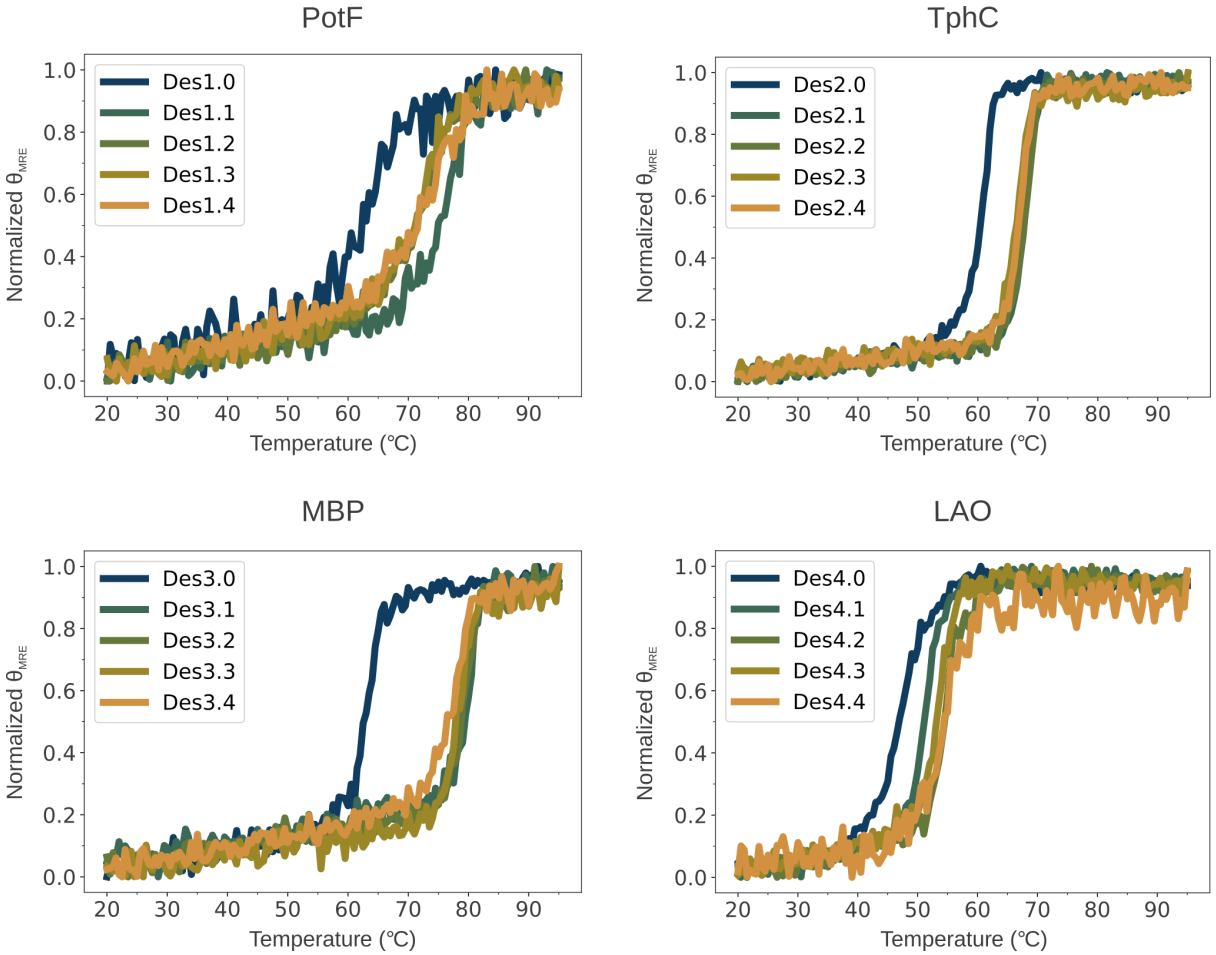
Supporting Figure 1. ΔΔ*G* values of PROSS mutations generated of PotF, TphC, MBP and LAO filtered by Rosetta LayerSelector categories: surface, boundary, and core. Open- and closed-state energies were plotted against each other.



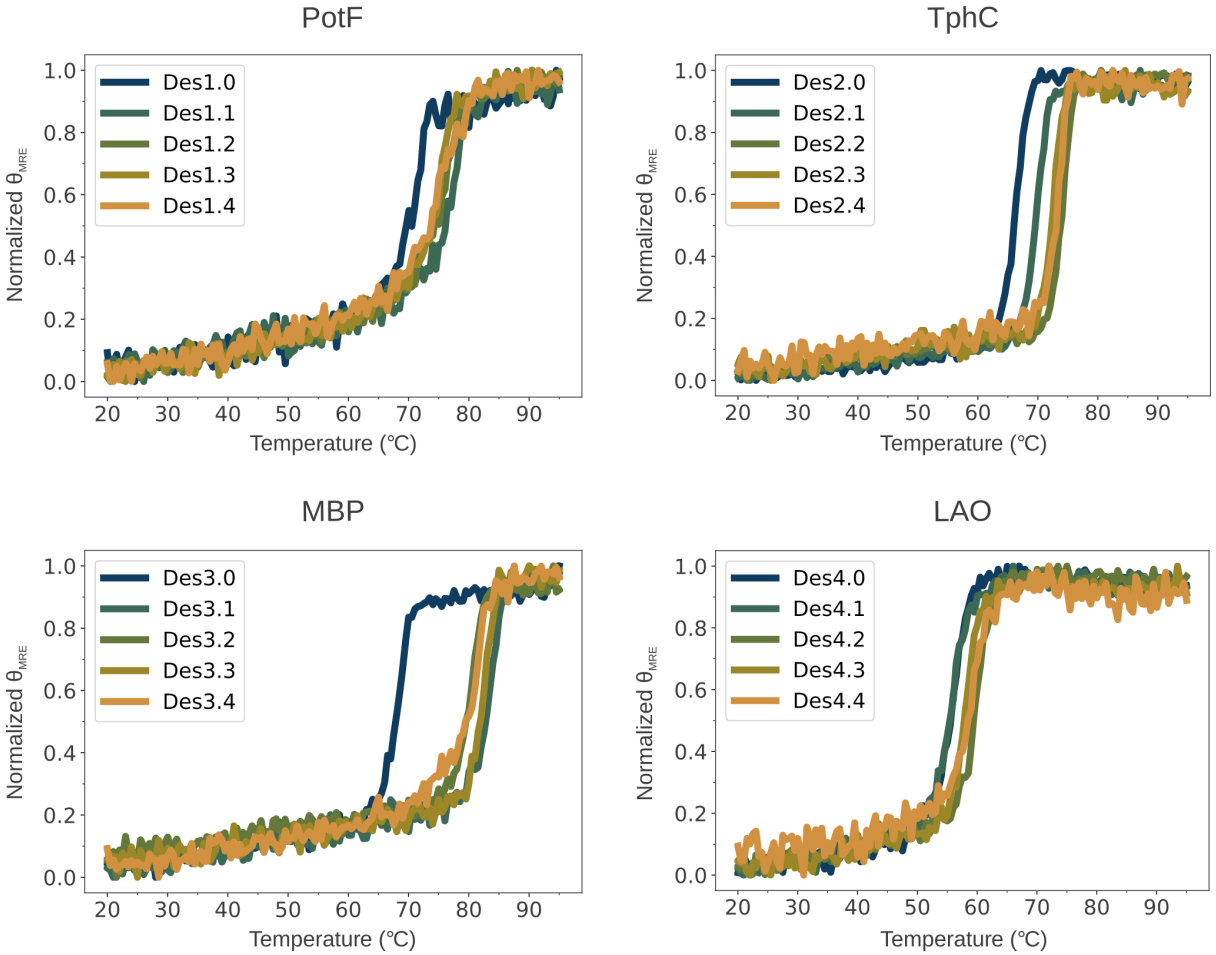
Supporting Figure 2. PBP interface residues that were fixed during PROSS design. (A-D) Closed-state PotF, TphC, MBP and LAO structures were split by predicted hinge residues and subdomain interface residues were highlighted in blue, green, orange and pink, respectively.



Supporting Figure 3. CD spectra of all PBP variants without ligands present plotted as mean residue ellipticity (θ_{MRE}) against the wavelength.

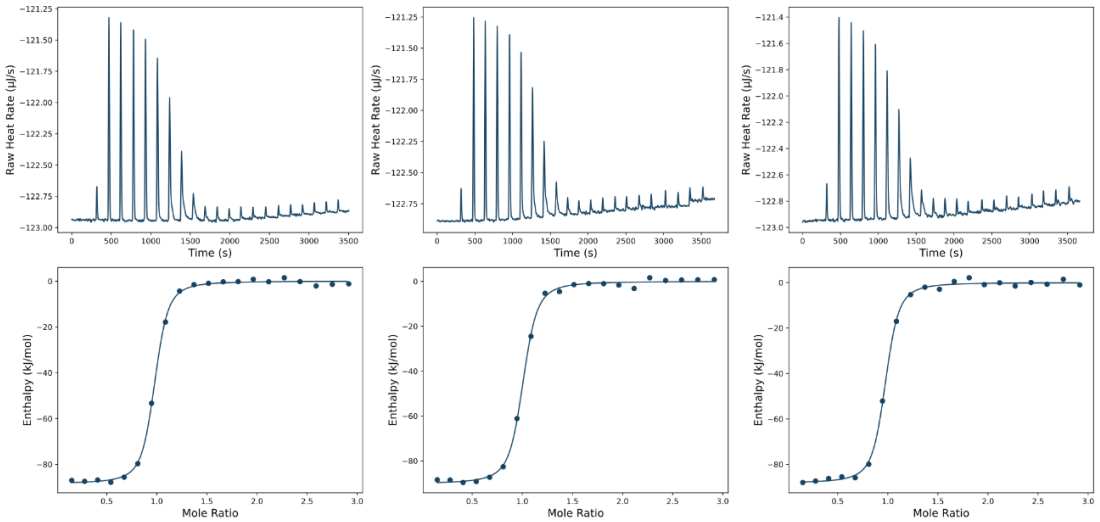


Supporting Figure 4. CD thermal melts of all PBP variants without ligands present plotted as normalised mean residue ellipticity (θ_{MRE}) against a temperature gradient.

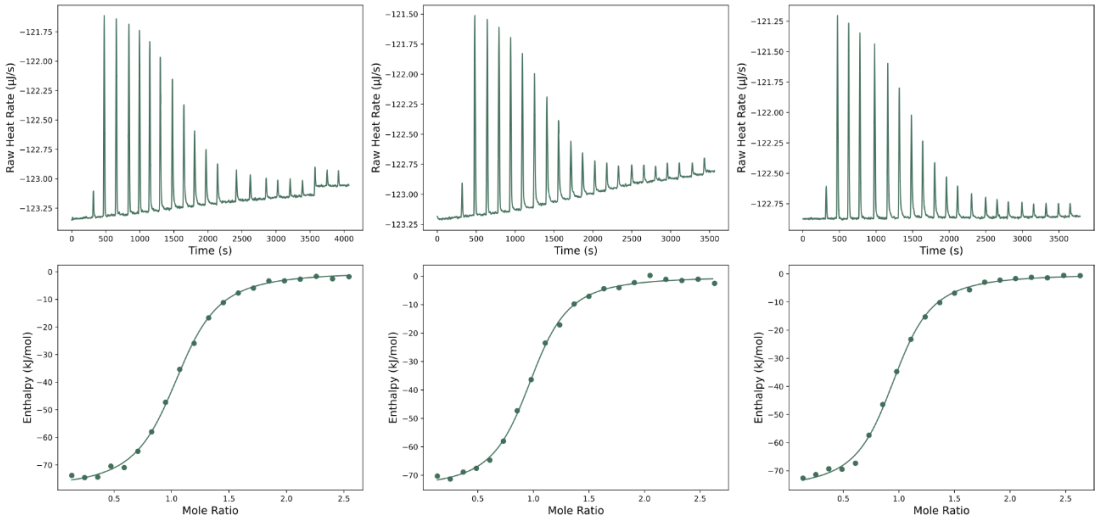


Supporting Figure 5. CD thermal melts of all PBP variants with ligands present plotted as normalised mean residue ellipticity (θ_{MRE}) against a temperature gradient.

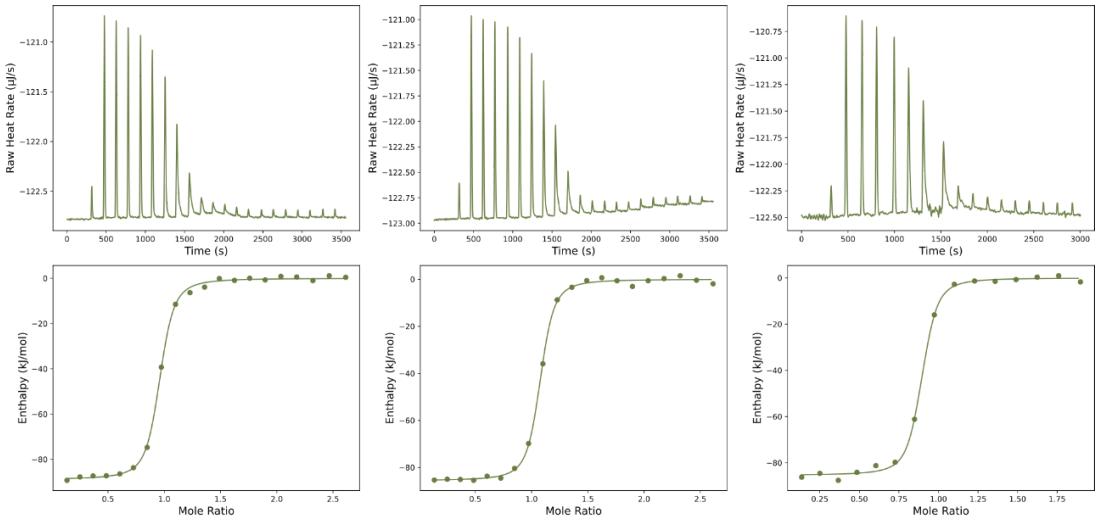
PotF Des1.0

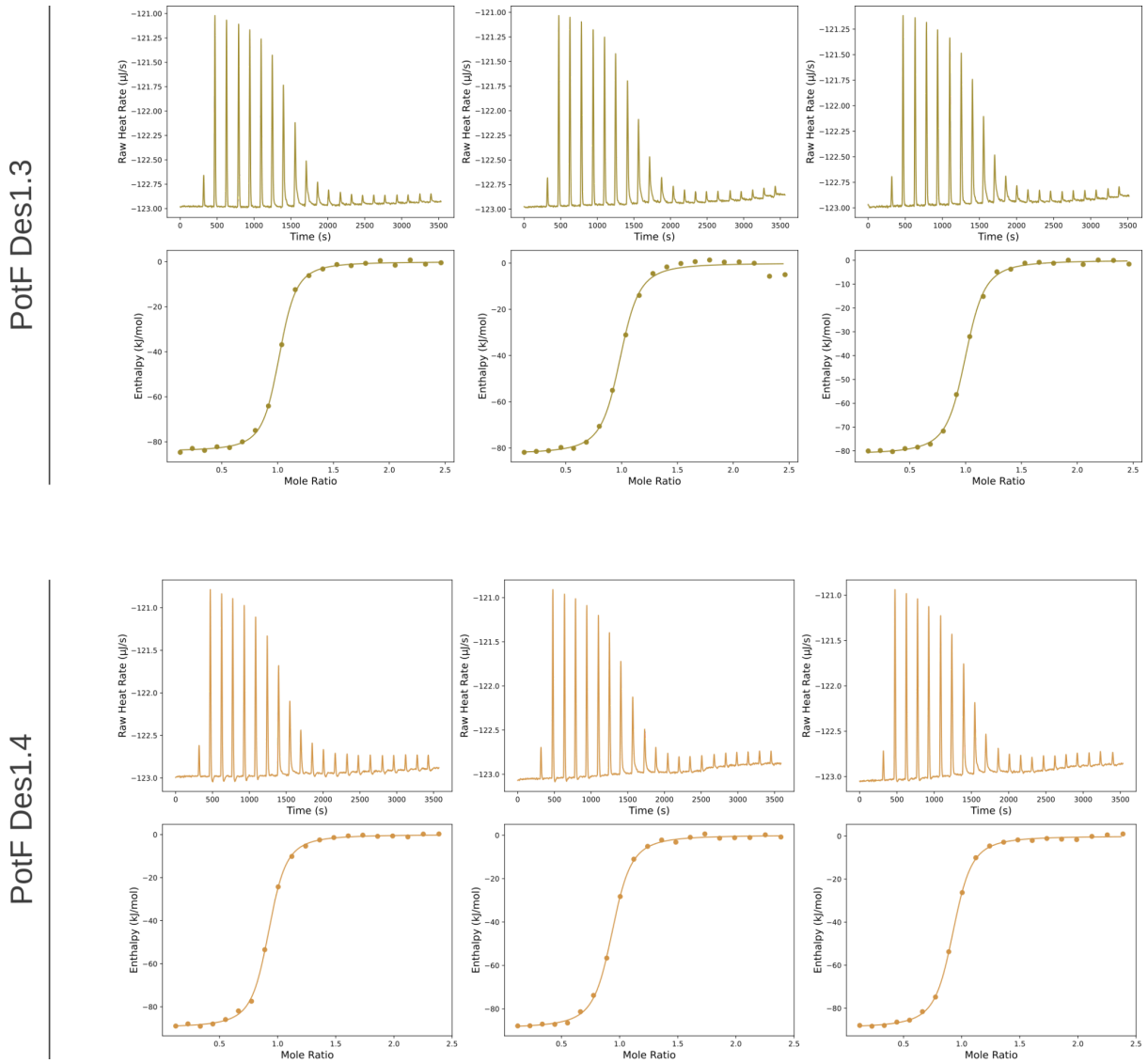


PotF Des1.1



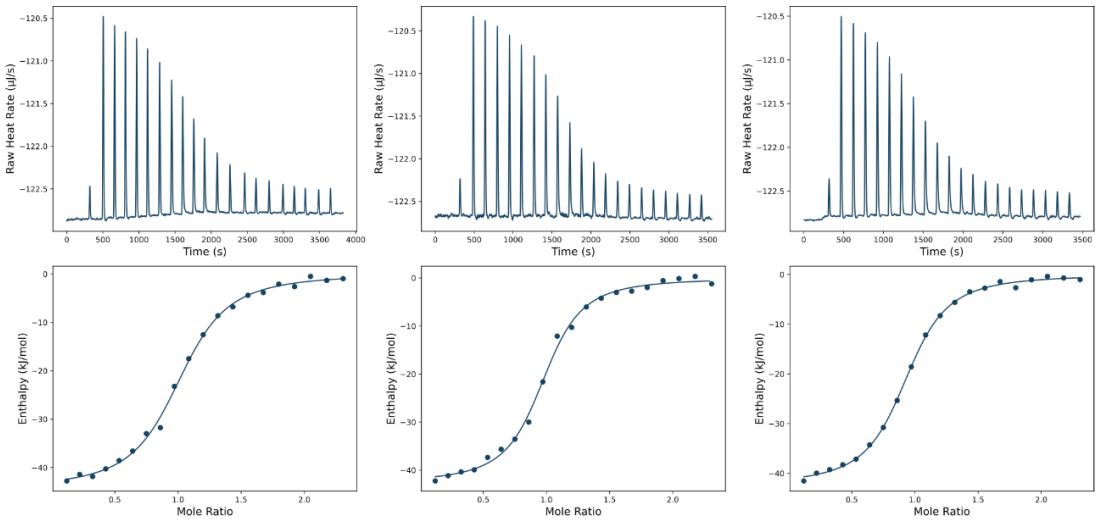
PotF Des1.2



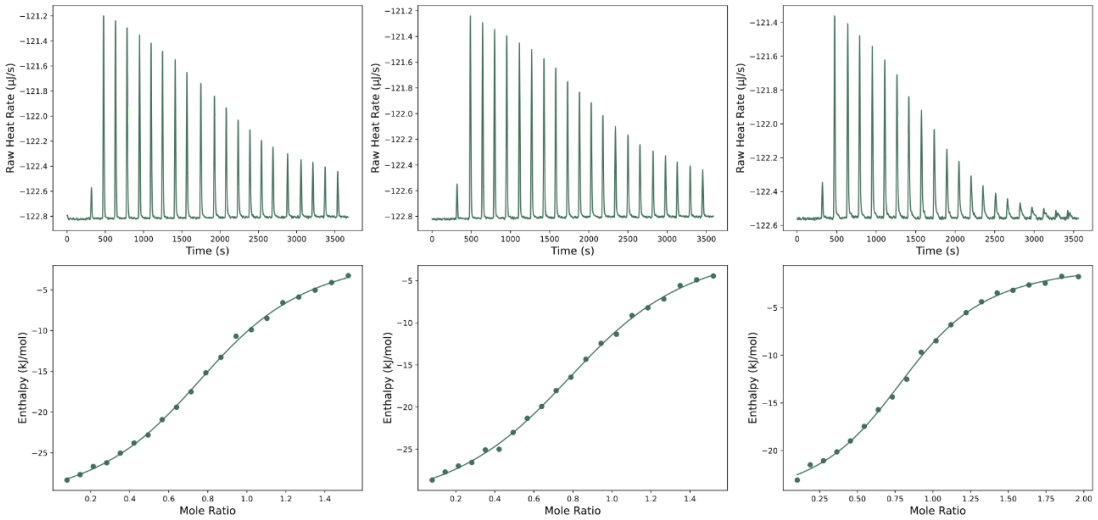


Supporting Figure 6. Thermograms and binding isotherms of PotF Des1.0-1.4 and putrescine in three technical replicates. Buffer-putrescine control titrations were performed to obtain baseline values that were subtracted from the integrated heats of PotF Des1.0-1.4.

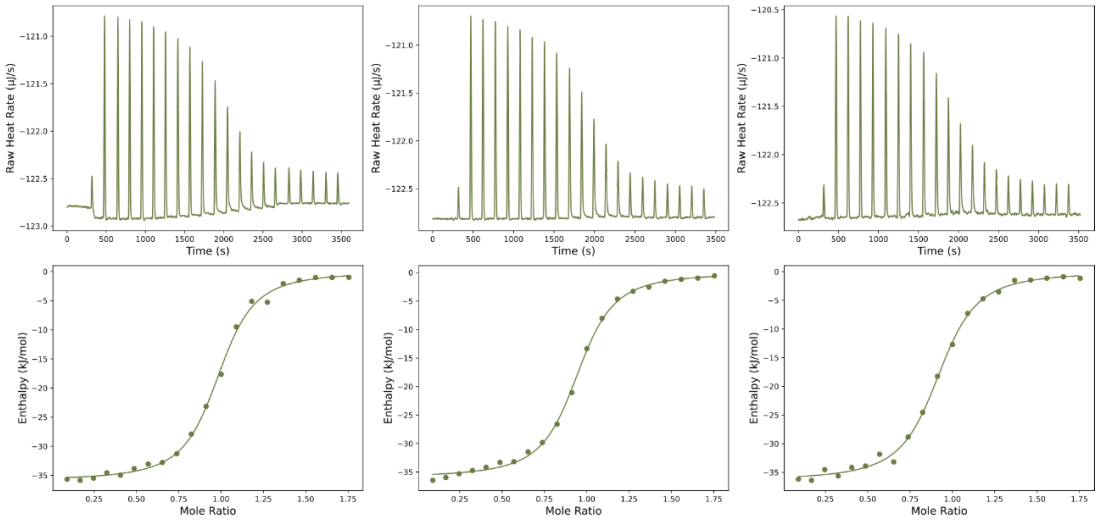
TphC Des2.0

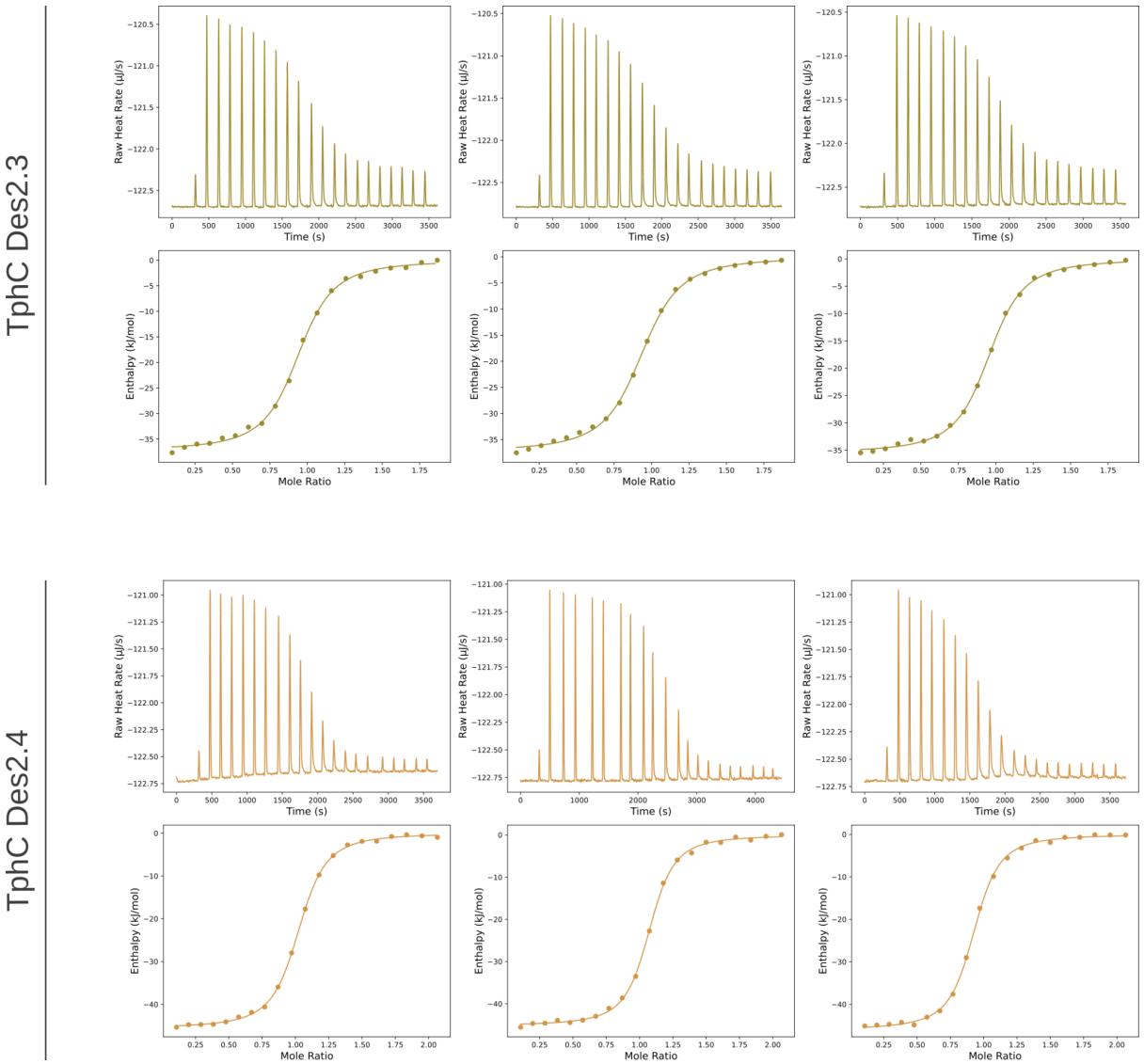


TphC Des2.1



TphC Des2.2



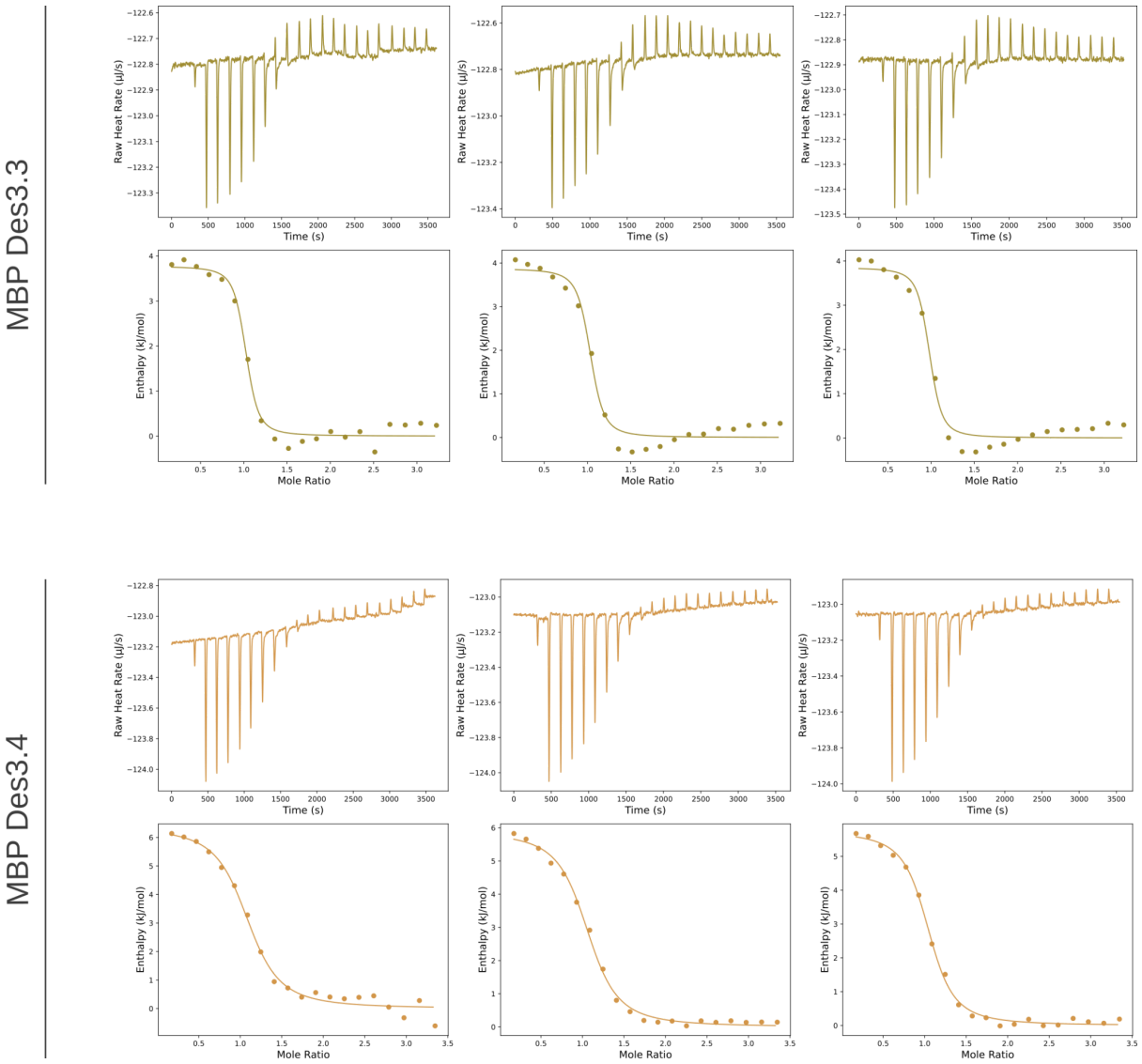


Supporting Figure 7. Thermograms and binding isotherms of TphC Des2.0-2.4 and terephthalate in three technical replicates. Buffer-terephthalate control titrations were performed to obtain baseline values that were subtracted from the integrated heats of TphC Des2.0-2.4.

MBP Des3.0

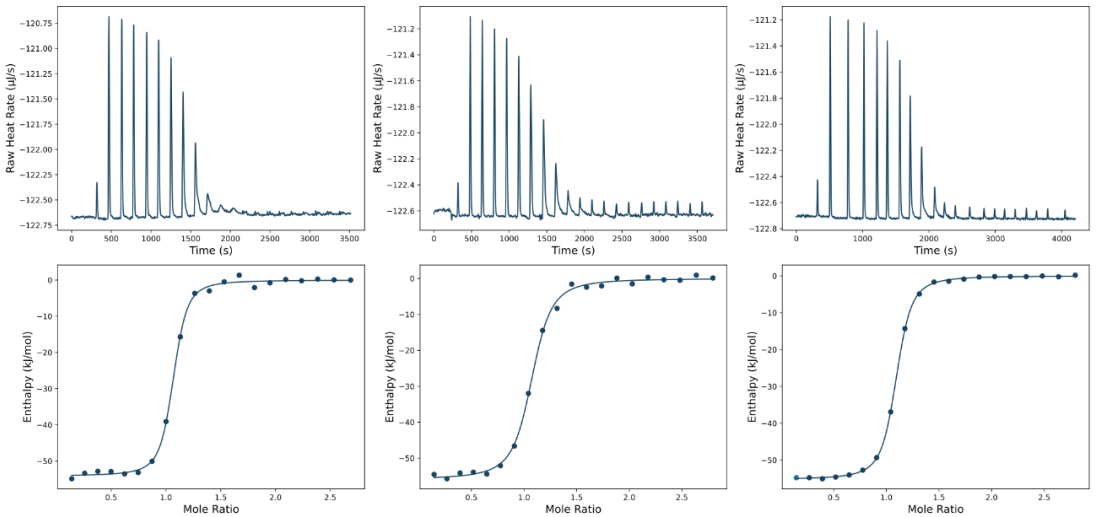
MBP Des3.1

MBP Des3.2

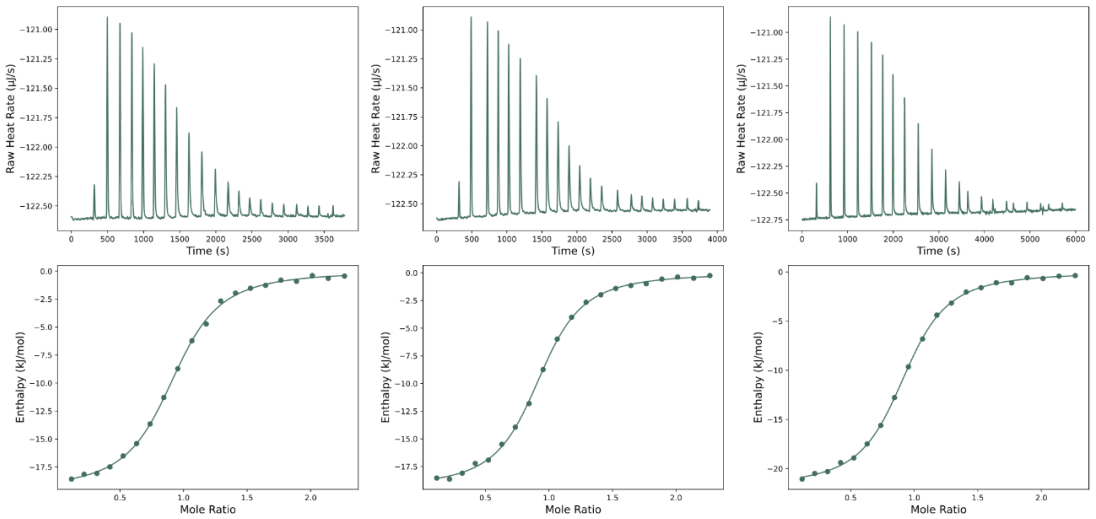


Supporting Figure 8. Thermograms and binding isotherms of MBP Des3.0-3.4 and maltose in three technical replicates. Buffer-maltose control titrations were performed to obtain baseline values that were subtracted from the integrated heats of MBP Des3.0-3.4.

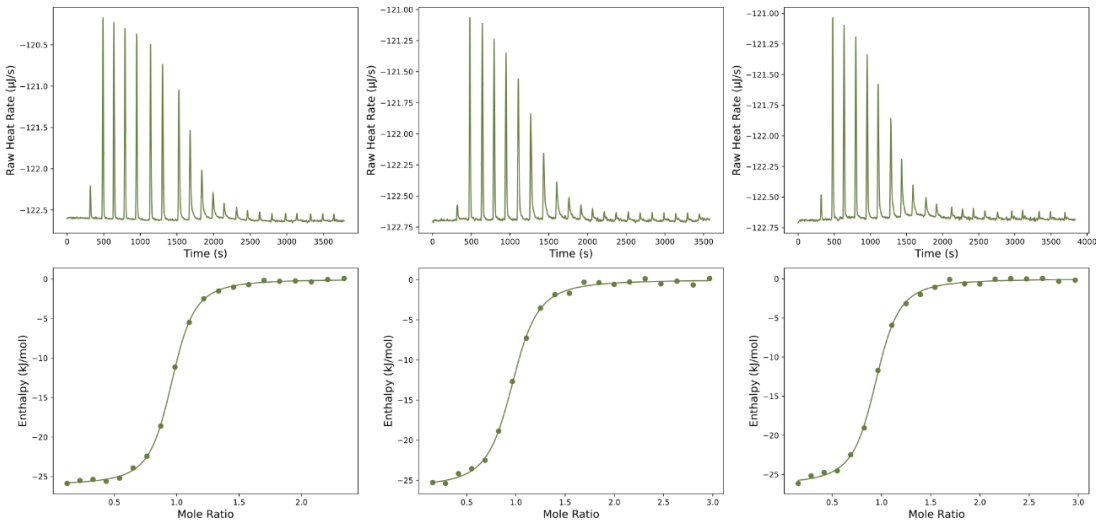
LAO Des4.0

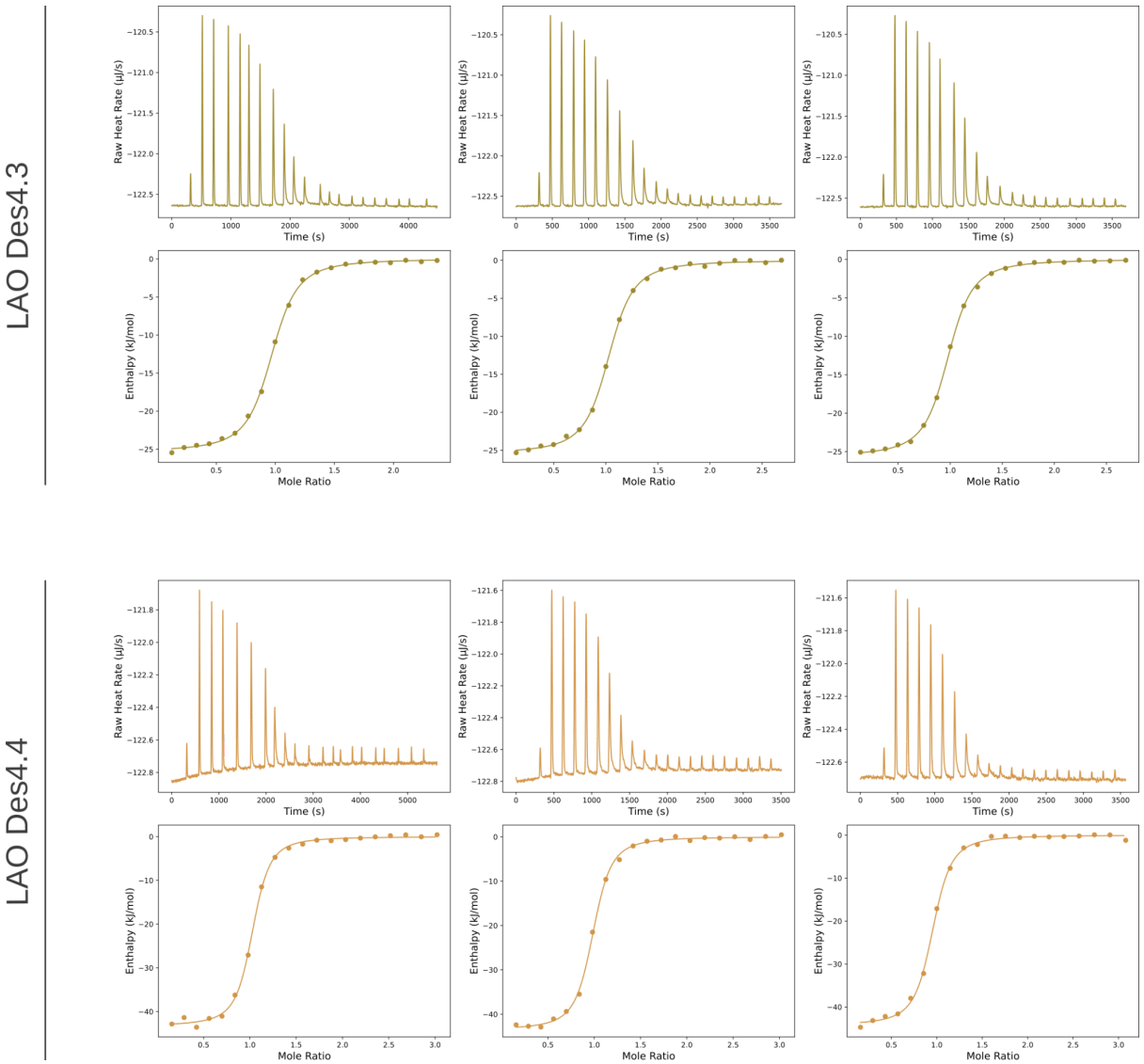


LAO Des4.1



LAO Des4.2





Supporting Figure 9. Thermograms and binding isotherms of LAO Des4.0-4.4 and L-lysine in three technical replicates. Buffer-lysine control titrations were performed to obtain baseline values that were subtracted from the integrated heats of LAO Des4.0-4.4.

References

1. Kröger, P., Shanmugaratnam, S., Ferruz, N., Schweimer, K. & Höcker, B. A comprehensive binding study illustrates ligand recognition in the periplasmic binding protein PotF. *Structure* **29**, 433-443.e4 (2021).
2. Gautom, T. *et al.* Structural basis of terephthalate recognition by solute binding protein TphC. *Nat. Commun.* **12**, 6244 (2021).
3. Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31**, 10657–10663 (1992).
4. Quiocho, F. A., Spurlino, J. C. & Rodseth, L. E. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997–1015 (1997).
5. Oh, B. H. *et al.* Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J. Biol. Chem.* **268**, 11348–11355 (1993).
6. Oh, B. H., Ames, G. F. & Kim, S. H. Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. *J. Biol. Chem.* **269**, 26323–26330 (1994).
7. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).